



UNIVERSIDAD NACIONAL DE CÓRDOBA

FACULTAD DE LENGUAS

DOCTORADO EN CIENCIAS DEL LENGUAJE

MENCIÓN EN LINGÜÍSTICA APLICADA

*El error de concordancia plural en español L2
desde una perspectiva emergente*

TESIS DE DOCTORADO

DOCTORANDO: PABLO EZEQUIEL MARAFIOTI

DIRECTOR DE TESIS: CARLOS JULIO REYNOSO (UBA)

CO-DIRECTOR DE TESIS: LAURA ALONSO ALEMANY (UNC)

CÓRDOBA

JUNIO DE 2020

Agradecimientos

A MI PADRE JUAN Y A MI HERMANO FERNANDO,
POR SU CONSTANTE APOYO DURANTE LOS AÑOS DE DOCTORADO.

Resumen

En esta tesis se analizan, desde una perspectiva emergente, los errores de concordancia plural en cuatro aprendientes italianos de español L2: SONIA (nivel A), NATI (nivel B1), JAKO (nivel B2), MIRKA (nivel C1). Los objetivos principales son: (i) examinar los factores relacionados con la chance / riesgo de error; (ii) analizar la dinámica del error a nivel microscópico y macroscópico desde el enfoque de los sistemas complejos; (iii) predecir el error local de las últimas sesiones de los aprendientes. La orientación de esta tesis es fuertemente cuantitativa. Se han utilizado técnicas del ámbito de la estadística, la minería de datos y la física de los sistemas complejos. En cuanto al primer objetivo, se crearon las siguientes variables predictoras: (i) tipo de modificador (artículo definido, artículo indefinido, determinantes, adjetivos atributivos); (ii) si es a larga distancia; (iii) de más de dos términos; (iv) presencia de -e- epentética (controladores terminados en consonante); (v) propiedades del controlador (animicidad, concretud, familiaridad, imaginabilidad, frecuencia), (vi) similitud entre las terminaciones del español y el italiano; (vii) similitud entre las raíces léxicas de ambas lenguas; (viii) errores acumulados hasta la instancia en cuestión; (ix) posibles estrategias de aprendizaje (de 1 a 7), (x) frecuencia de TYPES de instancias de concordancia en corpus *EsTenTen* y en el propio. Las variables respuesta fueron: (i) binaria [error / no error]; (ii) categórica [error de: género, -e- epentética, plural, mixto]; (iii) tiempo hasta que se produce un error; (iv) serie simbólica. En general se hallaron efectos de: (a) tipo de modificador: los errores suben con determinantes / adjetivos respecto del artículo definido; (b) género: plurales masculinos más fáciles que los femeninos; (c) familiaridad / frecuencia del controlador [errores bajan]; (d) animicidad [errores suben para animados]; (e) frecuencia de TYPES [error baja a más frecuencia]. El efecto de -e- epentética fue facilitador, contra las expectativas. Se interpretó que la estrategia 5, que hacía bajar el error, tuvo efecto benéfico para los plurales en *-es*. La distancia de terminaciones evidenció el efecto inverso al esperado, se lo explicó analizando las instancias del nivel de referencia. El efecto de errores acumulados resultó débil. Respecto del segundo objetivo, la hipótesis principal fue tratar al error como atractor. A nivel microscópico, se utilizaron los resultados de los análisis estadísticos para sesgar el flujo a atractores en tres simulaciones basadas en sistemas dinámicos. En general, se logró emular el patrón global de error pero se aproximó menos el patrón de error por sesiones. En cuanto al nivel macroscópico, se usaron medidas de detección de cambio de régimen y redes complejas. Se logró identificar grupos de sesiones con dinámica similar y regiones aproximadas de transiciones, usando la variable respuesta de serie simbólica. Las redes complejas arrojaron efectos a nivel de las palabras respecto de: modificador, familiaridad / frecuencia y de imaginabilidad / concretud del controlador, -e- epentética y similitud entre raíz y desinencia. También hubo efectos para las palabras terminadas en *-e*. Para el tercer objetivo se emplearon, entre otras predictoras, información proveniente de las redes complejas. En general no se logró superar el 80% de precisión. Los atributos derivados del grafo fueron seleccionados como influyentes para todos los alumnos.

Índice general

1. Introducción.	1
1.1. El emergentismo en el lenguaje.	1
1.2. Hipótesis general y Objetivos.	2
1.3. Resumen de los capítulos.	3
2. La concordancia en español.	7
2.1. Género y número en español e italiano.	7
2.2. Lingüística de la concordancia.	8
2.2.1. Encuadre tipológico del fenómeno.	8
2.2.2. Concordancia según el programa minimalista.	15
2.2.3. Concordancia según gramáticas con estructuras de rasgos: HPSG, LFG.	17
2.2.3.1. Lexical Functional Grammar (LFG).	17
2.2.3.2. Head-driven Phrase Structure Grammar (HPSG).	19
2.2.4. Concordancia según la gramática de las construcciones.	22
2.3. Adquisición de la concordancia (L1).	24
2.4. Adquisición de la concordancia (L2).	25
2.4.1. Estudios <i>offline</i>	26
2.4.2. Estudios <i>online</i>	30
2.4.3. Aprendizaje de dependencias no adyacentes.	37
2.4.4. Resumen de los hallazgos.	40
3. Sistemas Dinámicos Complejos en L2.	43
3.1. Atención y Memoria.	43
3.2. El modelo de producción bilingüe.	44
3.3. Teorías de adquisición de L2.	47
3.3.1. Gramática universal.	47
3.3.2. Funcionalismo orientado a los conceptos.	48
3.3.3. Teoría de la procesabilidad.	49
3.3.4. Aprendizaje basado en el uso y en el conocimiento declarativo / procedural.	50
3.4. Sistemas Dinámicos Complejos (SDC).	52
3.5. Pensar la adquisición desde los SDC	56
3.5.1. Relación entre SDC y adquisición de segunda lengua.	56
3.5.2. Implicancias metodológicas e investigación en SDC.	58
3.5.3. Relación con otros enfoques.	60
3.5.3.1. La naturaleza del aprendizaje: incidental, implícito y que va más allá del input recibido.	60
3.5.3.2. La naturaleza de la inter-lengua: sistemática y con variabilidad dentro y entre subsistemas.	61
3.5.3.3. El rol de la primera lengua.	62
3.5.3.4. El rol del input y el output (producción).	62
3.5.3.5. El rol de la instrucción formal.	64
3.6. La concordancia como sistema dinámico.	65

3.6.1.	Un modelo para L1.	65
3.6.2.	El error de concordancia en L2 como atractor. Estrategia de análisis.	67
3.7.	Apéndice de capítulo.	69
3.7.1.	Análisis del modelo logístico con «depredación».	69
4.	Recolección de datos y creación de variables.	71
4.1.	Recolección de datos.	71
4.2.	Transcripción y codificación del corpus.	72
4.3.	Variables como respuesta y atributos.	75
4.3.1.	Variables «respuesta».	75
4.3.2.	Variables «atributos».	76
4.4.	Pre-procesamiento.	85
4.4.1.	Datos faltantes.	85
4.4.2.	Transformaciones.	87
4.4.3.	Colinealidad.	88
4.4.4.	Discretización.	90
5.	Análisis exploratorio.	95
5.1.	Agrupamiento de casos.	95
5.2.	Gráficas descriptivas.	102
5.2.1.	Respuesta binaria.	104
5.2.2.	Respuesta categórica.	105
5.3.	Análisis de Correspondencias Múltiple [MCA].	106
5.4.	Selección de variables.	108
5.4.1.	Selección univariada.	109
5.4.2.	Selección basada en eliminación recursiva de atributos con modelos de minería de datos [RFE].	111
5.4.3.	LASSO.	113
5.5.	Resumen.	115
5.6.	Apéndice de capítulo.	117
5.6.1.	Selección univariada de variables.	117
6.	Análisis de respuesta binaria.	119
6.1.	El modelo logístico.	119
6.2.	El modelo logístico mixto.	124
6.3.	Selección de modelos.	125
6.4.	Expectativa de análisis.	127
6.5.	Análisis de datos.	129
6.5.1.	Modelo logístico.	129
6.5.2.	Modelo logístico mixto.	131
6.5.3.	Modelo final: chequeo de supuestos y resultados.	134
6.6.	Análisis de errores.	140
6.7.	Conclusiones.	142
6.8.	Apéndice de capítulo.	143
7.	Análisis de respuesta categórica.	147
7.1.	Marco teórico.	147
7.1.1.	Distribución multinomial.	147
7.1.2.	Test de chi-cuadrado.	147
7.1.3.	Modelo logístico multinomial (mixto).	148
7.1.4.	Modelo multinomial mixto bayesiano.	149
7.2.	Expectativa del análisis.	151
7.3.	Selección de modelos.	153

7.4.	Modelo Bayesiano. Resultados y chequeo de supuestos.	154
7.5.	Análisis de errores.	161
7.5.1.	Error de género vs. no error.	161
7.5.2.	Error de <i>-e-</i> epentética vs. no error.	162
7.5.3.	Error de <i>plural</i> vs. no error.	163
7.5.4.	Error <i>mixto</i> vs. no error.	164
7.6.	Conclusiones.	165
7.7.	Apéndice de capítulo.	166
7.7.1.	Gráficas de mosaico	166
7.7.2.	Ajuste del modelo multinomial bayesiano.	168
8.	Análisis de respuesta de tiempo hasta el error	171
8.1.	Nociones de análisis de datos de sobrevivencia.	171
8.1.1.	Función de sobrevivencia y de riesgo.	172
8.1.2.	El modelo semi-paramétrico de Cox.	172
8.1.3.	Residuos.	174
8.2.	Expectativa de análisis.	175
8.3.	El modelo de eventos múltiples.	176
8.3.1.	Pre-procesamiento y marco teórico.	176
8.3.2.	Selección de modelos.	178
8.3.3.	Modelo final: chequeo de supuestos y resultados.	179
8.4.	El modelo de riesgos competitivos.	182
8.4.1.	Selección de modelos.	183
8.4.2.	Modelo final: chequeo de supuestos y resultados.	185
8.5.	Conclusiones.	189
8.6.	Apéndice de capítulo.	191
9.	Simulación con sistemas dinámicos.	193
9.1.	Introducción.	193
9.2.	Modelo competitivo de Lokta-Volterra.	194
9.2.1.	Aplicación y resultados.	196
9.3.	Teoría de los juegos evolutiva.	198
9.3.1.	Dinámica del replicador.	202
9.3.2.	Aplicación y resultados.	204
9.4.	Ascenso de gradiente de armonía.	207
9.4.1.	Aplicación y resultados.	208
9.5.	Conclusiones.	212
9.6.	Apéndice de capítulo.	213
9.6.1.	Modelo de Lokta-Volterra: análisis de estabilidad.	213
9.6.2.	Dinámica del replicador: análisis de estabilidad.	214
9.6.3.	Ascenso de gradiente de armonía: análisis de bifurcación.	215
10.	Dinámica de series de error.	219
10.1.	Introducción.	219
10.2.	“Motivos”.	219
10.2.1.	Motif y LDA.	219
10.2.2.	Análisis y evaluación de resultados.	221
10.3.	El plano entropía - complejidad.	223
10.3.1.	Entropía de permutación.	224
10.3.2.	Significancia de las medidas.	225
10.3.3.	Resultados.	225
10.3.4.	Análisis de cambio de punto.	227
10.4.	Análisis cuantitativo de recurrencias (RQA)	229

10.4.1. Gráficos de recurrencias.	229
10.4.2. Medidas a partir de recurrencias.	231
10.4.3. Significancia de las medidas.	234
10.4.4. Cambios de dinámica en las sesiones.	235
10.4.5. Recurrencias sobre instancias correctas y erróneas.	239
10.5. Indicadores de cambio de régimen.	239
10.5.1. Cambios de régimen mediante ventanas móviles para recurrencias.	241
10.5.1.1. Coeficiente AR(1) y desvío típico.	241
10.5.1.2. Heteroscedasticidad condicional.	243
10.5.1.3. Modelo de Drift-Diffusion-jump.	244
10.5.1.4. Análisis de Potenciales.	247
10.5.1.5. Resumen de resultados de los indicadores.	247
10.6. Conclusiones.	248
10.7. Apéndice de capítulo.	250
11. Dinámica en redes complejas y predicción de error.	253
11.1. Nociones sobre grafos.	253
11.2. Redes de concordancia.	256
11.3. Análisis estadístico de evolución temporal.	260
11.4. Predicción de estatus de error.	265
11.5. Mixtura de Bernoulli y regularización.	271
11.5.1. Clustering por mixtura de distribuciones.	271
11.5.2. Modelo mixto de regularización.	273
11.5.3. Resultados.	274
11.6. Conclusiones.	279
11.7. Apéndice de capítulo.	281
11.7.1. Comparaciones múltiples.	281
12. Comentarios didácticos y conclusiones.	283
12.1. Feedback.	283
12.1.1. Interacción y Feedback.	283
12.1.2. Feedback para la concordancia.	287
12.2. Discusión.	288
12.3. A modo de cierre: algunos consejos para la enseñanza de la concordancia.	293

Capítulo 1

Introducción.

1.1. El emergentismo en el lenguaje.

La presente tesis se enmarca en la corriente de pensamiento emergentista (MacWhinney, 2015). La idea central consiste en que la estructura lingüística (en diferentes niveles) emerge a partir de patrones de uso repetido en el tiempo. El análisis de tipo emergentista se funda en tres enfoques principales derivados de la teoría de sistemas adaptativa.

El primero es el de la teoría de la evolución. Existen similitudes entre la genética de las poblaciones y el lenguaje (Croft, 2000; Blythe & McKane, 2007). Un gen puede verse como una variable que codifica una característica específica (p. e.: el color de los ojos, o un rasgo gramatical), sus diferentes variables se llaman alelos (p. e.: diferentes valores de dicho rasgo). La frecuencia de los alelos cambia debido a: (1) la replicación, en la que los hijos adquieren copias de alelos de sus padres (los rasgos gramaticales se replican en cada enunciado producido); (2) mutaciones: cambios estocásticos entre tipos de variantes (alelos) o creación de nuevas; (3) la selección: una determinada variante es favorecida y aumenta su frecuencia, eventualmente se convierte en la más usada y se fija en el hablante / comunidad. Los hablantes podrían constituir ‘islas’, y cuando conversen con otros hablantes / islas se producirán ‘migraciones’ de variantes. O sea que las frecuencias de las variantes en los hablantes se modifican por la propia producción o por la producción de los otros hablantes (migraciones). Las variantes compiten por el uso de los miembros de la comunidad lingüística como si fuera una competición de especies similares. Se nota inmediatamente el parecido con el proceso de adquisición en L2: el aprendiente replica variantes, genera variación / mutaciones (‘errores’), una variante es la más favorecida porque la usa la comunidad lingüística de L1, y, eventualmente, se fija en la producción del aprendiente adquiriéndola. Por otro lado, en cada producción individual compiten diferentes soluciones a ser proferidas, una correcta y otra(s) incorrecta(s).

El segundo enfoque propone analizar a los sistemas complejos en términos de una jerarquía organizada por niveles, en la cual las propiedades de los niveles superiores no se pueden predecir completamente a partir de los inferiores. La complejidad surge de la recombinación de pequeñas “piezas” en estructuras más grandes, que no constituyen la suma de sus partes. Por ejemplo, los patrones sintácticos pueden codificarse en el nivel elemental como ítems léxicos, agrupándose luego en construcciones de diferentes niveles de abstracción. Esto significa que no existe diferencia entre léxico y sintaxis. Los niveles del lenguaje se hallan interconexos y la descomponibilidad total resulta imposible.

El tercero es una teoría de marcos temporales, la cual establece que los procesos que suceden en los diferentes niveles se asocian a tiempos de diferente resolución. En primera instancia se encuentra el tiempo de procesamiento *online* en el momento del habla. En segundo lugar, el tiempo de la consolidación. El procesamiento *online* produce trazas en la memoria que pueden ser efímeras o durar por décadas. La memoria permite la emergencia de estructura en los niveles superiores a través de la generalización. Luego está el marco temporal de la difusión social. La consolidación de patrones lingüísticos depende de la interacción entre código neuronal y social. Los individuos consolidan formas que son usadas por la comunidad lingüística más amplia. El lenguaje sería desde este punto de vista como una colección de “memes” sociales internalizados por los miembros del grupo social. Dichos “memes” controlan interacciones sociales, como el cambio de código y el control de secuencias conversacionales. Por último aparece el tiempo a escala de

la difusión genética; que va de décadas a milenios. Todos los marcos temporales impactan en el momento del habla. Los patrones lingüísticos deben haberse consolidado en la memoria a largo plazo, permitiendo su acceso inmediato cuando las condiciones de uso del patrón lo requieran en el momento inmediato. Estas condiciones incluyen requerimientos sociales y afectivos desarrollados a través de relaciones interpersonales en la comunidad durante años.

El emergentismo adscribe a una visión “no cartesiana” de la cognición (Ellis, 2019), que está soportada por cuatro pilares. En primer lugar se trata de una cognición “encarnada” [*embodied*], modelada por aspectos corporales como el sistema perceptivo, las interacciones del cuerpo con el ambiente, y por expectativas integradas al organismo como resultado de la experiencia repetida. Es decir que la percepción y el sistema motor no constituyen aparatos periféricos. En segundo lugar, la mente se encuentra insertada en el ambiente [*Embeddedness*]. Por medio de capacidades sensomotoras se logran notar ciertos aspectos del ambiente que elicitan acciones de un organismo. El ambiente provee oportunidades para la acción [*affordances*]. Por ejemplo, una taza elicit que sea agarrada para tomar el café en su interior; las páginas web de los diarios ofrecen la oportunidad de acceder a noticias de actualidad. El individuo puede tomar dicha oportunidades de acción o dejarlas pasar. En tercer lugar, la cognición se asocia al enactivismo [*Enactivism*]: los procesos mentales no son solamente neuronales sino que también involucran rutinas de interacción del organismo con el ambiente. Los procesos cognitivos son dinámicos, adaptativos y sensomotores de variables en tiempo real. La cognición no es una maquinaria representacional que procesa información y está guiada por una sintaxis computacional. Por último, la mente está extendida [*extended mind*]; no se encuentra en el individuo “dentro de la cabeza” sino que es un sistema socio-técnico distribuido. Se considera que las tecnologías, redes sociales, instituciones, leyes, prácticas educativas y políticas sociales forman parte de órganos distribuidos de pensamiento. Por ejemplo, una agenda electrónica en el celular ayuda a compensar limitaciones de la memoria, extendiendo su capacidad.

Spivey (2007) invita a dejar la metáfora de la mente como si fuera una computadora que manipula símbolos discretos para pasar a una concepción en términos de un flujo continuo de procesos. Desde la perspectiva emergentista la gramática no es modular ni se encuentra encapsulada; sino que se adopta la visión conexionista de interrelación entre módulos permeables y emergentes. Se supone un input muy rico. Su riqueza permite el aprendizaje mediante mecanismos generales que abstraen propiedades estadísticas y distribucionales de dicho input. Los patrones del lenguaje surgen del uso, la generalización y la auto-organización. Las «reglas» de la gramática no se instancian como parámetros u operaciones con rasgos que preceden a la experiencia. Las «reglas» son epi-fenómenos, emergen de distribuciones de probabilidad que se estabilizan con el aprendizaje.

1.2. Hipótesis general y Objetivos.

El objetivo principal de la presente tesis es el análisis de los datos de campo recogidos sobre la adquisición de la concordancia de número plural en ELE (Español como lengua extranjera) en cuatro alumnos de lengua materna italiana. Se trató de cuatro estudiantes que tomaron un curso de niveles A1/A2, B1, B2 y C1 durante dos semestres en el *Instituto Cervantes* de Milán, en el año académico 2008/2009. La recogida de datos se realizó por medio de entrevistas de 30 minutos cada 15 días, de conversación no estructurada pero usando temas apropiados para el nivel. El resultado fue un total de entre 12 y 14 entrevistas. Ya que se pone el foco en comprender el cambio lingüístico en la adquisición de ELE como fenómeno complejo, no lineal y emergente; se utiliza como marco teórico general la teoría de los sistemas dinámicos (y en particular su aplicación a L2). El tema de análisis serán los errores de concordancia plural en el ámbito nominal. Se intentará relacionar varios factores influyentes en la comisión de dichos errores de concordancia mediante modelos estadísticos que tengan en cuenta la temporalidad y varias fuentes de variabilidad.

La hipótesis principal será que las instancias producidas “correctas” e “incorrectas” constituyen *atractores en un espacio de fase*. Se los considerará estados discretos emergentes de un proceso continuo subyacente de estados “intermedios” entre correcto / error. Se utilizarán las variables relacionadas con el error para diseñar mecanismos de sesgo hacia dichos atractores, que formarán parte de tres modelos de simulación basados en sistemas dinámicos. El objetivo será el de recrear los patrones globales de error para cada

alumno y el patrón de error a través de las sesiones.

Luego se examinará la dinámica temporal del error utilizando diferentes técnicas no lineales, con el propósito de evaluar regiones de cambio de dinámica; es decir, si los alumnos están pasando a una fase siguiente de aprendizaje. Asimismo se analizará la dinámica en redes complejas de concordancias. Además se utilizarán las variables de los modelos creados y aquellas derivadas de las redes (más otras) como atributos de clasificadores de minería de datos para intentar predecir localmente el error. También se aplicará un agrupamiento de las sesiones según dos niveles de intensidad de error y se examinarán las variables asociadas a la chance de error en cada grupo. Por último, se avanzarán algunos comentarios didácticos para la enseñanza de la concordancia según los resultados de la tesis y de la teoría de sistemas dinámicos aplicada a L2.

Los objetivos generales de la tesis son los siguientes:

- Comprender el cambio lingüístico en la adquisición de la concordancia plural español L2 en cuatro alumnos italianos como fenómeno complejo, no lineal y emergente.
- Realizar una contribución al estudio de la concordancia plural en español L2; investigando posibles factores relacionados con el error.
- Revisar críticamente los trabajos hechos sobre adquisición de la concordancia en L2.
- Extraer, a partir de los resultados, posibles sugerencias para la enseñanza de la concordancia en el aula.
- Analizar la dinámica temporal del error para detectar posibles cambios en el nivel de aprendizaje.
- Predecir el error de concordancia en las últimas sesiones de los individuos.
- Crear simulaciones con sistemas dinámicos que modelen la emergencia de la concordancia en L2 basándose en mecanismos de sesgo hacia atractores.

1.3. Resumen de los capítulos.

Se presenta a continuación una breve exposición del contenido de los capítulos de la tesis.

Capítulo 2. Se introduce el fenómeno de la concordancia desde la gramática descriptiva, la tipología, el minimalismo, la gramática sintagmática nuclear (HPSG), aquella léxico-funcional (LFG) y la gramática de las construcciones. Se la delimita al dominio nominal. Se revisa la bibliografía sobre psicolingüística de la concordancia en lenguas segundas. Se identifican en dicha bibliografía posibles parámetros que influyan en el aprendizaje de la concordancia en español como lengua extranjera. A saber: (i) etapas de adquisición; (ii) mayor facilidad del plural respecto al género; (iii) forma no marcada como default; (iv) facilidad de concordancia con artículos respecto de la de los adjetivos; (iv) morfología trasparente vs. no trasparente; (v) animicidad; (vi) distancia; (vii) similaridad tipológica entre lenguas; (viii) transfer.

Capítulo 3. Se introducen las nociones de «memoria» y «atención»; así como el modelo de producción bilingüe. Se resumen brevemente algunas teorías de la adquisición de lenguas segundas: gramática universal, funcionalismo, procesabilidad, adquisición de habilidades, declarativa / procedural, uso de las construcciones. Luego, se introduce la teoría de los sistemas dinámicos complejos (SDC) y se explican las nociones de «atractor», «repulsor» y «bifurcaciones». Se conectan los sistemas dinámicos con la enseñanza de ELE. Se revisa la bibliografía de SDC aplicada a la adquisición de L2. Se caracteriza la adquisición de lengua segunda como fenómeno no lineal, auto-organizado, emergente, sometido a la dinámica de atractores que se fortifican o debilitan. Se pone énfasis en cada trayectoria particular de los sujetos de aprendizaje. Se compara brevemente la teoría de los SDC con las introducidas previamente en aspectos como input, output, interlengua, instrucción formal, *transfer*. Se presenta una aplicación de los sistemas dinámicos a la concordancia en L1 y se propone otra aplicación para L2.

Capítulo 4. Se presentan los sujetos: SONIA, NATI, JAKO, MIRKA. Se detalla el procedimiento de recolección de datos, codificación y creación del Corpus. Se definen las variables (tipos de “respuesta” y “predictores”) y se explican métodos de pre-procesamiento para tratar problemas de colinealidad y datos faltantes; y para discretizar (y transformar) variables. Se delimitan las instancias de concordancia a estudiar. Se definen cuatro tipos de “respuesta”: (i) binaria [error / no error]; (ii) categórica [error de: género, -e-epentética, plural, mixto]; (iii) tiempo hasta que se produce un error; (iv) serie simbólica. Cada tipo de respuesta ameritará la aplicación de un método cuantitativo diferente en los capítulos 6 a 8 y 10. Se definen las predictoras que caracterizan las instancias de concordancia según: (i) tipo de modificador; (ii) si es a larga distancia; (iii) de más de dos términos; (iv) presencia de -e-epentética; (v) propiedades del controlador (concretud, familiaridad, imaginabilidad, frecuencia), (vi) similaridad entre las terminaciones del Español y el Italiano; (vii) similaridad entre las raíces léxicas de ambas lenguas; (viii) errores acumulados hasta la instancia en cuestión; (ix) posibles estrategias de aprendizaje (de 1 a 7), (x) frecuencia de TYPES de instancias de concordancia en corpus *EsTenTen* y en el propio.

Capítulo 5. Se aplica el *clustering* jerárquico para agrupar los casos de concordancia, definiéndose un predictor adicional con seis grupos. Se realizan gráficas descriptivas de las variables “respuesta” y “atributos”. Se utiliza un análisis de correspondencias múltiple para examinar la asociación entre las predictoras, sin tener en cuenta la respuesta. A continuación se lleva a cabo una selección preliminar de las variables predictoras más asociadas a la respuesta “error” y “tipo de error”. Para ello se utilizan medidas univariadas y basadas en modelos de *minería de datos*.

Capítulo 6. Se trata del primer capítulo con análisis estadístico. El objetivo consiste en identificar los predictores asociados a la respuesta binaria “comete o no error”. Se explica el modelo logístico (mixto) y se detalla la estrategia de selección de modelos a usar en este capítulo y los dos siguientes. Luego de llegar al modelo final, se analizan los errores de concordancia a la luz de los resultados.

Capítulo 7. En este segundo capítulo de análisis estadístico se presenta un modelo multinomial bayesiano para descubrir asociaciones entre las predictoras y cada uno de los cuatro tipos de error de concordancia definidos (respuesta categórica). Se analizan los errores de concordancia según tipo de error.

Capítulo 8. En el tercer capítulo de análisis estadístico se busca descubrir la posible asociación entre las predictoras con el tiempo que transcurre hasta que el alumno comete un error. Se introducen conceptos de análisis de “sobrevivencia” y se presenta un modelo de errores múltiples y otro de errores competitivos para cada tipo de error.

Capítulo 9. Se crean tres simulaciones basadas en sistemas dinámicos: (i) modelo competitivo de *Lokta-Volterra*; (ii) dinámica del replicador en un juego de coordinación según la teoría de los juegos evolutiva; (iii) modelo de ascenso por gradiente, en el marco de la *gradient symbolic computation*. Se utilizan los resultados de los capítulos 7 y 8 para crear un «contexto» / «ambiente» que sesga la activación de los modelos hacia alguno de los atractores «correcto» o «error». Se comparan resultados de patrones agregados de proporciones de error / acierto globales y por sesiones, para cada alumno, entre la simulación y los datos del corpus.

Capítulo 10. Se intentará cuantificar la pregunta general: ¿el alumno mejora en su aprendizaje? Desde el punto de vista de los sistemas complejos, esto implica detectar posibles transiciones entre regímenes de errores, sesiones de dinámica de error similar y aquellas con dinámica diferente. Se utilizan las siguientes técnicas: LDA [*Latent Dirichlet Allocation*], entropías, RQA [*Recurrence Quantification Analysis*], índices de cambio de régimen.

Capítulo 11. Se presenta la teoría de las redes complejas. Se la aplica a un análisis de cambio de la dinámica de la red a través de las sesiones, para cada aprendiente, testeando efectos estadísticos. Por otra parte, se aplican clasificadores para predecir el estatus de error de los enlaces en las últimas cuatro sesiones de cada aprendiente. Se practica un *clustering* por mixtura de distribuciones para agrupar las sesiones en dos niveles de intensidad de error. Se emplea un modelo mixto de regularización para seleccionar variables en cada grupo y se evalúa su relación con la respuesta binaria a través de un modelo logístico mixto.

Capítulo 12. Se introducen brevemente las nociones “Noticing Hypothesis”, “Focus on Form” y “Feedback”. Se sugiere una estrategia para corrección de errores de concordancia. Se cierra la presente tesis con un resumen de los hallazgos, discutiéndolos a la luz de la literatura; y se proponen algunos consejos para la enseñanza de la concordancia en el aula.

Capítulo 2

La concordancia en español.

En el presente capítulo se define el fenómeno de la concordancia y se lo encuadra tipológicamente. Se presentan explicaciones desde el punto de vista del minimalismo, gramáticas a unificación y gramática de las construcciones. Se comenta brevemente su adquisición en primera lengua. Luego se pasa a una exposición extendida del problema de la adquisición de la concordancia para español como lengua segunda.

2.1. Género y número en español e italiano.

La categoría gramatical del GÉNERO asigna sustantivos a dos (o más) clases y es una propiedad léxica inherente al nombre. Dicha asignación puede basarse en criterios semánticos, morfológicos o fonológicos (Corbett, 1991). En español algunos sustantivos animados poseen su género basado en alguna característica de su referente, como por ejemplo el sexo biológico (criterio semántico): (a) *el hombre / la mujer*; (b) *el gato / la gat-a*. En (a) la asignación a las clases “femenino” y “masculino” se expresa mediante bases léxicas diferentes; en cambio, en (b) la misma base léxica toma diferentes morfemas para expresarla; es decir que el criterio resulta también morfológico. En lo concerniente al criterio fonológico, a modo de ejemplo, las formas en *-i* y *-u* tienden a ser masculinas (*el espíritu, el taxi, pero: la tribu*). Además los segmentos *-ma* y *-ta* también determinan la pertenencia al masculino (*el problema, el planeta*). Formalmente, los nombres terminados en *-a* son en general femeninos (96.3%) y los terminados en *-o*, masculinos (99.8%)¹. Sin embargo existen formas menos prototípicas que asignan ambas clases; aquellas en *-e*: *el puente, la suerte*, y en consonante: *el camión, la canción*; así como excepciones infrecuentes del tipo: *la mano, la foto, el día, la radio*. Incluso hay casos donde el cambio de clase implica un cambio de significado: *el libro / la libra; el capital / la capital* (Ambadiang, 1999). Harris (1991) consideró que las terminaciones en *-a / -o*, representan marcadores de clase cuya realización fonética se produce en sincretismo con los marcadores de género; así por ejemplo, *[[perro]-jo]* reviste los rasgos: + *-o* marcador de clase / + masculino; mientras que *[[man]-jo]* conllevaría los rasgos: + *-o* marcador de clase / +femenino. Aun siendo la asignación de género léxica, ésta se manifiesta a nivel sintáctico a través de la concordancia con los artículos, adjetivos y pronombres. Es decir que la única pista realmente clara que tiene a disposición en el *input* un aprendiente es distribucional, por medio de relaciones de co-ocurrencia con modificadores que reflejan el género de modo “transparente”: *la man-o (está) suci-a*.

Si bien los nombres en español solo se asignan a una de las clases de género, (casi) todos ellos pueden exhibir la categoría gramatical de NÚMERO en sus rasgos «singular» o «plural». Formalmente, el plural de los nombres acabados en vocal no acentuada se forma añadiendo *-s* (*libros, barcos*); mientras que agregan *-es* aquellos acabados en consonante (*relojes, paredes*), a través de inserción fonológica de una *-e-* epentética. Los que finalizan en *-e* acentuada solo admiten plural en *-s* (*bebés, cafés*), mientras que las demás vocales acentuadas tienden a presentar dobles (*esquí / esquís; tabú / tabúes*). Por otro lado, los nombres terminados en vocal inacentuada seguida de */-s/* no varían (*lunes, tórax, crisis, hipótesis*). En general en los sustantivos concretos y contables, la oposición singular / plural indica cuantificación («uno» / «más de uno») de un *mismo* referente. En cambio en los nombres de sustancia, abstractos y colectivos, su plural-

¹Porcentajes citados en Alarcón (2011).

lización apunta en su mayoría a un referente distinto y/o conlleva un significado diferente: (a) contable / incontable: *algodón / algodones*; (b) abstracto / persona u evento: *autoridad / autoridades, locura / locuras*; (c) plural estilístico: *gente / gentes*. Asimismo los plurales que refieren a objetos complejos no son significativos (*tijera / tijeras*); mientras que en otros casos cambia completamente el referente al pluralizarse (*esposa / esposas*). Por último, existen algunas formas sin plural (*fe, salud*). A diferencia del género (inherente al nombre), el número es transcategorial, lo pueden manifestar sustantivos, verbos, adjetivos, artículos y pronombres. En general, adjetivos, artículos y pronombres forman el plural siguiendo las reglas morfológicas arriba descritas; a excepción de las formas invariables en los relativos: *lo que, lo cual, cuanto*; los interrogativos: *qué, cuánto*; y los indefinidos: *nada, ninguno/a, cada, más, menos, demás*. Todos los demostrativos, posesivos y artículos cuentan con formas de singular y plural, dejando de lado a las formas neutras *esto, eso, aquello, lo* (Ambadiang, 1999).

En italiano los nombres que terminan en *-o* suelen ser masculinos (*l'albero* ['el árbol']) y en *-a*, femeninos (*la stella* ['la estrella']). Sin embargo, al igual que en español, los terminados en *-ma* y *-ta* pertenecen al masculino (*il programma* ['el programa'], *il pirata* ['el pirata']) y, recuérdese también los de género común (*il / la cantante* ['el cantante' / 'la cantante']). Las formas terminadas en *-e* se asignan a ambas clases (*il leone* ['el león'], *la luce* ['la luz']). No obstante, los terminados en *-ie* e *-ione* suelen pertenecer a la clase de los femeninos (*la serie* ['la serie'], *la tensione* ['la tensión']). Por otra parte, los terminados en *-i* son generalmente femeninos (*l'analisi* ['el análisis'], *la crisi* ['la crisis']); al igual que aquellos que finalizan en *u* acentuada (*la tribù* ['la tribu'], *la virtù* ['la virtud']). Los nombres en consonante final son raros y de origen extranjero. A estos se les asigna la clase del masculino: *l'autobus* ['el colectivo' / 'el autobús'], *il bar* ['el bar'], *il film* ['el film', 'la película']. A los sustantivos masculinos en *-o* le corresponden femeninos en *-a* (*ragazzo / ragazza* ['chico' / 'chica'], *caballo / caballa* ['caballo' / 'yegua']). Al masculino en *-e* le pueden corresponder tres terminaciones de femenino: (i) en *-a*: *cameriere / cameriera* ['camarero' / 'camarera']; (ii) en *-essa*: *professore / professoressa* ['profesor' / 'profesora']; (iii) *-trice*: *attore / attrice* ['actor' / 'actriz']. En cuanto al adjetivo, merece la pena notar un par de diferencias con el español. Los adjetivos que finalizan en *-e* son invariantes: *il ragazzo inglese / la ragazza inglese* ['el chico inglés' / 'la chica inglesa']. Los adjetivos que en español terminan en *-e* son menos que en italiano, así: *allegro / allegra* ['alegre']; *arabo / araba* ['árabe']; *caldo / calda* ['caliente']; etc. El morfema de masculino plural de más frecuencia es *-i* (cualquiera sea su género): *il libro / i libri* ['el libro' / 'los libros']; *la mano / le mani* ['la mano' / 'las manos']; sin embargo, no siempre coincide con el masculino plural *'-os'* en español: *i turisti* ['los turistas']; *i tedeschi* ['los alemanes']. El morfema de femenino plural de más frecuencia es *'-e'* (con nombres femeninos en *-a*): *la casa / le case* ['la casa' / 'las casas']; sin embargo, no siempre coincide con el femenino plural *'-as'* en español: *le strade* ['las calles']; *le scarpe* [fem. pl.] / los zapatos [masc. pl.]. Los nombres terminados en *-e* siempre forman plural en *-i*: *il fiore / i fiori* ['la flor' / 'las flores']. Asimismo algunas palabras son invariantes: (i) nombres terminados en vocal acentuada: (*la / le*) *città* ['(la / las) ciudades']; (ii) monosílabos: *il re / i re* ['el rey' / 'los reyes']; (iii) nombres de origen extranjero, terminados generalmente en consonante: *il computer / i computer* ['el / los ordenador(es)'; 'la / las computadora(s)']. También resultan invariantes los terminados en *-o* (todos femeninos): (*la / le*) *radio* ['las radios']; (*la / le*) *foto* ['las fotos']. Los números de 200 a 900 carecen de género, como en español: *duecento case / monumenti* ['doscientas casas' / 'doscientos monumentos']. Los posesivos de 1ra. a 3ra. persona (singular / plural) tienen diferentes morfemas, marcando género y número, donde en español solamente se marca número: (1) *i miei libri - le mie case* ['mis libros' - 'mis casas']; (2) *i tuoi libri - le tue case* ['tus libros' - 'tus casas']; (3) *i suoi libri - le sue case* ['sus libros' - 'sus casas']. Algunas palabras tienen plural en *'-a'*: (*il*) *dito - (le)* *dita* ['el dedo' - 'los dedos']. Unos pocos nombres masculinos son invariables en *-a* también: *il cinema / i cinema* ['el / los cine(s)']; *il delta / i delta* ['el / los delta(s)'] (Carrera Díaz, 1989).

2.2. Lingüística de la concordancia.

2.2.1. Encuadre tipológico del fenómeno.

Defínase la concordancia como una relación entre rasgos sublexicales (pares 'valor: atributo') de los ítems léxicos (O' Grady, 2005). En español dichos rasgos son 'persona', 'número' y 'género', junto a sus

valores. Corbett (2006) denomina *controlador* al ítem léxico que determina la concordancia y *objetivo* al elemento cuya forma es determinada por aquel. Se denomina *dominio* al entorno sintáctico en el cual ocurre la concordancia. La concordancia se establece por covarianza sistemática de rasgos. Otros factores - como por ejemplo el orden de las palabras - pueden influir en la concordancia, aun sin estar representados como rasgos subléxicos; a estos se los llama *condiciones*. Dichos conceptos se caracterizan mediante criterios con respecto a un punto focal de comparación que se considera «canónico» o «prototípico» (lo cual no implica que sea más frecuente en las lenguas del mundo). Se resumen a continuación ejemplificando en español si los criterios se verifican en dicho idioma. La jerarquía marcada con « > » significa «más canónico que».

Controladores. En español: nombres y pronombres (personales, relativos, indefinidos, interrogativos).

- ausente > presente: el controlador (nominal o pronominal) canónicamente está presente (a) y es tácito (b).

(a) l-o-s [M-PL] niñ-o-s [M-PL] com-e-n [3P-PL]

(b) (ellos) cant-a-n [3P-PL].

- rasgos expresados > no expresados: en (a) el controlador (pronominal) marca canónicamente al menos tantas distinciones en la forma como su(s) objetivo(s); en (b), en cambio está subespecificado.

(a) ella / él está content-o / -a.

(b) yo estoy content-o / -a.

- controlador consistente > híbrido: en (a) el controlador (nominal) canónicamente determina los mismos rasgos en diferentes objetivos; en cambio en (b) el controlador toma diferentes concordancias (neutro en la frase nominal y masculino en la frase verbal).

(a) l-a-s [F-PL] ciudad-es [F-PL] fuero-n [3P.PL] destruid-a-s [F-PL]

(b) Búlgaro (Corbett, 2006: p. 12):

Negov-o Veličestv-o e došál.

su-N.SG majestad(N)-SG AUX.3SG llegar.PAS[M.SG]

'Su majestad ha llegado'.

- categoría gramatical irrelevante > relevante: en los dialectos del norte de Inglaterra, en el dominio del sintagma nominal sujeto, los pronombres (adyacentes al verbo) siguen la concordancia verbal del inglés estándar: *he sings / they sing* (*he, she, it + V-s; they, we, you + V*); en cambio si el núcleo es nominal o el pronombre no es adyacente el verbo lleva siempre *-s*: *birds sings* (*NP + V-s*); *They sing and dances* (Pietsch, 2005). Es decir que la categoría gramatical del controlador en un dominio específico es relevante para determinar la concordancia del objetivo.

Objetivos. En español: artículos, adjetivos, verbos, pronombres (demostrativos, posesivos, indefinidos).

- morfema ligado (afijo) > clítico > morfema libre: canónicamente el objetivo realiza la concordancia a través de afijos ligados a una base léxica, o sea por medio de morfología inflexional concatenativa, como en el ejemplo (a). En (b.1) el clítico «duplica» al objeto topicalizado y concuerda en género y número; en (b.2) el participio también concuerda. En (c) el clasificador genérico CLASS (para objetos que son armas) se repite en el mismo sintagma nominal constituyendo un ejemplo de concordancia entre morfemas libres.

(a) l-a-s [F-PL] cas-a-s [F-PL] blanc-a-s [F-PL]

(b.1) ¿l-o-s [M-PL] plat-o-s [M-PL], l-o-s [M-PL] compr-aste [2P.SG]?

(b.2) Italiano:

i piatt-i, l-i hai portat-i?

los-M.PL platos-M.PL, los-M.PL has-AUX.2P.SG traido-M.PL?

(c) Ngan'gityemerri (Corbett, 2006: p. 14)

(syiri) magulfu (syiri) marrgu.

CLASS palo.cilíndrico.de lucha CLASS nuevo

'un palo cilíndrico nuevo para luchar'

- obligatorio > opcional: en los ejemplos (a) y (b.1-2) anteriores los objetivos deben marcar obligatoriamente la concordancia; en cambio, en (c), en un caso menos canónico, el clasificador es optativo (por eso está entre paréntesis).
- morfología regular > supletiva: en el ejemplo siguiente el adjetivo usa dos raíces diferentes según concuerde en singular o en plural, un caso menos canónico de concordancia supletiva.

Noruego (Corbett, 2006: p. 15)

en lit-en bil / to småbil-er

un pequeñ-o-M.SG auto[SG] / dos pequeños.PL aut-os-PL

'un pequeño auto' / 'dos pequeños autos'

- morfología aliterativa > opaca: si la misma marca de concordancia se usa para diferentes objetivos hay aliteración completa, como en (a); en cambio en (b) es más opaca. En (a) '7' (el singular del género 7/8) se repite en los diferentes objetivos con la forma *-ki*; en cambio en (b), '1' (el singular del género 1/2) toma la forma *-a* como prefijo verbal y *-m* en el sintagma nominal. El español también observa grados de aliteración: *l-as[F-PL] man-os[F-PL] grand-es[F-PL] / l-as [F-PL] chic-as [F-PL] bonit-as [F-PL]*.

Swahili (Corbett, 2006: p. 15-16)

(a) ki-kapu ki-kubwa ki-moja ki-lianguka

SG-canasta (7/8) 7-grande 7-una 7-se cayó.

'se cayó una canasta grande'

(b) m-tu m-moja a-likuja

SG-persona(1/2) 1-una 1-vino

'vino una persona'

- marca de concordancia productiva > esporádica: cuántos más ítem léxicos o categorías manifiesten concordancia entonces más productiva será.
- objetivo con concordancia obligatoria siempre > solo si el controlador está ausente: la situación más canónica es cuando la concordancia se produce en el objetivo sin importar la presencia o ausencia del controlador, como sucede en español.
- objetivo concuerda con un solo controlador > con más de uno: en español se produce la situación canónica de objetivos concordando con un solo controlador. Si embargo en el ejemplo siguiente, el asociativo *-a-o* conlleva dos marcas de concordancia; una para cada nombre: *imbwa* (perro(s)) y *ana* (chico, joven). Obsérvese que éste último repite la marca de plural *-v* tanto en el nombre como en el asociativo; en cambio, el primer nombre no varía para número pero su interpretación es plural por la marca *-dz* en el asociativo.

Shona (Corbett, 2006: p. 18)

Imbwa na-v-ana v-a-dz-o

perro(s)(9/10) y-PL-joven(1/2) 2-ASOCIATIVO-10-ASOCIATIVO

'los perros y sus cachorros'

- objetivo sin elección de controlador > con elección: el ejemplo a continuación muestra la situación menos canónica en la cual el verbo matriz *-iy-* (saber) puede concordar en género (marcado con números romanos) tanto con el complemento en su conjunto (y en este caso lleva género IV, que es general) como con un nominal en caso absolutivo (concordando en el género de éste, aquí el III). O sea que un objetivo puede elegir controlador para la concordancia.

Tsez (Corbett, 2006: p. 18-19)

eni-r [už-ā magalu b-āc'-ru-li] r-iy-xo / b-iy-xo

madre(II)-DAT [chico(I)-ERG pan(III)[ABS] III-comer-PAS_PART-NMLZ]

IV-saber-PRES / III-saber-PRES

'la madre sabe que el chico se comió el pan'

- categoría gramatical irrelevante > relevante: ¿se comportan igual todos los objetivos sin importar su categoría gramatical? En el ejemplo siguiente se observa la situación menos canónica en la que los artículos definidos concuerdan siempre en género con su controlador nominal (a) pero los adjetivos solo lo hacen si están antepuestos a éste (en (a) pero no en (b)).

Michif (Corbett: 2006, p. 269)

(a) la gru-s tāt

ART.DEF.F:SG gran-F tienda(F)

'la gran tienda'

(b) la mǎzũ blā

ART.DEF.F:SG casa(F) blanca[M]

'la casa blanca'

Dominios. En español: sintagma nominal, sintagma verbal (predicativo), oración subordinada.

- relación asimétrica > simétrica: el controlador «determina» el rasgo del objetivo, o sea que éste último posee tantas marcas cuantos atributos posibles sean requeridos por el primero. Por ejemplo, en español la mayoría de los adjetivos tienen la marca *-o* y *-a* de género para el atributo «masculino» y «femenino» respectivamente, según requieran los sustantivos; sin embargo en el nombre las marcas pueden ser ambiguas o no existir: *mano blanca*, *camión blanco*. El objetivo se «adapta» a su controlador pero no al revés. La concordancia es pues una relación típicamente *asimétrica*.
- dominio local > no local: canónicamente la concordancia se da en un dominio local, por ejemplo dentro de un sintagma nominal (*[niños contentos]*SN); mas bien que cruzando a otro dominio; como en el caso de un predicativo concordando con el núcleo de un sintagma nominal en función de sujeto oracional (*[[los niños]SN [están contentos]SV*). La concordancia constituye una relación a distancia entre elementos no adyacentes (ya que aun estando un atributo «al lado» del nombre, entre las flexiones que concuerdan siempre existe material lingüístico, p. ej. las bases léxicas). Dicha distancia puede ser: (a) lineal: contando el número de elementos que intervienen entre el controlador y su objetivo (sílabas, palabras); (b) estructural: contando el número de nodos sintácticos que se cruzan entre controlador y objetivo (O'Grady et al., 2003). En el ejemplo siguiente, tanto en (a) como en (b) la distancia estructural es de un nodo, pero linealmente es de cuatro palabras en (a) y de ocho en (b).

(a) [La tienda de la esquina [está abierta]VP]CP

(b) [los zapatos de la tienda del tío de Pablo [son carísimos]VP]CP

- conjunto de dominios > dominio único: lo canónico es que haya concordancia en un conjunto de entornos sintácticos (atributo - núcleo nominal; antecedente-relativo; etc.), y no que se realice solo en un dominio único, por ejemplo, solamente entre el sujeto y el verbo.

Rasgos. En español: género: masculino, femenino; número: singular, plural; persona: 1ra., 2da., 3ra.

- rasgos léxicos > no léxicos: canónicamente los rasgos se consideran propiedades inherentes de un controlador (es decir, con los valores de sus rasgos listados en el léxico mental) y contextuales en sus objetivos. El género es, como se vio, claramente inherente al nombre, ya que las marcas de concordancia en sus objetivos se ajustan al valor dictado por aquel (son requeridos por el contexto). El número es ciertamente inherente en casos de nombres solamente plurales o singulares (*salud*, *dinero*, *esposas*). En lo que respecta a los nombres que poseen ambos valores de número Corbett (2006, p. 130) también lo considera léxico ya que el cambio en el significado en el nombre cuando se pluraliza afecta a éste pero no a sus modificadores. Por ejemplo en: *barco(s) bonito(s)* cambia la cantidad del referente pero no la propiedad que se le atribuye. Aunque en el verbo, el rasgo es contextual, existen casos de número verbal (Corbett (2006, p.131) cita un ejemplo en Georgiano). En cuanto al rasgo de persona, dicho rasgo es inherente a los pronombres personales y posesivos pero contextual en el verbo.
- valores de rasgos concuerdan > no concuerdan: canónicamente se esperaría que los valores de los rasgos sean los mismos en la concordancia. No obstante, los siguientes ejemplos muestran discordancia de valores en el número debido a razones semánticas o pragmáticas (Acuña-Fariña, 2009). En (a) el predicativo concuerda con el núcleo semántico y no con el sintáctico del sujeto; en (b) la construcción de «medidas» admite ambos valores en el verbo; (c) es un caso de resolución semántica de la construcción coordinada; (d) muestra discordancia en el valor del rasgo persona; por último en (e) el verbo toma concordancia plural aun habiendo un sujeto singular por razones de respeto.

(a) El [M-SG] fenómeno [M-SG] de tu hermana [F-SG] dice que está cansad-a [F-SG] / -o [M-SG] (?)

(b) tres huevos [PL] es [SG] mucho / son muchos

(c) Marta [F-SG] y su [3P.SG] hermana [F-SG] están [PL] contentas [F-PL]

(d) Los hombres [3P] somos [1P] unos desgraciados

(e) Ruso (Corbett, 2006: p. 3):

mamen'ka plač-ut

madre llorar-3P.PL

'tu madre está llorando'

Estos ejemplos demuestran que la concordancia no es únicamente una cuestión sintáctica. Corbett (2006, cap. 7) explica esta «interferencia» semántica en términos de una jerarquía de dominios, en la cual el movimiento hacia la derecha coincide con un aumento de probabilidad de injerencia de la semántica: atributo > predicado > pronombre relativo > pronombre personal. Lo ilustra con el uso de *committee* (comité) como sigue (p. 206-7):

(a) the committee / *these committee (atributo) 'el comité / * estos comités'

(b) the committee has decided / have decided (predicado) 'el comité ha / han decidido'

(c) The committee ... it / they (pronombre personal) 'el comité ... él / ellos'

En el dominio local atributivo (a) únicamente es posible la concordancia sintáctica en singular; sin embargo, una vez que dejamos dicho dominio, el plural es posible por interpretación colectiva del referente. Esta jerarquía demuestra, en primer lugar, que la concordancia es un asunto gradual; segundo, que el mismo controlador puede determinar valores diferentes de rasgos en objetivos diferentes al mismo tiempo; tercero, que la referencia pronominal puede verse como un fenómeno de concordancia (dado que los pronombres manifiestan los mismos rasgos que otros objetivos y están sujetos a los mismos mecanismos de resolución).

- sin elección del valor del rasgo > con elección: el ejemplo (2.11.b) demuestra que para sintagmas nominales que implican medidas (*dos, una docena, etc.*) ambos valores del número están a disposición en el verbo. En cambio en el ejemplo siguiente, a pesar de que *fiú* «chico» y el verbo *llegar* tienen formas de plural, son imposibles para el sintagma numeral.

Húngaro (Corbett, 2006: p. 25)

hat fiú érkez-ett

seis chico[SG] llegar-PAS[3.SG]

'llegaron seis chicos'

Condiciones: Si el controlador, el objetivo, el dominio y los rasgos han sido establecidos, lo canónico es que no se necesite condición alguna adicional para caracterizar una determinada construcción de concordancia. Si «algo más» determina la concordancia; esto se establece en términos de «condiciones». Dichas condiciones implican factores (tales como la animicidad o el orden de palabras) que no se realizan explícitamente como rasgos pero que pueden determinar de forma absoluta (presencia / ausencia) o relativa (preferencias estadísticas) la concordancia. Por ejemplo en *miya*, los controladores animados determinan concordancia en número en el objetivo; pero los inanimados no la exigen en absoluto. Es decir que aunque el controlador inanimado sea plural, su objetivo será singular, ya que la condición de animicidad prohíbe el plural.

Miya (Corbett, 2006: p. 177-8)

(a) níkin dzáfə

este.PL hombre.PL

'estos hombres'

(b) náκən víayúw-awáw

esto.M.SG chimenea(M)-PL

'estas chimeneas'

En cambio en *ruso*, con sintagmas coordinados en función de sujeto, la concordancia verbal puede ser plural (semántica) con ambos controladores o singular (sintáctica) con el controlador coordinado más cercano (aquí *brat* 'hermano'). Ambas posibilidades están a disposición del hablante; sin embargo, estadísticamente los hablantes prefieren el plural con controladores animados.

Ruso (Corbett, 2006: p. 179)

ogorči.1-i-s' / ogorči.1-i-sja brat i sestra

molestar-PAS-PL-REFL / molestar-PAS[M.SG]-REFL hermano y hermana

'hermano y hermana se molestaron'

En ambos ejemplos, el factor animicidad no es un rasgo manifiesto de concordancia; no obstante ello, condiciona su aparición (de modo absoluto o relativo).

A continuación se muestran ejemplos del fenómeno de concordancia bajo estudio en la presente tesis; caracterizado según los conceptos arriba expuestos de controlador, objetivo, dominio, rasgos y condiciones. Como se nota, la mayoría de los criterios son canónicos.

- L-os-MASC-PL libr-os-MASC-PL nuev-os / grand-es-MASC-PL;
- L-as-FEM-PL cas-as-FEM-PL nuev-as / grand-es-FEM-PL;
- L-os-MASC-PL libr-os-MASC-PL son nuev-os / grand-es-MSC-PL;
- L-as-FEM-PL cas-as-FEM-PL son nuev-as / grand-es-FEM-PL;

1. CONTROLADOR: Nominal. Presente, consistente en la concordancia que determina, expresa abiertamente los rasgos y su categoría gramatical es irrelevante.
2. OBJETIVO: artículos (definidos e indefinidos), adjetivos y pronombres (demostrativos, posesivos, indefinidos). Usa morfemas ligados (flexión) para expresar concordancia; con marcado obligatorio; usando morfología regular, productiva y con diferentes grados de aliteración. Concuerta con un solo controlador obligatoriamente sin importar si está presente o ausente. Además su categoría gramatical resulta irrelevante.
3. DOMINIO: sintagma nominal; sintagma verbal (predicativo); oración subordinada. Concordancia asimétrica (el género y número de los objetivos dependen del controlador nominal); dominio local y no local; y constituyen parte de un conjunto de dominios.
4. RASGOS: Género: valores «masculino» y «femenino»; Número: valor «plural». Rasgos léxicos en el controlador nominal; sus valores concuerdan (salvo en algún caso esporádico del tipo de (2.11.c)) y no hay posibilidad de elección.
5. CONDICIONES: sin condiciones.

En lo que atañe a la concordancia a larga distancia, se considerarán como tales aquellas del adjetivo con el controlador con verbos copulativos (*ser*, *estar*, *parecer*, *quedar*, etc.), también dentro de una oración subordinada relativa. El caso de elipsis en posición predicativa no será considerado como a larga distancia sino como local. Tampoco se considerará la concordancia en la voz pasiva con *ser* pero sí con *estar*. El Cuadro que sigue ilustra los casos.

EJEMPLO	¿LARGA DISTANCIA?
Los romanos <i>son</i> habladores / ...	SI
Los niños <i>están</i> despiertos / contentos / aburridos / dormidos / ...	SI
Las mesas (ya) <i>están</i> servidas / armadas ...	SI
Las cosas <i>que son</i> importantes	SI
Las costumbres son las mismas / otras [costumbres]	NO
Los libros interesantes <i>son</i> los [libros] rojos	NO
Esas teorías pueden <i>ser discutidas</i> más a fondo	NO (no incluido)

Figura 2.1: Casos de concordancia a larga distancia considerados.

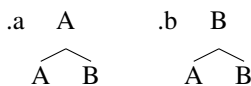
En el resto del capítulo la discusión se centrará en construcciones con estas características (dejando de lado, por ejemplo, la concordancia verbal).

2.2.2. Concordancia según el programa minimalista.

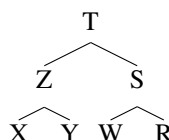
En el programa minimalista (Chomsky, 2000, 2001; ver también: Adger, 2003) se sostiene que la sintaxis manipula rasgos morfo-sintácticos abstractos. Estos rasgos a los que la sintaxis es sensitiva determinan aspectos de la forma y el significado de las palabras. Cada rasgo tiene valores asociados, por ejemplo: *Infl: past, gen: fem, N: +*. Los rasgos se dicen *interpretables* si generan un efecto en la interpretación semántica, de lo contrario son *no interpretables*. Además se postulan categorías léxicas y funcionales, a las cuales se asocian dichos rasgos. Tres operaciones manipulan categorías: *merge*, *move* y *agree*. La primera toma dos elementos/categorías *A* y *B*, y forma otro objeto, proyectando o bien *A* o bien *B*². Esta operación crea una relación de «hermandad» entre los elementos *A* y *B*. Además, los elementos a concatenar pueden ser otros objetos donde ya se ha aplicado *merge*. Sucesivas aplicaciones de *merge* van creando la estructura de frase. La operación *move* toma un elemento y lo desplaza para adjuntarlo al núcleo de un sintagma. Los valores de los rasgos se caracterizan también por ser «fuertes» o «débiles». Estos últimos desencadenan operaciones de movimiento. Por último, *agree* toma dos elementos del árbol y establece una relación de concordancia entre ellos. Los núcleos de los sintagmas funcionales contienen rasgos no interpretables cuyos valores deben ser chequeados y luego borrados antes de que la derivación se envíe a los módulos de interfaz. Un rasgo *no interpretable* *F* no especificado para su valor *v* que se aloja en un objeto sintáctico *X* se puede chequear via *agree* cuando *X* se encuentra en una relación de *c-comando*³ con otro objeto sintáctico *Z* que conlleva el mismo rasgo *interpretable* con su valor *F:v*. Cuando se produce el chequeo el rasgo no especificado en *F* recibe su valor correspondiente y es borrado. *X* se denomina «probe» y *Z* es el «goal». En suma, el rasgo interpretable y con un valor del «probe» chequea y borra el rasgo no interpretable/no valuado del «goal», valuándolo: $[uF:] \dots [iF:v] \Rightarrow [uF:v] \dots [iF:v]$ (donde «...» indica la relación de c-comando). Notar que se concibe la concordancia como una relación asimétrica entre rasgos no interpretables e interpretables, en la cual los primeros dependen de los segundos para especificar sus valores.

En lo que respecta al número, Ritter (1991, 1992) propuso el sintagma funcional de número (*Snum*), en el que son chequeados los valores $[\pm plural]$. En español el valor del rasgo de número es fuerte, lo cual hace que pueda desplazarse el rasgo léxico de *N* a la posición de especificador de *Snum* para chequear dicho

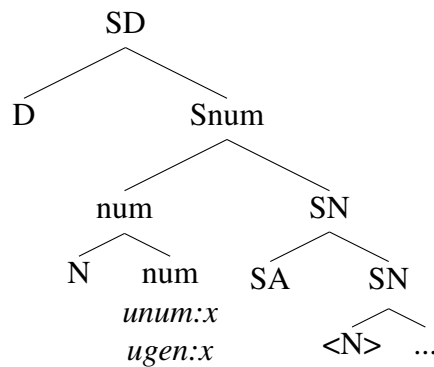
²O sea, la operación de *merge* entre las categorías *A* y *B* da lugar a cualquiera de los objetos en (a) o (b).



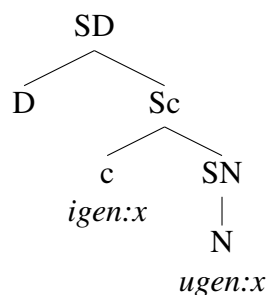
³Un nodo *A* tiene una relación de c-comando con otro nodo *B* si el «hermano» de *A* es o bien *B* o bien contiene a *B*. Así, en el siguiente árbol, *X* tiene una relación de c-comando con *Y* (su hermano) y con ningún otro nodo. En cambio, *Z* c-comanda a *S* (su hermano) y también a *W* y *R* (porque están contenidos en *S*).



valor⁴, como se muestra a continuación. Esto explica el orden de palabras en español respecto del adjetivo y el nombre: *una bonita casa / una casa bonita*. Si bien *num* es una categoría funcional, para Ritter (1993) el rasgo de género no lo es. Notar que aparece como rasgo no interpretable en el núcleo de *num*.



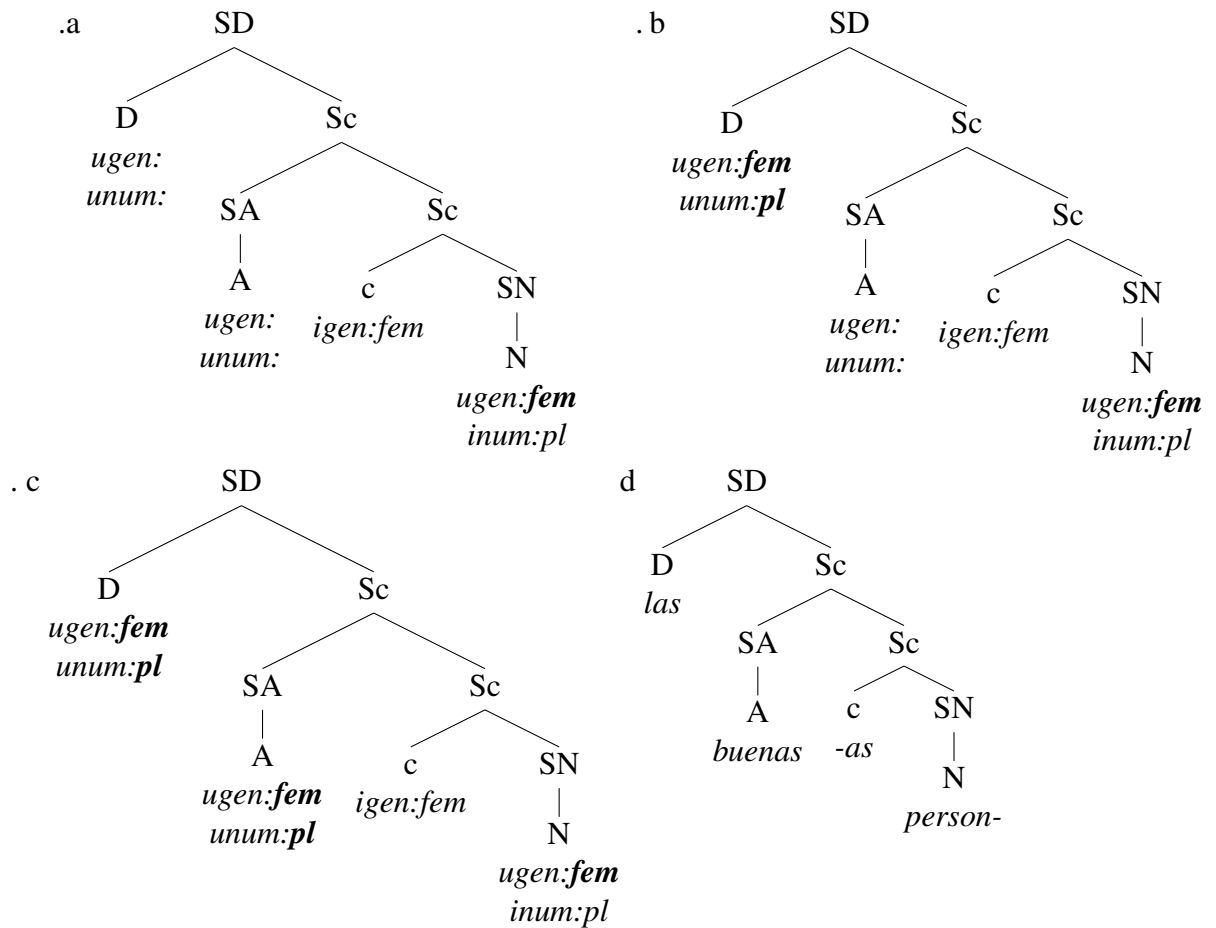
Harris (1991) advirtió la correlación existente en español entre marcador de clase y género gramatical (ver sección 2.1). Elaborando sobre ello, Picallo (2005) propuso una categoría funcional de marcador de clase de palabra, denominado *c* de «categorizador», como núcleo de la correspondiente proyección *Sc*, y que alojaría el rasgo de género. El núcleo de la categoría funcional conlleva la función de asignar una categoría a la entidad denotada por el nombre. Pero si el rasgo de género posee dicha función semántica entonces pasa a ser interpretable. El sintagma funcional *Sc* toma como complemento al *SN* y se aplica *agree* entre el rasgo *interpretable* de género alojado en el núcleo del sintagma y el rasgo de género *no interpretable* de *N*. Se aplica *merge* a la categoría funcional *c* y a (la proyección de) *N* antes de llevar a cabo otras aplicaciones de *merge*, por ejemplo, con adjetivos o determinantes. Es decir que la categorización tiene precedencia respecto de cualquier otro modificador.



Schoorlemmer (2009) adopta este enfoque pero restringe *agree* a la relación de dominancia⁵ entre la proyección de la categoría que aloja el rasgo no valuado y la categoría con el rasgo valuado. Considérese el ejemplo *las buenas personas*, cuya derivación se presenta a continuación. La categoría léxica *N* conlleva el rasgo no interpretable de género e interpretable de número. Su proyección se une a la categoría funcional *c* vía *merge* y se aplica *agree* al rasgo de género mediante el mecanismo propuesto por Picallo (2005). Luego el sintagma adjetival (*SA*) se adjunta a *Sc* cuyo núcleo léxico aloja los rasgos no especificados de género y número porque es «probe». Sin embargo, *SA* no entra en una relación de dominancia con *N*, y en consecuencia, *agree* no se puede aplicar. La derivación avanza aplicándose *merge* entre *Sc* y la categoría funcional de determinante. Dicha categoría también constituye un «probe» y conlleva rasgos no especificados para género y número [ver (a)]. Sin embargo, su proyección *SD* sí se halla en relación de dominancia (con el resto de los nodos). Primero se aplica *agree* entre *D* y *A*. Dado que los rasgos de ambos no están valuados, *agree* no «desactiva» a *D* como «probe». Entonces se sigue adelante con la búsqueda y se lleva a cabo *agree* entre *D* y *c* para obtener el valor de género y con *N* para que *D* reciba el de número [ver (b)]. En esta instancia, *D* y *A* comparten rasgos por haberse llevado a cabo *agree*. Como los rasgos en *D* se valoraron mediante *agree*, al compartir dichos rasgos, también *A* especifica los suyos vía concordancia indirecta [ver (c)]. Por último, se inserta el léxico y se lineariza para la salida fonológica [ver (d)].

⁴Dado que el número refiere al concepto de «pluralidad» se lo considera interpretable.

⁵Un nodo *N1* domina a otro nodo *N2* si existe un camino *siempre* descendente en el árbol entre *N1* y *N2*.



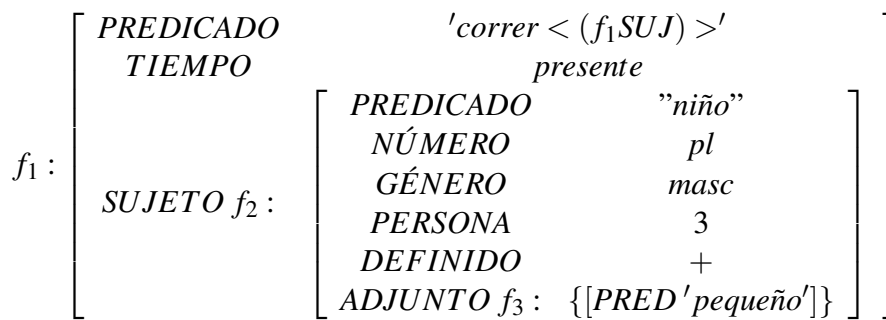
2.2.3. Concordancia según gramáticas con estructuras de rasgos: HPSG, LFG.

Esta sección trata brevemente el fenómeno de la concordancia en el ambiente nominal, según las denominadas *gramáticas a unificación de rasgos*. Se trata de gramáticas generativas no transformacionales; es decir que no postulan operaciones de movimientos. Formalmente, proponen un nivel de descripción en términos de estructuras de rasgos y mecanismos de «fusión» o unificación de dichos rasgos. En una palabra, los rasgos no se «mueven» sino que se «comparten».

2.2.3.1. Lexical Functional Grammar (LFG).

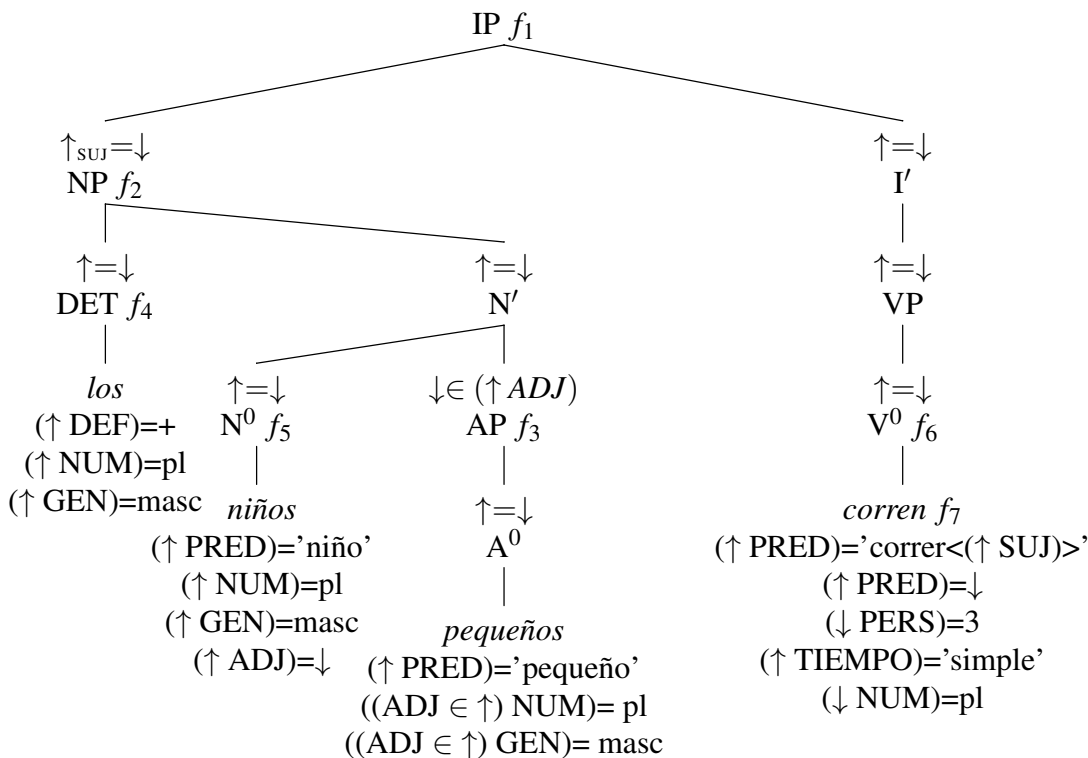
En primera instancia se presentará el análisis según la Gramática Léxico Funcional (Bresnan et al., 2016; Dalrymple, 2001; ver también: Tordera Yllescas, 2008). Se denomina «léxica» porque se considera que las posibilidades combinatorias de las palabras se hallan codificadas en el léxico. Así por ejemplo, **el mano* está mal formado porque en la entrada léxica de MANO se exige un determinante que sea femenino y singular. Asimismo, el atributo de «funcional» se debe a que no se supone que las funciones gramaticales (SUJETO, OBJETO, etc.) se deriven de la estructura sintagmática, sino que se las considera primitivos. También se relaciona con el hecho de que la buena formación de las *estructuras-f* se basa en funciones, en su sentido matemático. La teoría introduce varios niveles de representación y mecanismos de mapeo entre ellos. Se focalizará en los llamados: (i) «estructura-c»: que describe la estructura de constituyentes; y (ii) «estructura-f»: que representa las funciones gramaticales en términos de estructuras de rasgos⁶. Una estructura de rasgos es una matriz que incluye pares atributo-valor, donde los atributos deben conllevar como máximo un valor. La siguiente es la estructura-f de la oración *los niños pequeños corren*.

⁶Además se postulan otros niveles para dar cuenta de: los roles argumentales de los predicados [estructura-a]; aspectos del discurso [estructura-i]; semántica [estructura-s], prosodia [estructura-p].



La estructura está bien formada porque a cada atributo le corresponde solamente un valor. Los valores pueden ser simbólicos como «pl» o «+» o bien formas semánticas, como «niño». Además, los valores pueden estar constituidos por otra estructura de rasgos. En el ejemplo, la estructura f_1 contiene el rasgo *TIEMPO* con el valor *presente* y el atributo *SUJETO* posee como valor la estructura de rasgos f_2 , la cual aloja, por ejemplo, el atributo *NÚMERO* de valor *pl* y el atributo *ADJUNTO*⁷, cuyo valor es la estructura de rasgos f_3 . A su vez, la estructura-f f_3 contiene un atributo cuyo valor es la forma semántica «pequeño». Ahora bien, las estructuras-f son funciones y como tales se pueden representar mediante «ecuaciones funcionales» del tipo: $(f_i a) = v$; donde a es un atributo y v un valor de la función i . Por ejemplo: $(f_2 \text{NÚMERO}) = pl$ indica que la función / estructura-f asigna al atributo de número el valor «plural». Por otra parte, en la ecuación funcional $(f_1 \text{NÚMERO}) = f_2$ se denota que la estructura de rasgos f_1 asigna como valor del atributo sujeto a otra estructura de rasgos f_2 . Como se apuntó antes, la teoría es de corte lexicalista, con lo cual el léxico se halla anotado con esquemas funcionales. Por ejemplo a «niño» se asociarán las ecuaciones: $niñ- : (f_i \text{PRED}) = "niño"$; $-o : (f_i \text{GEN}) = masc$; $-s : (f_i \text{NUM}) = pl$.

La «estructura-c» describe la estructura de constituyentes. Se halla anotada tanto léxica como sintagmáticamente mediante ecuaciones funcionales que permiten relacionar cada nodo con una estructura funcional determinada. Dicho mapeo no es necesariamente «uno a uno». La estructura-c de la oración *los niños pequeños corren* es la que se muestra a continuación.



Las flecha «↑» indica que la estructura-f correspondiente es la asociada al nodo superior; «↓» indica que la estructura-f asociada es la del nodo corriente. Reemplazando las flechas por las estructuras-f a las que se refieren, se obtienen las siguientes ecuaciones funcionales: (a) $(f_1 \text{SUB}) = f_2$; (b) $f_2 = f_4$; (c) $f_2 = f_5$; (d) $(f_4 \text{DEF}) = +$; (e) $(f_4 \text{NUM}) = pl$; (f) $(f_4 \text{GEN}) = masc$; (g) $(f_5 \text{PRED}) = 'niño'$; (h) $(f_5 \text{NUM}) = pl$;

⁷En LFG, la función gramatical de *ADJUNTO* subsume a los atributos. Notar que se trata de una lista (indicada por {...}) de solo un elemento.

($f_5 GEN$) = *masc*; (i) ($f_3 PRED$) = 'pequeño'; (j) (($ADJ \in f_5$) NUM) = *pl*; (k) (($ADJ \in f_5$) GEN) = *masc*; (l) ($f_6 PRED$) = 'correr < (\uparrow SUJ) >'; (m) ($f_6 TIEMPO$) = *presente*; (n) ($f_6 SUB$) = f_7 ; (ñ) ($f_7 NUM$) = *pl*; (o) ($f_7 PERS$) = 3; (p) $f_6 = f_7$; (q) $f_1 = f_6$. Las ecuaciones proyectan las siguientes estructuras-f:

$$(c),(d),(e) \Rightarrow f_4 : \left[\begin{array}{cc} DEF & + \\ NUM & pl \\ GEN & masc \end{array} \right]; (f),(g),(h) \Rightarrow f_5 : \left[\begin{array}{cc} PRED & 'niño' \\ NUM & pl \\ GEN & masc \\ ADJ & f_3 \end{array} \right]; (i),(j),(k) \Rightarrow f_3 : \left[PRED \quad 'pequeño' \right];$$

$$(l),(m),(n) \Rightarrow f_6 : \left[\begin{array}{cc} PRED & 'correr < (\uparrow SUJ) >' \\ SUJ & f_7 \\ TIEMPO & presente \end{array} \right]; (\tilde{n}),(o) \Rightarrow f_7 : \left[\begin{array}{cc} PERS & 3 \\ NUM & pl \end{array} \right]$$

La concordancia de los atributos con el nombre utiliza una ecuación funcional de «inside-out» del tipo ($ADJ \in \uparrow$) *atributo* = *valor*, e indica que el valor del atributo del adjunto coincide con el valor del mismo atributo de la estructura-f al cual pertenece el adjunto. Por ejemplo el valor de NÚMERO del adjunto, es el mismo que el valor de número que está en la estructura f_5 , a saber, «plural». Por lo tanto, se co-especifica el valor de un rasgo del atributo con el valor del mismo rasgo en el sintagma nominal. Es una relación entre un elemento y el sintagma nominal que lo contiene. A diferencia del enfoque minimalista, no hay relación asimétrica (Börjars et al., 2019). Por medio de las funciones (a) y (b) las estructuras f_4 y f_5 se unifican en la estructura-f siguiente. Notar que el mecanismo de unificación toma dos estructuras-f y «pega» los atributos que se repiten. Tanto por medio de la función «inside-out» como por la unificación, la concordancia constituye esencialmente un intercambio de información en el seno del sintagma nominal.

$$f_2, f_4, f_5 : \left[\begin{array}{cc} PRED & 'niño' \\ DEF & + \\ NUM & pl \\ GEN & masc \\ ADJ & f_3 : \left[PRED \quad 'pequeño' \right] \end{array} \right]$$

Por otra parte, las funciones (p) y (q) producen la estructura-f:

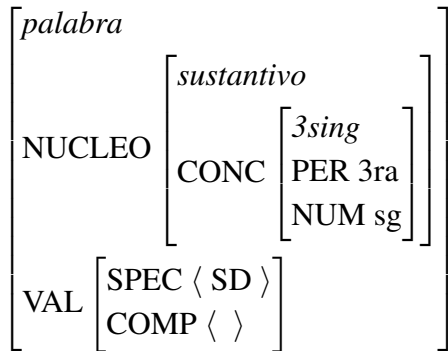
$$f_1, f_6 : \left[\begin{array}{cc} PREDICADO & 'correr < (f_1 SUJ) >' \\ TIEMPO & presente \\ SUJETO & f_7 : \left[\begin{array}{cc} NÚMERO & pl \\ PERSONA & 3 \end{array} \right] \end{array} \right]$$

Por último, la función (a) establece que el sujeto de $f_1 (= f_6)$ es la estructura-f $f_2 (= f_4 = f_5)$. Con lo cual, se obtiene la estructura-f f_1 inicial.

2.2.3.2. Head-driven Phrase Structure Grammar (HPSG).

Aquí se presenta el fenómeno desde el punto de vista de la llamada «gramática sintagmática nuclear» [Pollard & Sag, 1994; Sag, Wasow & Bender, 2003; Kim & Sells, 2008]. Al igual que el enfoque anterior, se trata de una teoría no derivacional y de corte lexicalista. En este enfoque, sin embargo, las estructuras de rasgos se encuentran tipificadas. La unidad básica de representación es el *signo*. Éste se describe como un complejo de tipos, los cuales conllevan información fonológica, sintáctica, semántica y discursiva. Los tipos representan entidades lingüísticas de la teoría y están estructurados con rasgos (pares atributo - valor), que describen sus propiedades y restricciones combinatorias. Los valores de los rasgos pueden ser ellos mismos otros tipos. De este modo se puede establecer qué rasgos se asocian a cuáles tipos y cuáles no. Los dos tipos principales son «palabra» y «sintagma». Tómese a modo de ejemplo, la palabra «perro», cuya representación se muestra a continuación. Es del tipo *palabra* y aloja los atributos NÚCLEO y VAL(encia). El primer atributo pertenece al tipo *sustantivo* y posee el atributo CONC(ordancia) que es en sí mismo un atributo complejo, de tipo *3sing*, y cuyos atributos son PERS(ona), de valor «tercera»; y NÚM(ero), con valor «singular». Asimismo, el atributo VAL(encia), aloja el atributo SPEC [especificador] que debe ser

uno de la lista de valores encerrada en <>, en este caso «SD» (sintagma determinante). Por otro lado, el atributo COMP(lemento) no lleva valores.

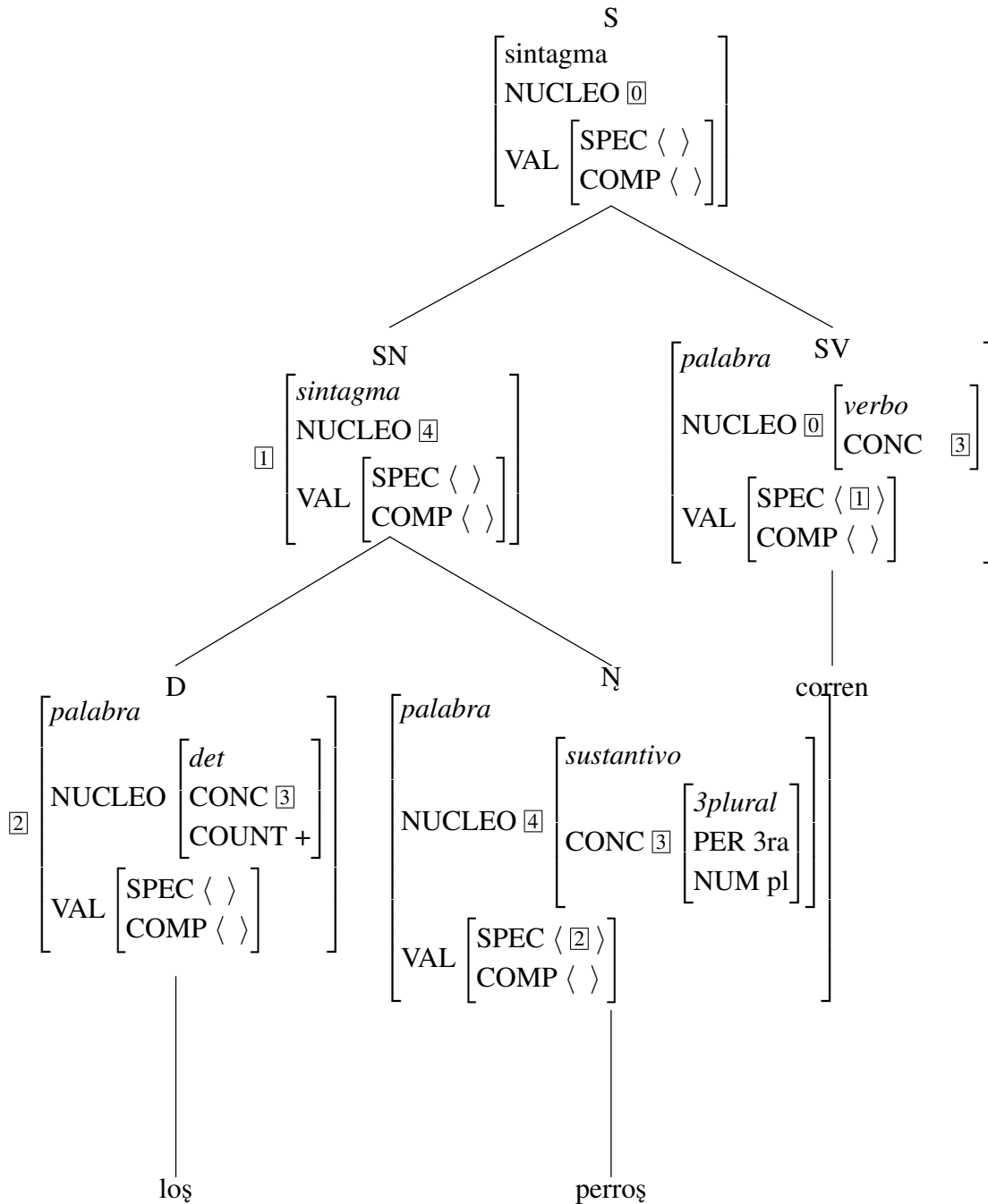


La teoría propone una serie de principios universales. Los siguientes son los pertinentes al fenómeno aquí descrito.

- Principio de atributo del núcleo [«Head Feature Principle»; HFP]: en la estructura sintagmática, el valor del núcleo de la estructura de rasgos «madre» debe coincidir con el valor del núcleo de la estructura de rasgos del «hijo».
- Principio de valencia [«Valence Principle», VP]: A excepción de que una regla diga otra cosa, los valores de los atributos SPEC y COMP de en la «madre» deben coincidir con los del núcleo del «hijo».
- Restricción de concordancia Núcleo-Especificador [«Specifier-Head Agreement Constraint», SHAC]: Los valores de CONC del núcleo deben coincidir con los valores de CONC de su especificador. Vale para verbos y sustantivos comunes⁸.

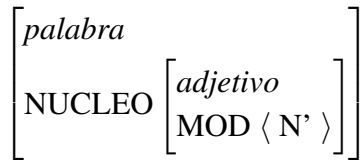
Pollard & Sag (1994, cap. 2) definen la concordancia como una variación sistemática en la forma derivada del hecho de que la información proveniente de dos elementos sobre un tercero debe ser compatible. Al igual que en el enfoque precedente, la concordancia constituye una relación simétrica de compatibilidad de información entre los valores de dos estructuras de rasgos diferentes, como se establece en el principio SHAC. La siguiente representación ilustra la oración *los perros corren*.

⁸Es decir que los verbos y los sustantivos comunes deben especificarse como: $\left[\begin{array}{l} \text{HEAD} \\ \text{VAL} \left[\text{SPEC} \left[\begin{array}{l} \text{AGR} \boxtimes \\ \langle \text{AGR} \boxtimes \rangle \end{array} \right] \right] \end{array} \right]$; cuyos valores en \boxtimes sean idénticos.

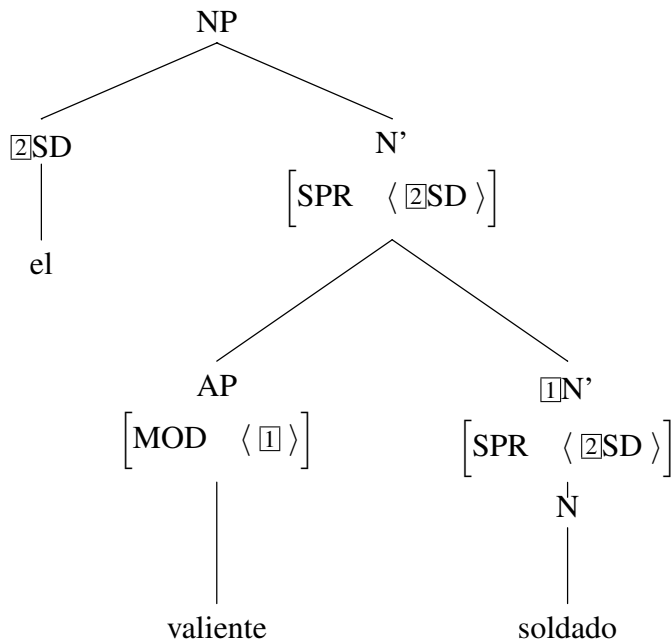


En el árbol, es necesario notar los siguientes aspectos. En primer lugar el valor de NÚCLEO del sustantivo *perros* ⁽⁴⁾ coincide con el del sintagma «madre». Lo mismo sucede en el caso del valor del NÚCLEO del verbo *corren* ⁽⁰⁾ y el sintagma «madre» superior. Esto satisface el principio HFP. Por otra parte, el valor de CONC del verbo ⁽³⁾ es idéntico al de CONC del SN, que está seleccionado por el verbo como su especificador ⁽¹⁾. Análogamente, el valor de CONC del N ⁽³⁾ es idéntico al de CONC del D, que es seleccionado por el N como su especificador ⁽²⁾. Estos dos hechos satisfacen el principio SHAC. Obsérvese que el atributo CONC en el N constituye, a su vez, un atributo de NÚCLEO ⁽⁴⁾. Por interacción de ambos principios, HFP y SHAC, se sigue que los valores de CONC de N, D y SN son idénticos. O sea que cuando el verbo selecciona un determinado sujeto NP como especificador (con ciertos valores en CONC), dicha selección restringirá la clase de N y, a través de las restricciones de selección del N, la clase de determinante que pueden co-ocurrir como sujetos. En suma, la concordancia se produce por la satisfacción mutua de las restricciones de selección.

En lo referente a los adjetivos en función atributiva, éstos conllevan el atributo MOD, cuyo valor es del tipo *sustantivo*. Por ejemplo la entrada léxica de *valiente* sería del estilo:



Para el sintagma *el valiente soldado* se obtendría un árbol como el siguiente :



2.2.4. Concordancia según la gramática de las construcciones.

Una construcción es la convencionalización de un apareamiento forma-significado, en varios niveles de complejidad y abstracción (Goldberg, 2013; Diessel, 2015). Por ejemplo, los siguientes constituyen ejemplos de construcciones: (i) palabra: *Casa* ; (ii) palabra (parcialmente completa): *pre-V* (*pre-calentar*); (iii) modismo: *ahogarse en un vaso de agua* ; (iv) modismo (parcialmente completo): *más A que NP* (más rápido que un bombero); (v) imperativa: *V + NP* (*¡cerrá la puerta!*); (vi) fórmulas: *no (lo) sé; ¡qué tal!*. Obsérvese que las unidades léxicas son construcciones, con lo cual no se postula diferencia alguna entre léxico y gramática. Al igual de los enfoques de la sección anterior, se trata de una gramática no derivacional; es decir, que no hace uso de operaciones de movimiento. Las construcciones son representaciones esquemáticas, con un polo del esquema para la forma y otro para el significado. Estos esquemas se hallan instanciados en expresiones léxicas concretas. Las construcciones *emergen* por un proceso de aprendizaje implícito en el cual se abstraen propiedades estadísticas y distribucionales de ejemplos concretos que conllevan forma y significado similares. Por ejemplo, en la construcción de «causa de movimiento de un objeto», un «agente» causa el movimiento de un «objeto» que recorre un «sendero» para llegar a «destino». Este «esquema de significado» se podría abstraer a partir de posibles instancias como: (*Marcos*) *le envió el libro a María por avión*; (*Marcos*) *envió el libro a Córdoba*; (*Marcos*) *tiró el libro por la ventana*; (*Marcos*) *tiró el libro al otro lado de la habitación*. Las construcciones forman parte de una red asociativa en donde las construcciones «hijas» heredan propiedades de construcciones «madres», que poseen mayor nivel de abstracción. Por ejemplo la construcción [*P + NP*], que conlleva una actividad asociada al *NP* (*¡a la mesa!* [a comer], *¡a la cama!* [a dormir], *¡a la escuela!* [a estudiar]) es «hija» de la construcción más general de sintagma preposicional y hereda de ella el orden de sus elementos (que haya pre-posiciones y no post-posiciones). Por otra parte, una misma construcción puede variar cros-lingüísticamente en forma, función y distribución. Tómese por ejemplo la «construcción pasiva», que en las diferentes lenguas puede tener diferentes propiedades como: presencia o elección de verbo auxiliar, caso o adposición; restricciones semánticas o discursivas; frecuencia de uso. Sin embargo, dicha variabilidad no se explica mediante principios universales innatos propios de la facultad del lenguaje, sino a través de mecanismos generales de la cognición, como por ejemplo, «fuerzas» funcionales (comunicativas) y restricciones de procesamiento y de aprendizaje. Vale la pena aclarar que los roles y categorías sintácticas (verbo, adjetivo, sujeto, etc.)

no constituyen primitivos de la teoría, sino que son propiedades que *emergen* con el uso (y la evolución diacrónica) de las construcciones. Al igual que las construcciones son generalizaciones a partir de enunciados concretos; las categorías y relaciones gramaticales se entienden como generalizaciones sobre el uso recurrente de partes de las construcciones (Croft, 2001). Asimismo, la relación entre las construcciones y los lexemas que pueden instanciarlas es semántica pero también probabilística. Hay lexemas más frecuentes que otros para una construcción en particular, que lo que se esperaría por azar. Esto sucede aunque dichos lexemas sean compatibles con el esquema de significado de la construcción en cuestión. También la estructura de constituyentes se entiende como un fenómeno emergente. Los sintagmas emergen a partir de dos factores. En primer lugar, el hecho de que elementos que son coherentes semánticamente tienden a colocarse cerca. En segundo lugar, a partir de probabilidades condicionales entre los elementos. Si dos o más elementos (léxicos o categorías abstractas) se usan de modo frecuente en un determinado orden, la fuerza de los enlaces que se establecen entre dichos elementos va aumentando con el repetido uso. Dicha secuencia de elementos se convierte así en una «unidad de procesamiento», con alta cohesión sintáctica. Ya que las unidades de procesamiento más pequeñas tienden a tener más frecuencia que las más grandes, se conforma un jerarquía de unidades que podría ser representada en términos de «árboles». Por ejemplo, en la secuencia $V : cuento \rightarrow DET : una \rightarrow N : historia$, la frecuencia de uso es más alta entre $[DET \rightarrow N]$ que entre el verbo y el resto, con lo cual emerge la estructura $[V : cuento \rightarrow [DET : una \rightarrow N : historia]]$. En cambio en $V : cuento \rightarrow P : con \rightarrow DET : tu \rightarrow N : ayuda$. El enlace es más frecuente entre el verbo y la preposición, emergiendo: $[[V : cuento \rightarrow P : con] \rightarrow [DET : tu \rightarrow N : ayuda]]^9$.

Acuña-Fariña (2018) propuso tratar a la concordancia como una construcción y enumeró una serie de principios cognitivos que compartiría con el resto de las construcciones. A saber:

categorización. Las lenguas pueden dividirse en aquellas que están orientadas al orden de palabras fijo con poca o nula concordancia y aquellas de orden de palabras flexible con presencia fuerte de concordancia. Tanto el orden de palabras fijo como la concordancia constituyen «marcas» que ayudan a identificar sintagmas. Se afirma que la concordancia se apoya en un mecanismo de co-categorización / co-clasificación. Entonces, en español los nombres pertenecerían a la clase o categoría «masculino», «femenino» y se consideraría al género como una marca referencial de dicha clase (sea nocional como en «niño» o gramatical como en «casa»). Dicha marca de clase aparecería replicada en los satélites modificadores del nombre, que co-clasificarían/categorizarían las palabras que forman parte del sintagma nominal. Por tanto, la concordancia en español es en esencia un fenómeno nominal y referencial.

extensión metafórica. Entiéndase la metáfora como un traslado de un dominio conceptual a otro. En lo que respecta al género, operaría un traslado de la distinción biológica «macho» / «hembra» [*niño* / *niña*] a dominios conceptuales no biológicos [*mesa* / *barco*]. Al principio se establecería una relación entre la marca de género en el nombre y la noción de sexo biológico, en el caso de controladores «prototípicos». El repetido uso de las marcas en estos nombres prototípicos y en sus modificadores (*esos niños traviesos*) crearía esquemas del tipo [*os - os -os*]. Luego dichos esquemas se extenderían a otros dominios no biológicos (*esos autos rojos*), para nombres «no prototípicos». Por alta frecuencia de uso, el esquema se haría completamente automático, y se aplicaría a cualquier tipo de categorización de nuevas palabras (*casa* como «femenino»).

consolidación por el uso («entrenchment»). El uso repetido del esquema de concordancia lo convierte en automático e incluso «ciego» a la semántica. Entra en acto una expectativa probabilística: si se vio un artículo (*los*) y un nombre (*soldados*) terminados en *-os*, el siguiente adjetivo deberá terminar en *-os* (*altos* / **-as*), según el esquema. Este efecto de «forma» debería aparecer aunque los nombres tengan género gramatical o nocional. Es decir que la construcción se aplicaría sin tener en cuenta siquiera la parte del significado. El esquema se vuelve un *gestalt*, un objeto perceptivo cuya función es unificar los rasgos de los elementos de un sintagma, indiferente a la semántica. Resulta interesante notar que los efectos de interferencia semántica en español (concordancia *ad sensum*), que es una lengua orientada a la concordancia

⁹Nótese que esta última no podría ser generada con reglas de estructura sintagmática.

y al orden de palabras flexible, son mucho menores que en inglés, una lengua orientada al orden de palabras rígido¹⁰. También se ha visto que las palabras con marcas de morfología transparente (por ej., palabras de género en *-o/-a*) elicitán menos errores de concordancia que aquellas de marcas no transparentes (palabras en *-e*, que pueden ser masculinas o femeninas).

metonimia y uso idiosincrático. El español codifica rasgos de género y número en varias partes del sintagma nominal que son adyacentes entre sí. La alta frecuencia de uso del esquema de concordancia permite eliminar uno de los elementos adyacentes y recuperar aun así los rasgos de todo el sintagma nominal. Se habla, por supuesto, del fenómeno de elipsis (*quiero las [remeras] blancas*) donde opera un mecanismo metonímico: la parte (*las blancas*) está por el todo (*las remeras blancas*). Por otro lado, se sabe que en la periferia de uso regular de una construcción, se hallan casos de usos no regulares o idiosincráticos, como en *el niño y la niña juegan*, donde el verbo es plural pero los controladores son singulares. En este caso sí opera la concordancia por referencia semántica a un sujeto colectivo.

En suma, la concordancia se encuentra regida por varios principios cognitivos: categorización (que lleva a marcadores de clase de palabra); extensión metafórica (que amplía el uso de las marcas de género hacia casos no prototípicos); metonimia (que lleva a la elipsis); formación de *gestalt* (que lleva a efectos de «forma», sin tener en cuenta el significado), usos idiosincráticos (que llevan a la interpretación semántica).

2.3. Adquisición de la concordancia (L1).

Serra et al. (2000) comentan la adquisición, en español L1, del sintagma nominal [*DET + N*]. Al principio, algunos niños producen una especie de proto-artículo usando una vocal delante del nombre (a menudo son */a/, /e/, /o/*; */u/* es rara; *e /i/* no se observa en los datos); por ejemplo: *e pie; a pie; a coche*. En otros niños hay directamente omisiones sin pasar por dicha etapa de proto-artículos. De cualquier modo, el periodo de omisiones y/o proto-artículos culmina hacia los 30 meses de edad, señalando que la posición del determinante ya está ocupada. Paralelamente aparecen los morfemas de género femenino y masculino (con variabilidad inter-individual respecto del orden de emergencia) y el número singular. Los errores de concordancia aparecen con la emergencia del género y se prolongan hasta los 33 meses aproximadamente. Luego de los dos años los niños comienzan también a sobre-regularizar. El morfema de plural emerge después de los de género pero se adquiere antes: los errores de concordancia plural tienden a desaparecer antes que los de concordancia de género. Asimismo, el plural en *-s* se adquiere antes que el plural en *-es*. A partir de los dos años, aproximadamente, aparecen los artículos determinados e indeterminados (orden de emergencia variable por individuo). Tras la adquisición de éstos, surgen los demás determinantes (indefinidos, demostrativos, posesivos, cuantificadores, etc.). En general se observa que para la edad de tres años los errores caen debajo del 10%. En suma, se verifica una adquisición gradual del determinante. Dado que el nombre co-ocurre con varios tipos de determinante, los niños comienzan con la parte fácil: la posición. Luego se concentran en los más frecuentes: los artículos (determinados e indeterminados); y en dominar su concordancia con el nombre. Solamente después siguen con el resto. El Cuadro que sigue ilustra lo expuesto.

¹⁰Lorimor, Bock, Zalkind, Sheymean & Beard (2008) llevaron a cabo un estudio de meta-análisis de las investigaciones psicolingüísticas sobre atracción de la concordancia. Por ejemplo, en: *la mayoría (N1) de los manifestantes (N1) fue / fueron*, la concordancia del verbo puede estar atraída al N2 (*ad sensum*) en lugar del N1 (por forma gramatical). Encontraron que los efectos de atracción se hacían cada vez menos fuertes cuánto más recursos tenía la lengua para expresar concordancias. La lengua más porosa a efectos semánticos de interferencia fue el Inglés, seguida del Holandés, Alemán, las lenguas romances (español, Italiano, Francés) y el Ruso. Siendo esta última la lengua de morfología más rica y en la que se verificó menos efecto. Se nota cómo el efecto depende de un *clíné* de «fuerza morfológica».

	18 meses	24 meses	30 meses	36 meses
determinante	(proto-artículos)	artículos	det./indet.	otros determinates
posición	omisiones		adquisición de la	posición
género		Femenino/ Masculino		
		Errores de concordancia de género		
			sobre-regularizaciones	
número		singular	plural	
			Errores de concordancia de número	

Cuadro 2.1: Esquema de adquisición del sintagma nominal [*DET* + *N*] según Serra et al. (2000)

Según el estudio longitudinal de Marrero & Aguirre (2003), el plural en el ambiente nominal aparece antes que en el ambiente verbal, con evolución gradual. Proponen tres etapas de adquisición del plural. En la primera etapa o no se observan marcas de plural, o se observan plurales inanalizados (o sea, sin el contraste singular / plural). En la segunda etapa, emerge la oposición singular / plural y se la marca o bien en el artículo (*e(l) / (l)o(s)*): *e huevo / o huevo*; o bien en el nombre ($\emptyset / -e(s)$ o $\emptyset / -s$): *pez / pece* [a menudo la marca de plural es *-e* porque se omite la *-s*], *a mano / a manos*. En la tercera etapa el niño tiende a producir más de una marca en el sintagma nominal (en el nombre y en el determinante). Aquí es donde aparece la concordancia y los artículos se comienzan a usar en forma plena [*los*] y emerge el plural en *-es*: *los ojo*; *señores*; *las hojas*. Los ejemplos que siguen ilustran cada etapa.

I etapa	*MAR: ostá [= no está] a bota [=la/s bota/s] (1;8)
	*MAR: e nene [el/la nene/a] a botas [= las botas] (1;8)
II etapa	*IDA: y sale [l] # y ya [ll] y sale # y sale e(l) huevo ... (3;0)
	*IDA: (l)o(s) huevo. (3;0)
	*MAG: mira pece(s). (1;9)
	*MAR: Caca a [=en] manos [...] (1;9)
III etapa	*MAG: los ojo [*]. (2;0)
	*IDA: co señores. (3;0)
	*IDA: las hojas [%pho: lar óhah]. (3;3)
	MAG: aquí están la [] llaves. (2;0)

2.4. Adquisición de la concordancia (L2).

Como se ha visto en la sección 1, muchas veces las palabras individualmente no ofrecen indicios sobre la asignación de género a la clase masculina o femenina: las terminadas en consonante (*la canción, el camión*), las terminadas en *-e* (*la calle, el aire*), palabras «engañosas» (*la radio, el teorema*); palabras que cambian de significado (*el capital / la capital*). El aprendiente únicamente tiene a disposición pistas distribucionales para asignar la clase: *el aire fresco; el teorema complicado; el capital financiero*. Es decir que en el caso del género el estudiante se enfrenta a un problema de asignación en el léxico mental y a otro computacional de concordancia. En cambio, en cuanto al número, aparte de formas que no varían (*crisis*) o que son siempre singulares (*dinero*), se debe representar poca información en el léxico y se espera que el problema sea mayormente computacional. Se distinguirán entonces errores de asignación (de valores de rasgos en el léxico) y errores de concordancia (de computación de valores de rasgos). De este modo, por ejemplo, son errores de asignación **muchos *dineros; *las teoremas *complicadas*; donde todas las palabras evidencian asignación equivocada del valor de género o número. Mientras que se consideran de concordancia: *los bolígrafos *amarillos; *los crisis repetidas*, ya que al menos uno de los objetivos demuestra asignación correcta. Bastante investigación sobre el género y el número (y su concordancia) en L2 se ha basado en la Gramática Universal (GU) [ver capítulo siguiente] discutiendo cuestiones de representación de rasgos. Sin embargo, en esta sección se pasará por alto dicho debate y se pondrá el acento en las variables encontradas que inciden en facilitar o no la adquisición. Se dividirán los estudios según empleen técnicas *online* u *offline* tanto en comprensión como en producción [para una revisión bibliográfica ver: Alarcón (2009); Kotz

(2009); van Hell y Kotowicz (2010); Rothman, Gonzalez Alonso & Miller (2018)]. También se considerarán estudios sobre aprendizaje estadístico de dependencias no adyacentes, como es el caso de la concordancia.

2.4.1. Estudios *offline*.

Estos estudios realizan análisis de errores a partir de conversaciones grabadas; usan diferentes tareas para elicitación de lenguaje o utilizan juicios de gramaticalidad.

Finnemann (1992) analizó los errores de tres aprendientes (L1: inglés) de nivel inicial cuyo aprendizaje siguió durante seis meses, grabando entrevistas cada 2 o 3 semanas. Encontró que, en general, la tasa de adquisición de la concordancia de número fue más rápida que la de género. Los estudiantes manifestaron también preferencias por el singular o el plural como opciones *default*. La concordancia de género fue mejor cuando los alumnos se referían a ellos mismos y con referentes humanos femeninos. Por otra parte, a los nombres que los estudiantes consideraban de referencia ambigua (*amigo* como «los amigos» y *persona* como «la gente») se les asignaba mayormente el masculino. En cuanto a los tipos de modificadores, y respecto al plural y al femenino, hubo relativamente pocos errores con los artículos (menos con los definidos que con los indefinidos). Los demostrativos fueron reducidos a la forma masculina «este». Los indefinidos obtuvieron menos errores de plural que de femenino (el autor especula que es así por la «pluralidad» semántica de formas como *todo* o *mucho*). Por otro lado, la morfología nominal no incidió para el género (no se notó tanta diferencia entre los errores para los sustantivos en *-o/-a* y el resto de morfología menos transparente).

En otro estudio de análisis de errores, Fernández-García (1999) investigó la concordancia de género en el sintagma nominal en siete aprendientes ingleses de nivel intermedio a partir de entrevistas grabadas. En general hubo menos errores en contextos masculinos. El masculino se usó mayormente como *default* en artículos indefinidos y adjetivos. Además se verificó más precisión con artículos que con adjetivos. Al contrario del estudio anterior, los sustantivos que marcan las clases en *-o/-a* (*libro*) lograron mayor precisión que los no marcados (*flor*). También existieron menos errores para aquellos motivados semánticamente en el sexo biológico. Asimismo, los alumnos produjeron menos errores con los sustantivos animados (*tío*) que con los inanimados (*tenedor*).

Bruhn de Garavito (2008) estudió el plural (pero no la concordancia específicamente) en hablantes de francés L1 divididos en dos grupos (de edades de entre 14 a 17) pertenecientes, respectivamente, al nivel 1 y 2 de español como lengua extranjera. Ambos grupos adquirieron el orden N – ADJ. Respecto a los errores, el uso de plurales en contexto singular (*un libros*) fue de frecuencia negligible; en cambio, la omisión del plural en contextos que lo requerían (*tres balón*) resultó alta en el primer grupo pero baja (aunque no negligible) en el segundo. El plural en */-s/* se adquirió antes que el plural en */-es/*. En ambos grupos los errores de omisión del plural en */-es/* fueron más altos que en */-s/*. Se hipotetizó que la dificultad del plural en */-es/* se debió a un *transfer* de la estructura silábica y prosódica del francés, en la cual la */-s/* no puede ser silabificada (excepto en la *liaison*), lo cual les traería problemas a los hablantes de francés, al menos en una primera etapa. Los pasos de adquisición del plural encontrados eran: (1) plural nulo (por influencia de la estructura silábica de la L1); (2) plural en */-s/* (se adquiere que la */-s/* puede formar parte del ritmo de la sílaba, al contrario del francés); (3) plural en */-es/*.

Liceras, Díaz & Mongeon (2000) compararon la adquisición de género y número en L1 y L2 en el contexto del sintagma nominal y de elipsis (*las faldas negras / las negras*) en cuatro niños (edades: 4 y 8 años). Los niños de L2, a diferencia de los de L1, no producían proto-determinantes (*a-PROT bota*). Los errores de concordancia de género (*el zapatos; el cama*) en L1 y L2 tenían porcentajes muy similares (variaban entre 1.5 % - 5 %); aunque la cantidad de omisiones de artículos (*_ cama*) eran mayores (alrededor de 8 % y 20 %). Sin embargo, estos errores sucedían dentro de la fase de proto-determinantes en L1 para desaparecer luego; en cambio, en L2 persistían. En cuanto al número, los errores en L2 eran siempre menos que los de género y eventualmente desaparecían, como sucede en L1.

Francheschina (2001) siguió a un hablante (L1: inglés) que había alcanzado nivel superior («Martin»). El hablante exhibía alto grado de precisión en la concordancia de artículos, adjetivos, demostrativos y pronombres (entre 85 y 92 %). Sin embargo los errores de género eran siempre más que los de número (de 257 errores solamente el 7 % eran de número y el resto de género), concentrándose los de número en los

adjetivos. Asimismo, Martin extendía más veces el uso del masculino a contextos femeninos que viceversa en adjetivos y artículos.

Bruhn de Garabito y White (2002) estudiaron el género en estudiantes de secundaria (L1: francés) divididos en dos grupos según estuvieran terminando el primer o el segundo año de español. Los estudiantes adquirieron el orden N - Adj. sin problemas. La precisión en general fue alta (arriba del 85 %). Se verificaron más errores en el artículo definido que en el indefinido. Contra-intuitivamente, los errores de asignación fueron mayores con controladores de género motivados semánticamente que con aquellos motivados morfológicamente (sugiriendo que los alumnos estaban prestando atención a la morfología). Nuevamente hubo más sobre-extensiones del masculino a contextos femeninos, es decir, una preferencia de masculino como *default*; sin embargo, hubo estudiantes que prefirieron más el femenino. En lo que atañe a los adjetivos, en cambio, la forma de *default* fue claramente el masculino tanto en contexto atributivo como predicativo. Comparativamente, la precisión resultó mejor en los artículos que en los adjetivos.

White et al. (2004) investigaron producción y comprensión en concordancia en género y número con estudiantes franceses e ingleses de español L2 divididos en grupos de nivel inicial, medio y superior. La producción fue elicitada por medio de un juego comunicativo y de una tarea de descripción de fotografías. Para la comprensión, en cambio, se crearon unas frases con elipsis nominal en las que el género y el número de determinantes y adjetivos se manipulaban para poder seleccionar la imagen correcta (por ejemplo: 'maleta' [imagen] → *ponlas ahí cerca de la _ roja*). En lo que atañe a la producción, tanto tomando en cuenta bigramas *Artículo + Nombre*, como trigramas *Art. + N + Adj.*, la precisión en general fue alta (arriba del 83 % para el número y del 71 % para el género). Se volvió a observar mayor facilidad para adquirir la concordancia del artículo respecto de los adjetivos por un lado y más precisión en el número respecto al género, por el otro. También se notó nuevamente la tendencia a utilizar el masculino como *default*. Los errores desaparecerían a medida que subía el nivel. En suma, pues, se detectó un efecto de competencia (la precisión difería significativamente de acuerdo al nivel) y uno de rasgo (diferencia significativa en la producción de (valores de) rasgos correctos requeridos por el contexto, por ej., masculino correcto en el adjetivo si el controlador era masculino). Sin embargo, no hubo efecto de L1 (los aprendientes ingleses no diferían significativamente de los franceses). En lo concerniente a la comprensión, los aprendientes de los tres niveles sacaron buenos resultados con respecto a las oraciones que manipulaban el número. Incluso el grupo inglés de nivel inicial, aunque difería significativamente del grupo de control de españoles, tenía menos errores en número que en género. Además los errores de género y número desaparecían hacia el nivel superior. No se detectaron diferencias entre las L1s. O sea que se verificó otra vez un efecto de competencia y uno de rasgo; pero no uno de L1. En cuanto al género, se encontró evidencia del masculino como *default* también aquí. Si la elipsis contenía artículos y adjetivos femeninos (*la _ roja*), los alumnos tendían a elegir una fotografía correspondiente a un ítem léxico femenino. En cambio, si contenía artículos y adjetivos masculinos, algunos tendían a elegir fotografías que correspondían a ítems femeninos. Es decir que el efecto de sobre-extensión del masculino (como *default*) también se detectaba en la comprensión. Es más, los alumnos de nivel inicial tuvieron precisión de alrededor del 50 % para el masculino (o sea, elegían de modo aleatorio) pero más alta para el femenino.

Echevarría y Prévost (2004) estudiaron la concordancia de plural en los adjetivos atributivos (prenominales) y en los cuantificadores *mucho*, *bastante* y *demasiado*, para alumnos (L1: francés) de nivel intermedio y superior. Utilizaron juicios de gramaticalidad (comprensión) y una tarea de producción para completar con las formas en cuestión adaptándolas al contexto (o sea, poniendo en plural los adjetivos y cuantificadores si el controlador lo requería). En lo que atañe a los adjetivos, la concordancia obtuvo altos niveles de precisión en las dos tareas (arriba del 95 %). En cuanto a los cuantificadores, sin embargo, surgieron diferencias entre *mucho* por un lado, y *bastante* y *demasiado*, por otro. La precisión en *mucho* resultó muy alta en ambas tareas; y en la tarea de producción se observó el orden *mucho* > *demasiado* > *bastante* en los niveles de corrección. En la tarea de juicio de gramaticalidad, la precisión para aceptar oraciones gramaticales fue más baja para *demasiado* y *bastante* (alrededor del 80 % para el nivel superior y del 66 % para el intermedio para ambos cuantificadores). Las formas usadas como *default* fueron *demasiado* (masc. sing.) y *bastante* (sing.).

Lichtman (2009) investigó la concordancia de género en estudiantes principiantes y de nivel intermedio (L1: inglés) manipulando la variable de distancia. Se construyeron dos tareas de comprensión: una más

implícita de juicios de gramaticalidad y otra más explícita de detección de errores (marcando la opción correcta entre las opciones provistas). En ambas tareas se controlaba la distancia entre el controlador y el objetivo, en una escala de cuatro grados: 1. *chico simpático* (atributivo, adyacente); 2. *el director es simpático* (predicativo, + 1 palabra); 3. *el chico de mi apartamento es simpático* (predicativo, + 4 palabras); 4. *el profesor que trabaja conmigo es simpático* (predicativo, + 4 palabras con oración relativa). En la tarea explícita no hubo diferencias entre los grupos: ambos se ubicaron en el rango más alto de corrección (8-12). Por otro lado, para la tarea más implícita, los aprendientes de nivel inicial evidenciaron un efecto estadístico de distancia: aumentando ésta, era menos probable la detección de errores de concordancia. Dicho efecto no se observó para el nivel intermedio; marcando un efecto de competencia. No obstante ello, una vez removidas palabras que podían confundir a los estudiantes de nivel inicial (*hay, está*), se notó un descenso del nivel de precisión con el aumento de la distancia en ambos grupos (más pronunciado en el grupo inicial que en el intermedio, el cual se mantuvo casi estable para los primeros tres grados de distancia).

McCarthy (2008) propuso que los morfemas no marcados de número singular y género masculino están subespecificados para sus valores (o sea que, por ejemplo, el morfema /-o/ posee el rasgo «género» y «número» pero no los valores 'masculino' o 'singular'). En cambio las formas marcadas (el femenino y el plural) sí especifican sus valores. Cuando dichos morfemas se insertan en la sintaxis, donde existe especificación completa de rasgos, pueden producirse errores de *default* como en (a), y errores de incompatibilidad como en (b).

(a) La casa es *viej-o / Los libros son *nuev-o.

(b) El libro es *viej-a / El libro es *viej-o-s.

En (a) el contexto sintáctico (el controlador) está especificado como 'género:femenino' y 'número:plural'; pero el morfema en /-o/ está subespecificado en los valores 'singular' y 'masculino'. Por lo tanto, se produce un error por inserción de una forma no marcada de *default*. En cambio en (b), cuyo contexto es 'número:singular' y 'género:masculino', los morfemas en /-a/ y en /-s/ sí están especificados, por eso, se produce un error por incompatibilidad de valores. Pues bien, McCarthy hipotetizó que en la variación en L2 habrá mayores instancias de errores de *default* (tipo (a)), es decir, prediciendo variación unidireccional. La tarea de producción consistía en responder preguntas sobre fotografías para elicitación de clíticos de objeto directo y adjetivos (atributivos y predicativos). La tarea de comprensión testeaba únicamente los clíticos en un diseño similar al utilizado en White *et al.* (2004). Participaron estudiantes de nivel intermedio y superior (L1: inglés). El grupo intermedio resultó significativamente diferente del grupo de control nativo en ambas tareas. En general la concordancia de género tuvo más errores que la de número. En cuanto al género, había significativamente mayor precisión en adjetivos y clíticos en contextos masculinos que en femeninos tanto en producción como en comprensión solo para el grupo intermedio. Respecto al número, no hubo variabilidad en comprensión ya que los grupos no diferían significativamente de los hablantes nativos. En producción, no se verificaron efectos para los clíticos; pero sí para los adjetivos con diferencias significativas entre singular y plural (el singular con más precisión que el plural indicando su mayor uso en contextos plurales). En suma, tanto para el género como para el número los errores de *default* resultaron más frecuentes que los de compatibilidad, como fue predicho por la hipótesis. McCarthy sugirió que el orden de adquisición de los morfemas de género y número es *sub-especificado* > *especificado*, lo cual explica que al principio los errores sean bidireccionales: por ejemplo masculino en contexto femenino y viceversa (ya que no hay valores para ningún rasgo); pero luego, una vez adquiridas las formas especificadas, que haya preponderancia de las formas de *default*. Esta suposición sugiere un déficit representacional inicial, lo cual dejaría afuera efectos de *transfer*.

Montrul *et al.* (2008) estudiaron las diferencias en producción y comprensión en la concordancia de género entre estudiantes de L2 de L1 inglesa (grupo 1) y aquellos nacidos en EEUU e hijos de padres mexicanos pero expuestos al inglés antes de los cinco años (grupo 2). El grupo de control eran hablantes nativos. La tarea de comprensión consistía en marcar la palabra correcta en respuesta a una pregunta con elipsis nominal (*no quiero llevar las _ de ese color > las maletas*). Se encontró un efecto de grupo: mayor precisión en el grupo 1 que en el 2; y otro (valor de) rasgo: ambos grupos obtuvieron mejores niveles de corrección en el masculino que en el femenino. Incluso los alumnos del grupo 1 con precisión mayor al 80% tenían

una competencia inicial / intermedia; más baja que la de los del grupo 2 en los mismos niveles de precisión. La segunda tarea consistía en subrayar el artículo / adjetivo (atributivo o predicativo) correcto basándose en nombres más transparentes (en *-o/-a*) o menos transparentes (en *-e*, consonante u opuestos del tipo de *la mano*): *el / la población de los / las países desarrollados / desarrolladas*. Se halló una mayor precisión en: (a) artículos respecto a adjetivos (efecto de modificador); (b) masculino respecto a femenino (efecto (de valor) de rasgo); (c) grupo 1 respecto al 2 (efecto de grupo); (d) los sustantivos más transparentes respecto de los menos transparentes (efecto morfológico). Además se verificaron las interacciones: (a) *grupo*rasgo*: grupo 1 mejor con el masculino que con el femenino y grupo 2 a la inversa; (b) *grupo*modificador*: grupo 1 sin diferencias entre la precisión de artículos y adjetivos; pero grupo 2 con mayor corrección en los primeros que en los segundos. La última tarea se trataba de elicitar producción de sintagmas nominales (*Art. + N + Adj.*) mediante la descripción de una fotografía (foto de una manzana > *veo una manzana roja / madura / ...*). Hubo un efecto de grupo, pero, a diferencia de las tareas anteriores, el grupo 2 obtuvo mejores resultados que el 1. Se verificó también un efecto de animicidad: mayor precisión en los sustantivos transparentes animados que en los inanimados. También se halló un efecto de morfología: los sustantivos terminados en consonante con mejores resultados que los terminados en *-e*; además, los opuestos (*la mano*) fueron los que se erraron más, especialmente con los adjetivos. Asimismo se observó un efecto de valor de rasgo: masculino mejor que femenino; pero ningún efecto de modificador. Luego de un análisis de los errores de producción se encontró una distribución similar para ambos grupos. En general, el porcentaje de errores de asignación y de concordancia fue similar (48% contra 52% respectivamente). Alrededor del 80% fue con el femenino, indicando el masculino como *default*. La mayoría de los errores de concordancia fueron con el masculino; mientras que más de la mitad de los errores con el femenino eran de asignación en sustantivos de morfología no transparente (*-e*, *-cons*, opuestos en *-o*).

Alarcón (2011) llevó a cabo 2 experimentos replicando las tareas 2 y 3 anteriores con estudiantes de L2 ingleses e hijos de inmigrantes (nacidos en EEUU) expuestos al español antes de los 5 años. Ambos grupos tenían una competencia similar de nivel superior; además, participaron hablantes nativos como control. El primer experimento fue similar a la tarea 2 de Montrul *et al.* (2008) usando artículos y adjetivos únicamente atributivos. Si bien los resultados mostraron alta precisión (al menos 92%) solo fue significativo un efecto de morfología: la concordancia de género con sustantivos transparentes mejor que con los menos transparentes, y con los hijos de inmigrantes significativamente mejores que los de L2. En lo que atañe al segundo experimento (tarea 3 de Montrul *et al.* (2008)) se halló significancia en los siguientes factores: (a) grupo: mejor en los hijos de inmigrantes que en los estudiantes de L2; (b) [valor de] rasgo: mejor en el masculino que en el femenino; (c) morfología: mejor en sustantivos transparentes que en menos transparentes; (d) *rasgo*grupo*: la diferencia en precisión entre masculino y femenino fue mayor en el caso de los aprendientes ingleses (masculino: 94%; femenino: 76%); (e) *rasgo*morfología*: la diferencia en precisión entre nombres más y menos transparentes resultó mayor para el femenino; *rasgo*modificador*: la precisión fue significativamente mayor para los artículos que para los adjetivos en el caso del masculino; *grupo*morfología*: ambos grupos con mejores resultados en sustantivos transparentes que menos transparentes pero, al igual que en (d), con diferencia mayor entre los aprendientes ingleses (transparentes: 97%; no transparentes: 73%). En cuanto al análisis de los errores de producción, ambos grupos produjeron más errores de asignación que de concordancia; más con el femenino que con el masculino y más con morfología no transparente que con transparente. En ambos grupos los errores de asignación solo se cometieron con morfología no transparente y en especial con el femenino. Los errores de concordancia fueron más frecuentes con nombres femeninos de morfología no transparente; sugiriendo un rol del masculino como *default* no solo a nivel léxico (asignación) sino también a nivel de la sintaxis (concordancia).

Montrul *et al.* (2013) crearon un experimento con sustantivos diminutivos y no diminutivos, en el cual los participantes veían una fotografía y debían contestar una pregunta, elicitando una respuesta del tipo *determinante indefinido + sustantivo (diminutivo / no diminutivo) + adjetivo*. Por ejemplo, se les mostraba una ilustración con una mesa y un pez amarillo pequeño arriba de ésta y un pez amarillo grande debajo. Se les preguntaba: *¿Qué hay arriba de la mesa?* y se esperaba que se elicitará la respuesta: *un pecesito amarillo*. Los participantes eran estudiantes de L2 ingleses e hijos de inmigrantes (nacidos en EEUU) expuestos al español antes de los 5 años. Hubo otro grupo hablantes nativos de español como control. Se sabe que los diminutivos ejercen un efecto facilitador en la adquisición de L1, respecto de la morfología no transparente

de los nombres. Es decir que al escuchar un diminutivo, por ejemplo, *crucecita*, los niños reciben una pista sobre el género del nombre, no presente si se hubiera usado la forma no diminutiva *cruz*. En concreto, los diminutivos «regularizan» la morfología de género no transparente del sustantivo. La pregunta de investigación fue si los hijos de inmigrantes, expuestos al español desde niños y, por ende, con acceso a más input de diminutivos, lograban una ventaja en la concordancia *determinante-nombre* y *adjetivo-nombre* en los controladores con morfología no transparente respecto de aquellos con morfología transparente; comparativamente a los estudiantes de L2 (con mucha menos exposición a diminutivos). Se manipularon las siguientes variables: (i) «tipo de controlador»: no diminutivo / diminutivo; (ii) «modificador»: determinante / adjetivo; (iii) «morfología del controlador»: transparente / no transparente; (iv) «género del controlador»: masculino / femenino; (v) «grupo»: hijos de inmigrantes / estudiantes de L2. En primera instancia se encontró un efecto de grupo: los hijos de inmigrantes tenían, significativamente, mejor desempeño que los estudiantes de L2. En segundo lugar, hubo efecto de género: ambos grupos lograban mayor precisión con el masculino respecto del femenino. La interacción entre *grupo*morfología* resultó significativa: ambos grupos hacían mejor la concordancia de género con nombre de morfología transparente respecto de los de morfología no transparente. Sin embargo, en el caso de los estudiantes de L2 dicha diferencia de precisión fue significativa, mientras que en el otro grupo no lo fue. Es decir que, los hijos de los inmigrantes alcanzaban mayor precisión respecto de los estudiantes de L2 respecto de la morfología del nombre. No se verificó significancia estadística en la interacción *grupo*tipo de modificador*. En general, los participantes tuvieron menor precisión en la concordancia con los nombres de morfología no transparente y de género femenino, tanto para los diminutivos como para los no diminutivos. En consecuencia, los diminutivos *no* confirieron ventaja alguna para los hijos de inmigrantes, respecto de la concordancia de género. En lo que atañe al análisis de los errores, ambos grupos produjeron más errores de asignación que de concordancia. Si bien los hijos de inmigrantes produjeron en general menos errores que los estudiantes de L2, los primeros cometieron, significativamente, más errores de asignación que los segundos. Además, en ambos grupos, los errores afectaron más a los sustantivos de morfología no transparente, y dentro de éstos, a los femeninos. La frecuencia de errores de concordancia resultó similar en ambos grupos.

Footo (2015) investigó el rol de la transparencia morfológica del controlador en la concordancia de género (singular), con aprendientes ingleses de español L2 de nivel intermedio y superior. En el experimento se utilizaba una tarea de completamiento del tipo: *la pared (N1) del baño (N2) está... limpia/*o*. Las variables manipuladas fueron: (i) «género»: el género del núcleo N1 del SN (masculino / femenino); (ii) «morfología»: la transparencia de marca de género en el N1 (transparente / opaca); (iii) «correspondencia»: la igualdad o desigualdad entre los géneros del N1 y el N2 (igual / distinto); (iv) «grupo»: el nivel de competencia (intermedio / superior / nativo). Los resultados arrojaron una interacción significativa *grupo*correspondencia*: los aprendientes de nivel intermedio cometían significativamente más errores que los de nivel superior; y éstos, más que los nativos. Además había más errores en la condición de no igualdad de género entre N1 y N2. No hubo efecto de transparencia morfológica ni de género.

El Cuadro 2.2 resume los resultados más importantes de cada estudio expuesto más arriba.

2.4.2. Estudios *online*.

Metodológicamente estos estudios usan tiempos (latencias) de reacción, tiempos de lectura, *eye tracking* o potenciales relacionados a eventos. Involucran únicamente comprensión.

Alarcón (2009) investigó siete posibles variables de la concordancia de género en adjetivos predicativos, en calidad de predictoras de los tiempos de reacción. Participaron alumnos de nivel inicial, intermedio y superior (L1: inglés). Dichas variables eran: (a) clase con (sin) motivación semántica (animado en correspondencia con sexo biológico / no animado) del núcleo del SN: *el doctor / servicio del consultorio*; (b) clase con (sin) motivación semántica del núcleo del sintagma preposicional [SP]: *el empleado del doctor / consultorio*; (c) congruencia semántica entre el núcleo del SN y el del SP: *el pariente del psicólogo / de la montaña*; (d) morfología (no) transparente del núcleo del SN: *el arquitecto del museo / el detalle del Cuadro*; (e) valor del rasgo de género (femenino / masculino) en el núcleo del SN: *la comida / el almuerzo del secretario*; (f) valor del rasgo de género (femenino / masculino) en el núcleo del SP: *la inyección de la enfermera / del enfermero*; (g) congruencia en valor de género entre el núcleo del SN y del SP: *la tarea /*

ESTUDIO	DISEÑO	PARTICIPANTES	FOCO	RESULTADO
Finnemann (1992)	L; PR; En	Np=3; P; L1: Ing.	G, N	N > G; default: Masc y Fem PL y Fem: Art. Def. > Art. Indef. IND: PL > Fem; DEM: solo «este»
Fernandez-Garcia (1999)	T; PR; En	Np=7; I; L1: Ing.	G	default: Masc; Art. > Adj.; MT > MNT; An > InAn
Liceras, Díaz & Mongeon (2000)	T; PR; En	4 niños (4 y 8 años) 2 L1 ing. y 2 L1 Esp	G, N	Nativos: proto-determinantes; Pres. G: L1 = L2 menos Errores de N con más Competencia N > G; omisiones de Art.: persisten en L2
Francheschina (2001)	Ca; PR; En	Np=1; S; L1: Ing.	G, N	G, N: Precisión alta; default: Masc. Errores de G > Errores de N
Bruhn de Garabito y White (2002)	T; PR; TDF	Np=42; P; I; L1: Fr.	G	G: Pres. alta; Default: Masc. Art. > Adj.; Art. Indef > Art. Def.
White <i>et al.</i> (2004)	T; PR; CO; TDF; SFE	Np=116; P; I; S ; L1: Ing., Fr.	G, N	G, N: Pres. alta; Default: Masc. N > G; Art. > Adj.;
Echevarría y Prévost (2004)	T; PR; CO TC; TJG	Np=29; I; S ; L1: Fr.	N	Precisión en PR: <i>mucho</i> > <i>demasiado</i> > <i>bastante</i> Precisión en CO: <i>mucho</i> > <i>demasiado</i> / <i>bastante</i> Precisión en Adj.: alta en PR y CO
Bruhn de Garabito (2008)	T; PR; TDF	Np=42; P; I ; L1: Fr.	Plural	etapas: plural nulo > /-s/ > /-es/
Lichtman (2008)	T; CO; TJG; TDE	Np=38; P; I ; L1: Ing.	G	TJG: efecto de distancia fuerte en P y débil en I
McCarthy (2008)	T; PR; CO; TDF; SFE	Np=24; I; S ; L1: Ing.	G, N	N > G; default: Masc y SG
Montrul <i>et al.</i> (2008)	T; PR; CO; TDF; SFE; TOM;	Np=163; P ; L1: Ing. BTe; BTa	G	BTe: CO > PR; BTa: PR > CO MT > MNT (-cons > -e > opuestos); An > InAn; default: Masc; Art. > Adj.; ErrAsig (mayor en MNT y Fem) = ErrConc
Alarcón (2011)	T; PR; CO; TDF; TOM	Np=53; S ; L1: Ing. BTe; BTa	G	MT > MNT; Art. > Adj.; default: Masc; ErrAsig (mayor en MNT y Fem) > ErrConc
Montrul <i>et al.</i> (2013)	T; PR; TDF	Np=53; S ; L1: Ing.	G	Masc > Fem; MT > MNT; HIn > EsL2 ErrAsig (mayor en MNT y Fem) > ErrConc
Foote (2015)	T; PR; CO	Np=24; I; S ; L1: Ing.	G	grupo: I > S; Correspondencia: diferente > igual

Leyenda: «TDF»: Tarea de descripción de figuras; «En»: Entrevistas grabadas; «Ca»: Estudio de caso; «PR»: Producción; «CO»: comprensión
«SFE»: Tarea de selección de figuras a partir de oraciones con elipsis nominal; «TC»: Tarea de completamiento; «L»: estudio longitudinal;
«TJG»: Tarea de juicios de gramaticalidad; «TDE»: tarea de detección de errores; «TOM»: tarea de opción múltiple; «T»: estudio transversal;
«Np»: número de participantes; «P»: nivel inicial; «I»: nivel intermedio; «S»: nivel superior; «Ing»: inglés; «Fr»: francés; «G»: género;
«N»: número; «BTe»: bilingües tempranos; «BTa»: bilingües tardíos; «Masc»: masculino; «Fem»: femenino; «PL»: plural; «SG»: singular;
«Art. Def.»: artículo definido; «Art. Indef.»: artículo indefinido; «IND»: adjetivos indefinidos; «DEM»: adjetivos demostrativos;
«Art»: artículos; «Adj»: adjetivos; «An»: animados; «InAn»: inanimados; «ErrAsig»: errores de asignación; «ErrConc»: errores de concordancia;
«MT»: nombres de morfología transparente; «MNT»: nombres de morfología no transparente; HIn: hijos de inmigrantes; EsL2: estudiantes de L2.

Cuadro 2.2: Estudios *offline*

el *trabajo de la compañera*. Los estímulos tenían la forma *SN complejo + verbo ser + adjetivo predicativo: el arquitecto del museo es... orgullosa (orgullosa)*. Un análisis de regresión múltiple con las variables anteriores como predictoras y los tiempos de reacción de la concordancia como respuesta, reveló como significantes las siguientes: (1) nivel inicial: variables (a), (b), (f); (2) nivel intermedio: variables (a), (b); (3) nivel superior: variables (a), (b), (e), (f); nativos: variables (a), (d). A partir de éstas se observó la direccionalidad de los tiempos de reacción TR (más rápido que '>') arrojando lo siguiente: (1) variable (a): TR motivación semántica > no semántica para todos los niveles y los hablantes nativos; (2) variable (b): TR motivación no semántica > semántica para todos los niveles; (3) variable (d): TR morfología transparente > no transparente en nativos; (4) variable (e): TR femenino > masculino para el nivel superior; (5) variable (f): TR incongruente > congruente. Sin embargo, nótese que en el estudio el coeficiente R^2 de la regresión resultó en general muy bajo, es decir que las variables, aunque significativas, explicaban una pequeña porción de la variabilidad de los tiempos de reacción; y por lo tanto, resultaron ser predictoras débiles. En suma, la motivación semántica [+/- animicidad] del núcleo de SN y SP fue el único predictor presente en todos los niveles, apuntando a un rol facilitador de los nombres animados en el aprendizaje de la concordancia.

Con aprendientes de nivel inicial (L1: inglés), Sagarra (2007) estudió el procesamiento de la concordancia de adjetivos con o sin distancia estructural midiendo latencias de lectura. Los estímulos se distribuían según cuatro condiciones: (a) concordancia y error de concordancia del adjetivo en el SN (sin distancia estructural): *la mujer lava la blusa blanca / *blanco en la cocina*; (b) concordancia y error de concordancia en el adjetivo posicionado dentro de una relativa (con distancia estructural): *la mujer lava la blusa que era blanca / *blanco en la cocina*. Además se dividió a los participantes según hubieran tenido alto o bajo rendimiento en un test de memoria de trabajo. Se midió el tiempo de lectura específicamente en el adjetivo *N* y en la palabra siguiente *N + I* (para controlar latencias residuales). Luego de cada estímulo se presentaba una pregunta (con respuesta SI / NO) de comprensión sobre la oración recién leída con el fin de verificar si se estaba leyendo el contenido semántico. Los resultados arrojaron un efecto de distancia: los tiempos promedio de lectura en *N* y *N + I* eran más altos en la condición sin estructura (SN) que en aquella con ella (relativa). Esto se explicó aduciendo que estudiantes de baja competencia no logran procesar varias fuentes de información a la vez. Es decir que, a larga distancia, cuando leían las relativas procesaban información sintáctica pero no morfológica (concordancia). No hubo efecto de capacidad (alta / baja) de memoria de trabajo ni de concordancia (correcta / error). Sin embargo en un análisis de la precisión obtenida en las preguntas de comprensión se obtuvo un efecto de memoria de trabajo: los estudiantes con alta capacidad de memoria de trabajo lograron más precisión en las preguntas de comprensión en la condición de distancia estructural (tanto en concordancia correcta como en el error de concordancia). Ahora bien, la insensibilidad observada para los errores de concordancia podía deberse a un factor de competencia (los estudiantes principiantes simplemente no han adquirido todavía la concordancia, de ahí que no detectaran una violación).

Sagarra & Herschensohn (2010) replicaron el experimento con alumnos de nivel inicial e intermedio. Testearon las latencias de lectura en oraciones con las siguientes condiciones: (a) concordancia correcta de género y número: *el ingeniero presenta el prototipo famoso en la conferencia*; (b) error de género: ... *el prototipo *famosa ...*; (c) error de número: ... *el prototipo *famosos*. Además se incluyó una tarea con juicios de gramaticalidad en la que los participantes debían juzgar oraciones según fueran correctas o no, identificar la causa del error y expresar su grado de confianza en la respuesta según una escala. Un grupo de hablantes nativos funcionó como control. Se evaluó a los alumnos según un test de memoria de trabajo. En cuanto a las latencias en *N*, *N + I* y la precisión en las preguntas de comprensión, no se observaron diferencias en las latencias de lectura ni en la precisión de preguntas en principiantes. En cambio, los de nivel intermedio y los nativos leían más lentamente oraciones con errores de género y número que las correctas. La tarea con juicios de gramaticalidad arrojó, en primer lugar, que los principiantes y los de nivel intermedio evidenciaban mayor precisión con concordancia (género / número) que con discordancia (errores). En segundo lugar, todos los aprendientes fueron más precisos y expresaron mayor grado de confianza con la concordancia de número respecto de la de género; y eran mejores identificando errores de número que de género (O sea que los alumnos procesan más fácilmente la discordancia de número que la de género). Además los alumnos de nivel intermedio con mayor capacidad de memoria de trabajo resultaban ser más sensitivos a los errores de género y no a los de número. Nótese, igualmente, que los de nivel intermedio tenían en general mayor capacidad de memoria de trabajo que los principiantes y los nativos. En suma, los resultados apuntaron a

una diferencia en el procesamiento entre el género y el número (el primero más demandante cognitivamente que el segundo) en alumnos con más competencia (nivel intermedio); y a un rol facilitador de la memoria de trabajo en aras de ganar sensibilidad a la morfología del adjetivo (concordancia).

Sagarra & Herschensohn (2013) re-analizaron los datos de Sagarra & Herschensohn (2010) separando las variables de (no) concordancia de género y animicidad. Se observaron *mayores latencias* de lectura (en N y $N + 1$) en estudiantes de nivel intermedio y nativos en: (a) oraciones con discordancia de género respecto a las de concordancia, con nombres [+/- animados] (efecto de concordancia); (b) oraciones con sustantivos [+animados] respecto a los [-animados], ya sea con o sin concordancia de género (efecto de animicidad). En lo referente a los juicios de gramaticalidad, los principiantes y los de nivel intermedio lograron más precisión en oraciones con concordancia de género respecto a las discordantes. No obstante, los de nivel intermedio fueron más precisos que los principiantes en las discordantes. Además, en línea con las latencias, los de nivel intermedio evidenciaron mayor precisión en oraciones con sustantivos inanimados respecto a los animados. En lo que atañe a la memoria de trabajo, dicha variable correlacionó positivamente en estudiantes de nivel intermedio (de mayor capacidad de memoria de trabajo) con oraciones cuyos controladores [+/- animados] tenían discordancia con adjetivos más que con los que concordaban correctamente. En resumen, el procesamiento en L2 con los nombres animados parece ser más demandante cognitivamente que los inanimados, quizás porque en el caso de una discordancia, el procesador necesita más tiempo para elegir entre dos sufijos en los animados (*espos-o/-a*) que en los inanimados donde solo es necesario uno (*mes-a*).

Mediante la técnica de *eye-tracking*, Keating (2009) abordó el procesamiento de la concordancia / discordancia de género en adjetivos según tres condiciones de distancia: (a) adyacente al controlador dentro del SN; (b) en posición predicativa; (c) en una subordinada con CUANDO. Por ejemplo: (a) *Un sombrero nuevo /*-a cuesta mucho cuando se compra en esa tienda*; (b) *un libro es bastante pequeño /*-a cuando tiene solo treinta páginas*; (c) *un libro recibe bastante atención cuando es nuevo /*-a y popular*. Participaron estudiantes de nivel inicial, medio, superior; y nativos como control. Luego de la lectura de cada estímulo, se debía responder una pregunta por SI / NO de contenido sobre lo leído. La única condición significativa resultó ser la (a): en el caso de los aprendientes de nivel superior, los adjetivos discordantes conllevaron más tiempo de lectura que los que concordaban con el controlador en el dominio local. Los nativos, por otra parte, demostraron sensibilidad a las violaciones de concordancia en las tres condiciones. En suma, los aprendientes avanzados son comparables a los nativos en procesamiento de la concordancia de género local pero no a larga distancia; ésto último, se hipotetizó, podría deberse a limitaciones de memoria de trabajo fuera del dominio local.

Keating (2010) repitió el experimento con aprendientes avanzados y nativos como control, aunque está vez manteniendo constante la distancia estructural (en el experimento anterior la distancia lineal aumentaba con la estructural confundiendo los efectos de ambas). Incluyó además un test de memoria de trabajo a fin de correlacionarlo con la sensibilidad de los aprendientes a la discordancia según el aumento de la distancia lineal entre controlador y objetivo. Los estímulos mantenían constante la distancia estructural en un nodo (SV) pero aumentaban la lineal según una, cuatro o siete palabras entre el adjetivo predicativo (no local) y su controlador: (a) *la tienda está abierta /*-o (1 palabra)*; (b) *la mochila de la estudiante está llena /*-o de libros (4 palabras)*; (c) *la falda en la tienda de ropa femenina es roja /*-o (7 palabras)*. Los resultados arrojaron que los nativos eran sensibles a la distancia de 1 y 4 palabras pero no a la de 7; y los aprendientes avanzados, únicamente a la de 1 palabra. La diferencia, sin embargo, fue que los nativos procesaron las discordancias antes (durante *first pass reading*) y los aprendientes de modo tardío (durante *second pass reading*). Por lo tanto, se detectó esta vez sensibilidad no local en los avanzados (nótese que la falta de detección en el primer estudio pudo deberse a que en posición predicativa la distancia era de 2 palabras y de 1 en éste). La correlación entre memoria de trabajo y distancia (en *first pass reading*) fue significativa para 1 palabra y 4 palabras en el caso de los estudiantes; pero no hubo significancia para los nativos en ninguna condición. Es decir que era más probable que aquellos aprendientes con mayor capacidad de memoria de trabajo pasasen más tiempo leyendo adjetivos discordantes en comparación con los controles, lo cual redundó en la falta de sensibilidad a errores en *first pass reading*.

Foot (2011) midió tiempos de lectura en la concordancia de género según dos condiciones de distancia: (a) adyacencia: *dicen que el libr-o blanc-*a está en esa mesa*; (b) larga distancia: *el poll-o del taco está ric-*a pero picante*. Participaron bilingües tempranos (hijos de inmigrantes) y tardíos, ambos con competencia

de nivel superior; así como nativos (control). Se tomaron tiempos de lectura en la palabra anterior al adjetivo ($N-I$), en el adjetivo (N) y en la posterior al adjetivo ($N+I$). Se incluyó un test de memoria de trabajo. Los resultados arrojaron sensibilidad (latencias de lectura mayores) para las oraciones agramaticales en ambas posiciones de distancia. No obstante, en los nativos y en los bilingües hijos de inmigrantes dicha sensibilidad fue temprana (posiciones $N-I$ y N), mientras que para el otro grupo fue tardía ($N+I$). Además las diferencias entre los tiempos de lectura de las discordancias y las concordancias fueron mayores en la condición de adyacencia que en la de larga distancia; es decir, se registró mayor sensibilidad a las violaciones en (a) que en (b). Se llevaron a cabo tests de correlación en los bilingües tardíos, con el fin de verificar si la reducción de sensibilidad en las discordancias a larga distancia se relacionaba con la capacidad de memoria a largo plazo. No evidenciaron significancia.

Lew-Williams y Fernald (2010) investigaron si aprendientes de nivel intermedio (L1: inglés) lograban usar información del género de los artículos (*la, el*) para procesar más eficientemente controladores nominales conocidos y nuevos. Nativos actuaron como control. Los sujetos miraban figuras con dos imágenes cuyas palabras tenían: (a) artículos con géneros iguales: *la pelota, la galleta* (no informativos sobre el posible referente); o (b) diferentes: *la pelota, el zapato* (con información sobre el referente). Al mismo tiempo, escuchaban una oración con una instrucción / pregunta: *encuentra la pelota / ¿dónde está la pelota?*. Se midieron tiempos de reacción dependiendo de cuán rápido los aprendientes cambiaban su mirada hacia el referente correcto (si habían comenzado mirando el opuesto). El primer experimento involucró solamente palabras muy frecuentes. Los resultados no arrojaron significancia en el caso de los aprendientes: si bien los nativos sacaron ventaja del valor de género del artículo para identificar el referente más rápidamente cuando los referentes pertenecían a clases de género diferente; dicha información no influyó en los aprendientes. En otro experimento los sujetos fueron expuestos la misma cantidad de veces en una fase de entrenamiento a figuras con cuatro objetos no familiares de género en *-a* y *-o* (*mira, es una catela / una pifa / un durino / un tebo*); para luego realizar el test de igual modo que en el primer experimento: *¿dónde está la catela?*. Nótese que la fase de entrenamiento involucraba artículos indefinidos, mientras que en el test se usaban definidos. Es decir que para realizar correctamente el experimento, los sujetos debían generalizar los valores de género de un tipo de artículo a otro y usarlos como información predictiva del referente. Los resultados fueron, como antes, no significativos para los aprendientes. Mientras que los nativos usaban el valor de género del artículo para identificar el referente inmediatamente, los aprendientes *esperaban* a oír el nombre antes de cambiar la mirada al referente correcto (aun conociendo metalingüísticamente la diferencia entre artículos definidos e indefinidos).

Grüter *et al.* (2012) combinaron los experimentos anteriores con alumnos de nivel superior. Había una fase de entrenamiento con los mismos objetos nuevos (*una catela, etc.*) usando artículos indefinidos y el experimento propiamente dicho que combinaba artículos definidos (*el, la*) con palabras cuyos referentes eran conocidos (*la pelota*) o nuevos (*la catela*). En general los aprendientes y los nativos obtuvieron más rapidez en la condición de diferente género respecto a la de igual género; y en la de nombres conocidos respecto a los nuevos. No se registró un efecto de grupo: nativos e aprendientes procesaron igual. Sin embargo, dicho resultado nulo no necesariamente constituye evidencia de que ambos grupos sacan igual ventaja del género de los artículos para el procesamiento. De hecho, un análisis intra-grupo reveló una diferencia significativa entre las condiciones de igual y diferente género para los nativos pero no para los aprendientes en el caso de los nombres conocidos. Este resultado implica que cualquiera sea la ventaja que se obtiene de los artículos para identificar referentes de nombres conocidos resulta más débil en los aprendientes respecto de los nativos. A pesar de ello, en el caso de los nombres nuevos, se hallaron efectos marginales para las condiciones anteriores. O sea que los aprendientes usaban posiblemente información del artículo en el caso de los nombres nuevos pero no de los conocidos. Los autores atribuyeron esto último al hecho de que las condiciones de aprendizaje de los nombres nuevos simulaban a las del aprendizaje de los niños; es decir, por medio de pares co-ocurrentes inanalizados *art + nombre*, permitiendo abstraer información sobre clases de género a partir de información distribucional (co-ocurrencias). En cambio, los nombres conocidos fueron aprendidos presumiblemente uno por uno; y entonces, la aparición del artículo no agregaba información adicional para el mapeo de nombres a referentes. Ésto explicaría por qué los aprendientes no lograban extraer información de género en el artículo en el caso de los nombres conocidos pero sí con los nuevos. Se llevaron a cabo también otros dos experimentos *offline*: (a) uno de comprensión

para identificar figuras a partir de oraciones con elipsis (similar a White *et al.*, 2004); (b) una tarea de elicitación de producción (similar a Montrul *et al.*, 2008), en la cual los sujetos escuchaban una pregunta sobre referentes que diferían en una cualidad (por ej., color) y respondían elicitando el sintagma *art. + (nombre) +adj.*: [figura de mariposas de diferentes colores] pregunta: *¿Cuál te gusta más?*; respuesta: *la (mariposa) roja*. En el experimento (a) ambos grupos obtuvieron marcas de 96% o más, por lo cual no había diferencia entre nativos y aprendientes de nivel superior en comprensión *offline*. En cuanto al (b), se analizaron los errores según: (a) errores de concordancia: *la (mariposa) *rojo*; (b) errores de asignación: *el (mariposa) rojo*. La gran mayoría de los errores en los aprendientes fueron de asignación (17% contra 1.5% de concordancia). Dichos errores fueron más frecuentes: (a) con nombres de morfología no transparente (-e, -cons) que con los transparentes (-o, -a); (b) con nombres femeninos respecto a los masculinos. Estos resultados replicaron a los de Alarcón (2011) y Montrul *et al.* (2013).

Mediante eye-tracking, Dussias *et al.* (2013) replicaron el experimento de Lew-Williams y Fernald (2010) usando artículos definidos y nombres femeninos en -a (*pera*) de morfología transparente y masculinos con morfología transparente en -o (*pozo*), y opaca en -cons (*caracol*, *arroz*, *reloj*) y -e (*punte*). Mientras que miraban las figuras con los objetos, los sujetos escuchaban oraciones que contenían los pares *art + nombre* en posición intermedia o final (*el estudiante estaba dibujando el reloj que vio ayer*). Participaron estudiantes ingleses de nivel superior e inicial; e italianos de nivel inicial (nativos como control). Los resultados arrojaron que tanto los nativos como los aprendientes de nivel superior sacaban provecho de la información del género masculino y femenino del artículo para acelerar el acceso al léxico e identificar más rápidamente el referente. En cambio, los principiantes italianos procesaron únicamente el género femenino del artículo; mientras que en el caso de los principiantes ingleses no hubo efecto de anticipación. A diferencia de Grüter *et al.* (2012), los nativos y los de nivel superior procesaron de modo similar el género del artículo de palabras conocidas; o sea, cambiando la mirada inmediatamente una vez escuchado el artículo en la condición de género diferente pero esperando a oír el nombre en la de igual género. Los autores hipotizaron que el resultado mixto de los italianos (procesar el género femenino del artículo para anticipar el del nombre pero no en el caso del el masculino) se debía a una combinación de un efecto facilitador (la L1 que también marca dos valores de género y una gran similaridad de vocabulario con el español) para procesar el género del artículo; y uno inhibitorio (competencia baja) que no permitía dirigir recursos cognitivos para capitalizar la información del masculino; el cual, en contraste con el femenino, en el experimento tenía también morfología no transparente, conllevando mayor esfuerzo cognitivo.

Por último se consideran los estudios basados en potenciales relacionados a eventos (ERPs), muy importantes porque se consideran una vía de acceso al conocimiento implícito o procedural.

Tokowicz y MacWhinney (2005) investigaron efectos de *transfer* entre inglés (L1) y español (L2) con estudiantes de nivel inicial. Consideraron tres tipos de construcciones con sus respectivas violaciones de concordancia: (a) progresivo: *his grandmother *cooking very well / su abuela *cocinando muy bien* (violación de ausencia de auxiliar similar en L1 y L2); (b) género del artículo: *ellos fueron a *un / una fiesta; they went to *a-MASC / a-FEM party* (violación de género ausente en inglés por falta del rasgo en la lengua); (c) número del artículo: **el / los niños están jugando; *The-SG / The-PL boys are playing* (violación de número ausente en inglés por falta de rasgo en el artículo, pero presente en el nombre). En el caso (a) se espera que, mediante *transfer* positivo, los aprendientes detecten agramaticalidades; en el (b) hay ausencia de *transfer* (o sea de influencia) debido a que el inglés no codifica género, por eso se espera que los estudiantes también detecten errores. En cambio, en el caso (c) habrá *transfer* negativo ya que los estudiantes ingleses están acostumbrados a ignorar la información de número en el artículo, por lo que se espera que no detecten las violaciones en español. Se encontró precisamente un efecto de P600 (marcador de violación gramatical) en las construcciones (a) y (b), pero no en (c). Por lo tanto la similaridad entre las lenguas influencia la probabilidad de que haya o no procesamiento implícito de concordancia en principiantes. Aunque las lenguas compartan un rasgo (número: plural); si éste se implementa de diferente modo en las lenguas (produciéndose un conflicto de competición) el procesamiento resulta más difícil. El estudio también incluyó juicios de gramaticalidad como medida explícita (ya que los aprendientes pueden recurrir a la memoria declarativa y recordar reglas metalingüísticas para resolverlos). Lo interesante fue que, para el género, la precisión resultó debajo del 50%. Es decir que, explícitamente los estudiantes no detectaban las violaciones pero implícitamente sí. Bond *et al.* (2011) replicaron el experimento (estudiantes principiantes)

con construcciones de concordancia de número sujeto-verbo (presente en inglés y español); concordancia de género en el adjetivo (única al español) y concordancia de número en el adjetivo (presente en español pero ausente en inglés, que lo marca en el controlador nominal). Hallaron el efecto de P600 para discordancias en las tres condiciones. Foucart y Frenck-Mestre (2011) aplicaron ERPs a la concordancia de género en artículos y adjetivos en el sintagma nominal, para estudiantes de francés (L2) de nivel superior cuya lengua materna era el alemán. La concordancia entre artículo y nombre es similar en ambas lenguas; en cambio, en el caso del adjetivo no se produce concordancia de género en alemán. Además en francés es canónico el orden *nombre + adjetivo*; mientras que en alemán únicamente existe el orden *adjetivo + nombre*. En las discordancias en francés hallaron en los aprendientes un efecto de P600 para el artículo pero no para el adjetivo (tanto en posición post- como pre-nominal). Interpretaron los resultados en términos de similaridad en las reglas de concordancia en ambas lenguas, considerando que el orden de palabras no resulta influyente. En resumen, estos estudios muestran que es posible procesar implícitamente en L2 un rasgo no presente en la L1 (género) similarmente a los nativos (que evidencian el efecto P600). Además el procesamiento del artículo parece más fácil que el del adjetivo. Respecto de los rasgos presentes en ambas lenguas pero implementados de modo diverso, la evidencia es inconclusiva, ya que Bond *et al.* (2011) halló sensibilidad a errores en construcciones con reglas de concordancia diferentes; mientras que los otros estudios ofrecen evidencia de procesamiento diferenciado.

Dowens *et al.* (2010) estudiaron el procesamiento de la concordancia de género y número (y su violación) en posición atributiva (con el artículo) y predicativa (con el adjetivo): *el suelo está plano y bien acabado*. Los aprendientes eran de nivel superior (L1: inglés). En general, el patrón de ERPs frente a las discordancias resultó similar en nativos y aprendientes: (a) una negatividad temprana entre 300-500 ms (LAN), indicadora procesos *automáticos* de integración morfosintáctica; (b) seguida de una positividad entre 500-900 ms (P600), interpretada como indicadora de procesos más *controlados* de re-análisis y reparación de error; (c) seguida de una negatividad tardía entre 900-1300 ms, indicadora de esfuerzo de memoria de trabajo (nótese que, en posición predicativa, se debe mantener el valor del rasgo del controlador en la memoria para concordarlo luego con el del adjetivo a larga distancia). Sin embargo, se verificaron diferencias entre aprendientes y nativos: el efecto LAN solo se observó en posición atributiva, con duración y distribución más amplia para las violaciones de número; el efecto P600 y la negatividad tardía también tuvieron más amplitud en el caso de discordancias de número respecto a las de género, indicando mayor «profundidad» de procesamiento. Nótese asimismo que el hecho de que el efecto LAN se hallara solo en posición atributiva aboga por un tipo de procesamiento más automático del artículo (como un hablante nativo) en comparación con el adjetivo. En suma, el factor competencia y el de distancia claramente influyen el procesamiento. Además el procesamiento más «profundo» del número apunta a efectos de *transfer* aun en estudiantes avanzados (rutinas de procesamiento de número ya operantes en L1 que facilitan su procesamiento en L2). Dowens *et al.* (2011) repitieron el experimento con aprendientes chinos de nivel superior. No se halló efecto LAN, pero sí uno de P600 seguido de negatividad tardía; sin diferencias significativas de posición (atributiva / predicativa) o de rasgo (género / número). Los resultados se interpretaron a favor del efecto de transfer: en el estudio anterior hay diferencia entre número y género porque los estudiantes ingleses transfieren rutinas de procesamiento de número; en los aprendientes chinos, al no existir el rasgo, no hay transferencia, y por ello no se observa ningún efecto. Por otra parte, la falta de una etapa de LAN en los estudiantes chinos reflejaría un procesamiento menos automatizado de los rasgos que en el caso de los ingleses.

Alemán Bañón, Fiorentino & Gabriele (2014) tomaron aprendientes de nivel superior (L1 inglés) y estudiaron el procesamiento de (las violaciones de) la concordancia de género y número en español L2. En primera instancia se deseaba averiguar si había diferencias generales de procesamiento de la concordancia entre género y número; visto que el inglés no posee el rasgo de género, pero en ambas lenguas se marca el número. En segundo lugar, se quería indagar sobre el impacto en el procesamiento de la concordancia de número entre adjetivo-nombre y demostrativo-nombre. Esto es porque ambas lenguas no marcan el número de igual modo: en inglés el número no se realiza en adjetivos pero sí en demostrativos en el ambiente nominal. Por último, se investigó el impacto de la distancia estructural entre el nombre y el adjetivo predicativo (a través del sintagma verbal) y controlando la distancia lineal. En lo que atañe a la primera pregunta, los resultados arrojaron efecto significativo de P600 tanto para las violaciones de género como para las de

número, en los aprendientes y en los controles nativos. También se registró que el efecto resultó ser más fuerte en el caso de las violaciones entre nombre y adjetivo. En lo que respecta a la segunda pregunta, el efecto de P600 se volvió a observar en ambos contextos de realización del número; pero siendo más robusto para nombre-adjetivo (donde el inglés no marca el número) que para nombre-demostrativo. Dicha asimetría no se encontró en el caso de las violaciones de género. Por último, se vio un efecto de P600 tanto a corta distancia (dentro del sintagma nominal) como a larga distancia (cruzando el límite del sintagma verbal). Al igual que en los nativos, el efecto resultó más robusto a corta distancia que a larga distancia, indicando que con la distancia disminuye la sensibilidad a las relaciones de concordancia. En suma, los aprendientes lograron procesamiento similar al nativo aunque uno de los rasgos no estuviera presente en su lengua materna (el género) o estuviera realizado de forma diferente (el número). El procesamiento a larga distancia también fue parecido.

MacLaughlin *et al.* (2010) argumentaron que existiría un camino progresivo de procesamiento en L2, desde una etapa dominada por el efecto N400 (relacionada con el procesamiento de la semántica y las probabilidades del input) hacia otra dominada por el P600 (procesamiento gramatical); o sea que los aprendientes van desde un aprendizaje basado en *chunks* y co-ocurrencias probabilísticas hacia otro basado en reglas de generalización. Reportan evidencia de que dicho cambio N400 > P600 se observa incluso inmediatamente, en el primer año de aprendizaje.

Se resumen los principales resultados en el Cuadro 2.3.

2.4.3. Aprendizaje de dependencias no adyacentes.

Se ha sugerido que el aprendizaje estadístico - el descubrimiento de estructura a partir de propiedades estadísticas del input - juega un rol importante para adquirir dependencias no adyacentes como la concordancia tanto en niños como en adultos (para revisiones bibliográficas, ver: Sandoval & Gómez (2013), Aslin & Newport (2012)). A partir del aprendizaje en adultos de gramáticas artificiales, se detectaron varios factores. En primer lugar, la frecuencia de co-ocurrencia. Esta cuenta las veces que dos elementos de una dependencia co-ocurren: Cuanto más alta es la frecuencia conjunta, más fácil es adquirir dicha dependencia. Otra medida relacionada es la probabilidad condicional o de transición (PT) entre dos elementos, que cuenta la frecuencia conjunta de dichos elementos relativa a uno de ellos: $P(A|B) = \frac{P(A,B)}{B}$. Por ejemplo, para los morfemas que concuerdan en *coch-es blanc-os*, la PT «hacia adelante» mide la probabilidad de encontrar el segundo elemento dado el primero: $P(A = -os|B = -es)$; mientras que si la PT es «hacia atrás», se considera el primero dado el segundo: $P(A = -es|B = -os)$. Cuanto más alta la probabilidad, más fuerte es la relación entre *A* y *B*, y más fácil resulta la dependencia (Saffran *et al.*, 1996; Pelucchi *et al.*, 2009). Estos estudios utilizan en general lenguas «en miniatura» inventadas para el experimento en cuestión.

Gómez (2002) propuso que el aprendizaje estadístico está orientado por *default* hacia dependencias adyacentes, pero que puede orientarse hacia las no adyacentes cuando la PT entre los elementos no adyacentes es mayor que con los adyacentes. Precisamente es así con el lenguaje, donde elementos funcionales pertenecientes a clases con pocos elementos se hallan separados por elementos de clases abiertas con muchos elementos. Por ejemplo, en el progresivo inglés *is eat-ing* [está comiendo], la PT entre *is* y *-ing* es más alta que entre *-ing* y la base léxica del verbo. Adultos (y niños) aprendieron estructuras del tipo *aXb* (similar a la del progresivo del inglés), donde las PTs («hacia adelante») $P(b|a) = 1$ (no adyacente) y $P(X|a), P(b|X) < 1$ (adyacente). En efecto, la precisión promedio era mayor (90%) si la clase *X* tenía muchos elementos ($X = 24$), reduciéndose así la PT entre elementos adyacentes y aumentando la diferencia entre PTs adyacentes y no. En cambio si *X* tenía muchos menos elementos ($X = 2, 6, 12$), lo cual aumentaba la PT adyacente y disminuía la diferencia entre PTs, la precisión era menor (60.5%, 66.5%, 65.5%, respectivamente).

Saffran y Romberg (2013) usaron las mismas estructuras del tipo *aXb*, pero, a diferencia de Gómez (2002), variaron las probabilidades de los elementos de *X* (que antes eran uniformes). Había una condición no adyacente, en la cual se debían detectar los patrones *a_b* no presentes en el input; y otra condición adyacente, en la que se detectaban los elementos de *X* no presentes en cada patrón *a_b*. Por ejemplo, el elemento $X = loga$ tenía alta probabilidad en el patrón *dak_tood*, baja en el patrón *vot_jic* y nula en *pel_rud*. Entonces, habiendo oído el input *loga* la PT era más alta para *dak* que para *vot*, y nula para *pel*. La condición de adyacencia consistía en determinar que *dak loga tood* era posible pero *pel loga rud* no;

ESTUDIO	DISEÑO	PARTICIPANTES	FOCO	RESULTADO
Tocowicz y MacWhinney (2005)	T; CO; ERP	Np=34; P; L1: Ing.	G, N	P600 asociado a construcciones de transfer positivo y falta de transfer
Sagarra (2007)	T; CO; TL	Np=311; P; L1: Ing.	G	TL: [-DIST] > [+DIST]; MT <> [+DIST]
Alarcón (2009)	T; CO; TR	Np=161; P; I; S; L1: Ing.	G	An: predictor en todos los niveles
Keating (2009)	T; CO; ET	Np=62; P; I; S; L1: Ing.	G	S: mayor TL en DISC en dominio local
Keating (2010)	T; CO; ET	Np=31; S; L1: Ing.	G	S: mayor TL en DISC en dominio no local detección: temprana en Nativos; tardía en S MT <> [+DIST]
Sagarra & Herschensohn (2010)	T; CO; TL; TJG	Np=196; P; I; L1: Ing.	G, N	TL en I: DISC > CONC; MT <> DISC de G TJG en I y P: N > G; CONC > DISC
Lew-Williams y Fernald (2010)	T; CO; TRO; TSF	Np= 52; I; L1: Ing.	G (Art)	palabras nuevas y conocidas: no procesan G del Art
MacLaughlin <i>et al.</i> (2010)	T; CO; ERP	P; I; S; varias L1	G, N	etapas: N400 > P600
Dowens <i>et al.</i> (2010)	T; CO; ERP	Np= 46 ; S; L1: Ing.	G, N	ERP en DISC: S = Nativos S: LAN solo en Art / [-DIST]; P600: N > G
Dowens <i>et al.</i> (2011)	T; Co; ERP	Np= 26; S; L1: Chino	G, N	S: sin LAN; P600: N = G.
Foote (2011)	T; CO; TL	Np=60 ; BTe ; BTa; L1: Ing.	G	TL: DISC > CONC; [-DIST] > [+DIST] detección: temprana en BTe; tardía en BTa
Bond <i>et al.</i> (2011)	T; CO; ERP	Np=32 ; P; L1: Ing.	G, N	P600 asociado a construcciones de transfer positivo transfer negativo y falta de transfer
Foucart y Frenck-Mestre (2011)	T; CO; ERP	Np= 32 ; S; L1: Al.; L2: Fr.	G	P600 asociado a transfer positivo y falta de transfer
Grüter <i>et al.</i> (2012)	T; CO; PRO; TRO; TSF; SFE; TDF	Np= 38; S; L1: Ing.	G (Art)	procesan G del Art solo en palabras nuevas ErrAsig (mayor en MNT y Fem) > ErrConc
Dussias <i>et al.</i> (2013)	T; CO; ET; TSF	Np=50 ; S; P; L1: Ing. Ita	G (Art)	S: procesan G del Art en palabras conocidas P(Ita): procesan G del Art solo en el Fem
Sagarra & Herschensohn (2013)	T; CO; TL; TJG	Np=196 ; P; I; L1: Ing.	G	TL en I: DISC > CONC; An > InAn; MT <> DISC TJG en I y P: CONC > DISC; I > P; InAn > An
Alemán Bañón, Fiorentino & Gabriele (2014)	T; CO; ERP	Np= 26; S; L1: Ing.	G, N	P600 asociado a ErrConc de: G, N; N en nombre-Adj, nombre-Dem.; CONC [+/- DIST]

Leyenda: «TDF»: Tarea de descripción de figuras; «TR»: tiempos de reacción; «PR»: Producción; «CO»: comprensión; «TL»: tiempos de lectura; «TSF»: Tarea de selección de figuras; «L»: estudio longitudinal; «TRO»: Tiempos de reacción ocular; «ET»: Eye Tracking; «T»: estudio transversal; «TJG»: Tarea de juicios de gramaticalidad; «SFE»: Tarea de selección de figuras a partir de oraciones con elipsis nominal; «G»: género; «N»: número; «Np»: número de participantes; «P»: nivel inicial; «I»: nivel intermedio; «S»: nivel superior; «Ing»: inglés; «Fr»: francés; «Al.»: alemán; «BTe»: bilingües tempranos; «BTa»: bilingües tardíos; «Fem»: femenino; «MNT»: nombres de morfología no transparente; «MT»: memoria de trabajo; «Art»: artículos; «Adj»: adjetivos; «An»: animados; «InAn»: inanimados; «ErrAsig»: errores de asignación; «ErrConc»: errores de concordancia; «Dem»: demostrativo; «+/- DIST»: mayor / menor distancia; «CONC»: concordancia; «DISC»: discordancia (error); «<>»: correlaciona con.

Cuadro 2.3: Estudios *online*

en cambio, la de no adyacencia, en ver que *dak loga tood* era posible pero *pel loga tood* no, ya que el patrón *pel_tood* no existía. Además, se hizo de manera que las PTs adyacentes fueran más fuertes o más débiles. Por ejemplo, *loga* puede ser cuatro veces más frecuente como ítem de probabilidad alta en un patrón respecto a su ocurrencia de probabilidad baja en otro patrón; reduciendo dicho ratio, disminuye su PT en el caso de alta ocurrencia, causando que la diferencia en las PTs altas y bajas sean menores. Hallaron que los sujetos resultaban sensibles a ambos tipos de dependencias aún con poca exposición al input, aunque se necesitaba más exposición para mejorar la precisión. O sea, no era el caso de que las dependencias adyacentes fueran *default* y que los aprendientes luego cambiaran a las no adyacentes porque habían detectado alta variabilidad (PT baja) entre elementos adyacentes; sino que ambas se adquieren en simultáneo. Se encontraron diferencias según el orden de administración del test (primero con estructuras adyacentes y después con no adyacentes o viceversa). Cuando la condición adyacente se testeaba primero y contenía PTs fuertes, la precisión era más alta con estructuras adyacentes que con no adyacentes; mientras que, con PTs débiles no había efecto. Es decir que el aprendizaje de las dependencias no adyacentes está afectado por la fuerza de las PTs de dependencias adyacentes.

En un estudio de segmentación de lenguaje (Newport y Aslin, 2004) los sujetos debían discriminar palabras trisilábicas del tipo *aXb* a partir de un continuo de input oral. Las PTs no adyacentes entre el primer y tercer elemento *a_b* dentro de las palabras eran mayores (PT = 1) que entre el último de una palabra y el inicial de la siguiente; o bien entre el elemento intermedio y los otros (*aX, Xb*), siempre con PTs < 0.5. Los patrones no adyacentes que conformaban las palabras eran: (a) silábicos, por ejemplo: *po_ga*, donde *X = ki,gu*, en palabras como *pokigu, poguki*; (b) consonánticos, por ejemplo *d_k_b*, donde *X = o,u,e*; en palabras como *dokube*; (c) vocálico, por ejemplo: *_a_u_e*, donde *X = p,g,t*; en palabras como *pagute*. Ahora bien, los aprendientes aprendieron a discriminar palabras correctamente en el caso de (b) y (c), pero no en (a), aun habiendo iguales PTs en los tres tipos de estructuras. Para explicar los resultados se sugirió que un factor de similaridad perceptiva entre los elementos que forman dependencias a larga distancia (aquí consonantes o vocales) ayuda en su adquisición.

Lany *et al.* (2007) encontraron que la experiencia previa de aprendizaje con una estructura del tipo *aX* de dependencias adyacentes (por ejemplo: *a = alt; X = tam-oo*; donde *a* predice palabras de clase *X* terminadas en *-oo*) aumentaba luego la precisión en estructuras no adyacentes tipo *acX*, que contaban con un elemento *c* intermedio variable (por ejemplo: *alt hes /kaf/ sij tam-oo*). En pocas palabras, el aprendizaje se facilitaba si se realizaba por etapas de menor a mayor complejidad.

Pacton y Perruchet (2008) propusieron que el factor que guía el aprendizaje de uno u otro tipo de dependencia es la atención selectiva (consciente). Les pidieron a los participantes que encontrarán un determinado número en una secuencia de dígitos y que sustrajeran o bien los dos dígitos que seguían al número en cuestión (dependencia adyacente) o bien los dígitos que lo rodeaban (dependencia no adyacente). Cada secuencia poseía ambos tipos de dependencias. Sin embargo, aquellos participantes que dirigieron la atención a las restas de números subsiguientes, solo aprendieron dependencias adyacentes; en cambio, los que se concentraron en las restas de números que rodeaban, únicamente aprendieron las no adyacentes. O sea que la atención sobre un tipo de dependencia ejerció un efecto de bloqueo para el aprendizaje del otro tipo.

Ahora bien, hasta ahora, los estudios trataron dependencias no adyacentes deterministas (donde $P(A|B) = 1$). Sin embargo, en el lenguaje tienden a ser probabilísticas, en las que más de un elemento puede ocurrir dado un primer elemento. Por ejemplo, en el progresivo *está cant-ando / com-iendo* la dependencia se establece entre el verbo y los morfemas *-ando* (si la raíz pertenece a la clase de conjugación en *-ar*) o bien *-iendo* (si la conjugación es en *-er* o *-ir*); es decir que, por ejemplo, $P(-ando|está) < 1$. Van der Bos *et al.* (2012) expusieron a los participantes a secuencias de 3 palabras del tipo *aXb* en dos condiciones: (a) dependencias no adyacentes entre *a* y *b* deterministas: por ejemplo el patrón *dak_tood* donde $P(b = tood|a = dak) = 1$; (b) probabilísticas: por ejemplo los patrones *dak_tood* y *dak_rud*, donde $P(b = tood|a = dak) = 0,5$. El elemento del medio era variable y pertenecía a la clase *X* de 24 elementos. La PT de los elementos adyacentes era constante: $P(X|a) = P(b|X) = 0,4$. La precisión de las dependencias deterministas resultó significativamente mayor que las probabilísticas y arriba del 50%; en cambio las probabilísticas obtuvieron una corrección promedio debajo de 50%, es decir no fueron aprendidas (los participantes respondían aleatoriamente). Luego se incluyeron en el input pistas fonológicas creando palabras terminadas en *-o, -in/im, -a, -er* y haciendo que concordaran entre el primer y tercer elemento (*pravin_wadim, nilbo_tisso*). Como

resultado, se aprendieron ambos tipos de dependencias sin diferencias significativas en la precisión entre ellas. En el caso de las dependencias probabilísticas la diferencia entre las PTs adyacentes y no adyacentes era inicialmente muy baja (0.1), por lo cual se necesitaban pistas fonológicas adicionales para aumentar las PTs no adyacentes que posibilitaran la formación de clases y se pudieran aprender. En suma, aprender dependencias no adyacentes probabilísticas es posible con la ayuda de «pistas» [cues] adicionales. Las lenguas ofrecen pistas fonológicas, semánticas y morfológicas para facilitar que el aprendiente pueda reducir el espacio de hipótesis de aprendizaje, y consecuentemente, disminuir la cantidad de input necesaria para encontrar una solución.

Van Heugten & Johnson (2010) investigaron la adquisición del diminutivo y el plural en holandés L1 y llevaron a cabo un estudio estadístico de corpus. Las dependencias no adyacentes ($a-X-b$) eran entre el artículo definido singular y el marcador de diminutivo, como en *het hondje* [el perrito]; y entre el artículo definido plural y el marcador de plural, como en *de honden* [los perros]. A los 17 meses los niños no habían adquirido ninguna dependencia; pero a los 24, habían adquirido el diminutivo pero no el plural. Entonces se analizó el input que los padres dirigían a los niños con el fin de verificar si las estadísticas distribucionales de la dependencia con el diminutivo favorecían la adquisición. Se consideraron las combinaciones entre artículos y las formas *-je* y *-en* de las dependencias. O sea, por ejemplo, se contó tanto la secuencia «gramatical» *de-X-en* (plural) como la «agramatical» *het-X-en*, como en *het eten* [la comida], donde la co-ocurrencia no involucra el plural. Las medidas para las secuencias «gramaticales» y «agramaticales» fueron: (a) distancia lineal en sílabas entre los elementos de las dependencias; (b) las probabilidades de transición «hacia adelante» y «hacia atrás»; (c) la plausibilidad estadística de los marcadores como predictores de diminutivo o plural, o sea, qué tan frecuente el marcador *-je* funcionaba efectivamente como diminutivo, y *-en* como plural; (d) la frecuencia conjunta de los elementos de las dependencias no adyacentes; por ejemplo, la frecuencia de la secuencia «gramatical» *het-X-je* y de la «agramatical» *de-X-je*, cuyo *odds ratio* medía la fuerza de la discrepancia entre ambas; (e) la variabilidad del material interviniente *-X-* medido como ratio de número de sílabas únicas sobre no únicas. Con excepción de la última medida, todas indicaban que el diminutivo contaba con mayor regularidad estadística que el plural; favoreciéndose así su adquisición temprana.

En resumen, la investigación en dependencias no adyacentes (deterministas y probabilísticas) ha echado luz sobre los factores que intervienen en su adquisición: las frecuencias conjuntas; las probabilidades de transición (diferencias entre las PTs de los elementos adyacentes y no adyacentes); la variabilidad de los elementos intervinientes; la similaridad perceptiva; la atención; la ayuda de «pistas» (fonológicas, morfológicas, semánticas); la regularidad estadística del input. Ambos tipos de dependencias se adquieren en modo simultáneo, pero la fuerza de la PT de las dependencias adyacentes incide en las no adyacentes. Asimismo, favorece la adquisición el aprendizaje por etapas: *adyacentes > no adyacentes*.

2.4.4. Resumen de los hallazgos.

La literatura de adquisición, más abundante para (la concordancia) de género, que la de número ha arrojado los siguientes resultados:

- La adquisición del plural tiende a seguir las etapas: plural nulo > plural en /-s/ > plural en /-es/ [Bruhn de Garavito (2008)].
- La concordancia de número es más fácil de adquirir que la de género ya que, en general, los errores en el primero tienden a ser menos que en el segundo. Si bien el nivel de competencia hace que los errores disminuyan, los de género tienden a persistir incluso luego de muchos años de práctica de producción oral [Liceras, Díaz & Mongeon (2000), Francheschina (2001), White *et al.* (2004)].
- Se observa un efecto de marcado. La forma de *default* tiende a ser no marcada, o sea, el singular para el número y el masculino para el género (no obstante también existe alguna evidencia del femenino como *default*) [Finnemann (1992); Fernandez-Garcia (1999); Bruhn de Garavito & White (2002); White *et al.* (2004); Montrul *et al.* (2008); Alarcón (2011); McCarthy (2008)]. Esto se verifica en cualquier nivel de competencia; tanto en bilingües tempranos como tardíos; y en tareas de comprensión y producción.

- La concordancia plural del cuantificador *mucho* parece más fácil de adquirir que la de *bastante* y *demasiado* [Echevarría y Prévost (2004)].
- La concordancia de género y número del artículo es más fácil de adquirir que la del adjetivo. Ésto parece ser así para cualquier nivel de competencia; en bilingües tempranos y tardíos; y tanto en producción como en procesamiento [Fernandez-Garcia (1999); White et al. (2004); Montrul et al. (2008)]; Alarcón (2011); Dowens et al. (2010)].
- En cuanto al género, resulta más fácil producir y procesar la concordancia : (a) de masculino respecto al femenino (reflejando el hecho de que las formas de masculino se utilizan en contextos femeninos, o sea, como *defaults*) [Fernandez-Garcia (1999); Bruhn de Garabito & White (2002); White et al. (2004); Montrul et al. (2008); Alarcón (2011); McCarthy (2008)]; (b) con controladores de morfología transparente (-o/-a como en *vaso*, *maestra*) respecto a los menos transparentes (en -e, como en *el puente*, *la suerte*; en consonante: *el camión*, *la canción*; u opuestos, como en *la mano*), éstos últimos en el orden de facilidad *-cons* > *-e* > *opuestos* [Fernandez-Garcia (1999); Montrul et al. (2008); Alarcón (2011)]; (c) con controladores inanimados (como en *hospital*) respecto de los animados (en correspondencia con el sexo biológico, como en *doctor/a*) [Sagarra & Herschensohn (2013); aunque Alarcón (2009) encontró el efecto contrario en núcleos de SN complejos]; (d) En SN complejos del tipo *N1 de N2*, cuando el género de N1 coincide con el N2 [Foote, 2015]. Además, los errores de asignación resultan más altos que los de concordancia, especialmente con controladores de morfología no transparente (*-e*, *-cons*) y con los femeninos [Montrul et al. (2008); Alarcón (2011); Grüter et al. (2012); Montrul et al., 2013].
- El aumento de distancia causa que disminuya la sensibilidad a las violaciones de género y que la concordancia se procese más lentamente en el dominio no local. Es decir, la concordancia de género se procesa mejor en el dominio local que en el no local [Sagarra (2007), Lichtman (2009), Keating (2009, 2010), Foote (2011), Dowens et al. (2010)]. Se ha propuesto que dicho efecto de distancia se relaciona con la capacidad de memoria de trabajo, ya que el aprendiente debe mantener el valor del rasgo del controlador (nominal) en la memoria para concordarlo luego con el del objetivo (adjetivo) a larga distancia. Sagarra (2007) y Keating (2010) hallaron correlaciones positivas entre distancia y memoria de trabajo; en cambio, en Foote (2011) no hubo evidencia de ello. Claramente se necesita más investigación al respecto.
- En el sintagma nominal, los aprendientes avanzados (al igual que los nativos) pueden utilizar la información de género presente en el adjetivo para acelerar el acceso al léxico mental e identificar con más celeridad referentes, procesando así más rápidamente el nombre. Por otro lado, aprendientes de otros niveles no aprovechan dicha información y esperan a oír el nombre para procesarlo. Esto sucede tanto para referentes conocidos como para los nuevos [Lew-Williams y Fernald (2010); Grüter et al. (2012); Dussias et al. (2013)]. Sin embargo, el factor de nivel de competencia interactúa con el *transfer*, ya que Dussias et al. (2013) detectaron también el efecto anterior de anticipación para el femenino en aprendientes italianos de nivel inicial (que pueden transferir rutinas de procesamiento de una lengua a otra).
- Existe evidencia de que el procesamiento del género implica más esfuerzo cognitivo que el de número [Sagarra & Herschensohn (2010)]. No obstante, es plausible que las diferencias encontradas dependan también de la similaridad entre las lenguas, y de efectos de *transfer*. Si no existe el *transfer* (porque la L1 no posee un rasgo, como el caso del género en inglés) o si es positivo (porque el rasgo está implementado de manera similar en L1 y L2, como el caso del presente progresivo en inglés y español), entonces los aprendientes de cualquier nivel de competencia logran procesamiento implícito (ERPs). O sea, se logra detectar implícitamente violaciones de gramaticalidad aunque explícitamente (por ej., en juicios de gramaticalidad) no se las reconozca [Tokowicz & MacWhinney (2005)]. En cambio, si el rasgo existe en ambas lenguas pero se encuentra implementado en las construcciones de modo diferente (como el caso del número, que en el SN está codificado en los modificadores en español pero no en inglés), se ha hallado tanto evidencia de procesamiento implícito igualmente

[Bond *et al.* (2011)] como de su ausencia por efecto del transfer negativo [Foucart & Frenck-Mestre (2011)]; lo cual amerita más investigación. Dada suficiente competencia gramatical, el procesamiento del género y del número resulta similar al de los nativos; sin importar si el rasgo está presente o no en la L1 [Alemán Bañón, Fiorentino & Gabriele, 2014]. A pesar de ello, el hecho de que se encuentren diferencias de «profundidad» o «esfuerzo» de procesamiento entre dichos rasgos sí depende de si el rasgo se halla en la L1; y por lo tanto, que los aprendientes puedan procesarlo «mejor» en la L2 reclutando rutinas ya presentes su L1 [Dowens *et al.* (2010, 2011)].

- Por último, se han aislado diversos factores relacionados con la facilitación del aprendizaje de dependencias no adyacentes (deterministas y probabilísticas): las frecuencias conjuntas y las probabilidades de transición (PT) [Saffran *et al.* (1996); Pelucchi *et al.* (2009)]; especialmente las diferencias entre las PTs de los elementos adyacentes y no adyacentes; la variabilidad de los elementos intervinientes [Gómez (2002)]; la similaridad perceptiva [Newport & Aslin (2004)]; la atención [Pacton & Perruchet (2008)]; la ayuda de «pistas» (fonológicas, morfológicas, semánticas) [Van der Bos *et al.* (2012)]; la regularidad estadística del input [Van Heugten & Johnson (2010)]. Ambos tipos de dependencias se adquieren en modo simultáneo, pero la fuerza de la PT de las dependencias adyacentes incide en las no adyacentes [Saffran & Romberg (2013)]. Asimismo, favorece la adquisición el aprendizaje por etapas: *adyacentes* > *no adyacentes* [Lany *et al.* (2007)].

Capítulo 3

Sistemas Dinámicos Complejos en L2.

Este capítulo presenta la teoría de los sistemas dinámicos complejos y su aplicación a la adquisición de lenguas segundas. En primera instancia se mencionan las nociones de «memoria» y «atención». Luego se introduce el modelo de producción bilingüe¹. A continuación, se comentan algunas teorías de adquisición de lenguas segundas. Seguidamente, se introduce la teoría de sistemas dinámicos complejos y su aplicación a la adquisición de L2. Después, se compara dicha perspectiva con los enfoques introducidos previamente. Por último, se muestra un modelo propuesto que trata la concordancia como sistema dinámico en L1; y se plantea, para L2, considerar al error de concordancia como atractor. Se ilustra la estrategia de análisis de los capítulos tesis sub-siguientes.

3.1. Atención y Memoria.

Si representáramos la consciencia como un teatro (Baars, 1997, 2002; Dehaene & Naccache, 2001), el escenario correspondería a la memoria de trabajo, es decir, el sistema de registro inmediato en que, bajo control voluntario, nos hablamos interiormente, visualizamos lugares, personas y planificamos la acción. La atención actuaría como la luz de un reflector que nos hace conscientes de los elementos que ilumina (p. ej., el significado de una oración que leemos). Este sector iluminado está rodeado por la penumbra; apenas tenemos consciencia de los elementos que allí se mueven (p. ej., la tipografía de la letra). Así, como en un escenario, la acción es limitada; sólo somos conscientes de una pequeña parte de la información procesada en el cerebro. La consciencia actúa como un espacio de coordinación e intercambio de la información producida por sistemas implícitos distribuidos (descentralizados). Estos representan a la ‘audiencia’, con un rol clave de la memoria a largo plazo (MLP), donde se representan los conceptos necesarios para entender la oración. Detrás del escenario está el ‘contexto’, los factores inconscientes que rodean la experiencia, como el color de la habitación en la que leemos.

¿Es necesaria la atención para aprender una lengua? Sin duda, pero ¿a qué nivel? Schmidt (2001, 2010) propuso la hipótesis del ‘notar’ (*noticing hypothesis*), argumentando que el aprendizaje se logra cuando detectamos algo y además nos ‘damos cuenta’ de que nos fijamos (*awareness*), es decir, ubicamos el input en el foco consciente de atención. Así, la atención a nivel de la detección no es suficiente; debe acompañársela de la experiencia subjetiva del ‘notar algo’ conscientemente. Van Patten (2015) afirma que, cuando procesan input de L2, los estudiantes orientan la atención hacia el contenido. Es decir, procesan primero palabras de contenido y luego marcadores gramaticales. Por ejemplo, al leer la oración mañana voy a ir al cine, se procesarán primero los lexemas *mañana*, *ir* y *cine* y se tenderá a ignorar el marcador de futuro *voy a*, ya que codifica información similar al adverbio (decir *mañana voy al cine* es igual). Por esta razón, para favorecer el aprendizaje de la perífrasis de futuro, el docente debe orientar la atención hacia la forma [*focus on form*]. Los alumnos no prestan atención selectiva ni se fijan en rasgos del input infrecuentes, comunicativamente redundantes, similares a su lengua materna, o sin saliencia perceptiva. Por ello, los programas de inmersión lingüística basados en actividades comunicativas solamente orientadas al significado generan beneficios en la fluidez y la comprensión pero no en la precisión. El profesor debe, pues, guiar la atención para enfocarla

¹Para este tema y el anterior se siguió a Marafioti (2016)

en la forma mientras los alumnos realizan actividades significativas y comunicativas (Long y Robinson, 1998).

Usamos la memoria para aprender nueva información, organizarla y recuperarla cuando queremos recordar algo. El modelo estandarizado de procesamiento en la memoria propuesto por Atkinson y Shiffrin (1968) comprende tres tipos de (almacenes de) memoria que procesan la información del mundo externo y operan de manera específica respecto de la capacidad de almacenaje, tipo de código que reconocen y velocidad de deterioro (olvido). Primero hay un registro sensorial de modalidad específica con deterioro inmediato (3 segundos en promedio), con capacidad ilimitada (p. ej., visual o icónica, auditiva o eólica, táctil o háptica, gustativa, olfativa), que extrae las características físicas de los estímulos. En una segunda instancia, un registro de corto plazo (MCP) toma información del registro sensorial a partir de procesos de atención/detección con una capacidad de procesar 7 ± 2 unidades de información y que olvida, en promedio, en 30 segundos (si no media la repetición). Por último, habría un registro de largo plazo (MLP), con capacidad y duración ilimitadas, en el que se introducen y/o refuerzan huellas de memoria automáticamente a través del proceso de repetición (cuanto más tiempo un ítem se mantiene en la MCP, más probable es su transferencia a la MLP). La MLP se divide en dos sistemas de memoria. Por un lado está la memoria declarativa (saber “qué”), que incluye información semántica (conceptos y relaciones sobre el lenguaje y el mundo) y episódica (experiencias autobiográficas vividas en un tiempo y lugar específicos). Esta información puede hacerse intencionalmente explícita, o sea consciente, a través de la recuperación voluntaria verbalizada. Por el otro tenemos la memoria procedimental o procedural (saber “cómo”), que opera con relaciones predecibles de estímulo-respuesta, destrezas y hábitos automáticos, implícitos e inconscientes. En pocas palabras, en este modelo los registros de memoria forman la estructura básica y los procesos (atención, reiteración, extracción de información de la MLP, etc.) controlan el flujo de información.

3.2. El modelo de producción bilingüe.

Kormos (2006, 2011) propuso un modelo de producción del lenguaje en L2 basado en la teoría modular de Levelt (1989; levelt et al., 1999). Consta de tres módulos especializados: el conceptualizador, el formulador y el articulador. La información circula entre ellos por medio de activación extendida² (*spread activation*) y sin posibilidad de flujos de información ‘hacia atrás’ (*feedback*). Entonces, el procesamiento funciona allí serialmente, aunque de modo incremental³. La memoria a largo plazo incluye los siguientes subcomponentes relacionados con la producción del habla: (a) un registro de memoria episódica; (b) un registro de memoria semántica organizado como un lexicón jerárquico de tres niveles: conceptos (información semántica), lemas (información morfo-sintáctica) y lexemas (información fonológica); (c) el silabario: registro de los movimientos articulatorios automáticos para producir sílabas; (d) un registro declarativo de reglas de L2 no automatizadas (las que sí lo están forman ya parte de los módulos de producción). En las etapas de la producción del habla las memorias episódica y semántica se relacionan y producen la coactivación mutua de conceptos y experiencias. En la memoria semántica (lexicón) se produce un flujo de información top-down del nivel conceptual al léxico, pasando por el de los lemas. Excepto el registro de reglas, el resto de las memorias está compartida (en mayor o menor medida) entre la L1 y la L2. El modelo se esquematiza en la Figura que sigue.

²Mecanismo por el cual los nodos de una red, al superar un umbral, diseminan ‘energía’ a otros nodos activándolos en menor medida (p. ej., la activación de «cuchillo» puede causar alguna activación en «cortar»).

³La incrementalidad posibilita que un módulo comience a procesar aunque el anterior no haya terminado (p. ej., una vez que la primera sílaba de una palabra se codifica, su articulación ya comienza en el articulador). Si bien esto implica un cierto procesamiento en paralelo, también se asume que en un nivel bajo de automatización se requiere del control consciente de la atención. Entonces, el procesamiento funciona en ese caso solamente de forma serial.

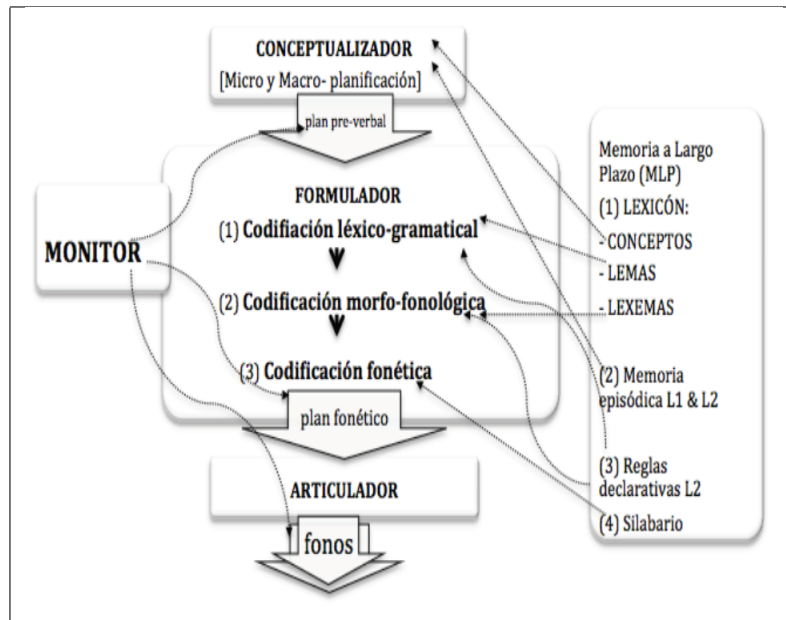


Figura 3.1: El modelo de producción bilingüe (adaptado de Kormos, 2011)

Conceptualización. El módulo conceptualizador planifica el mensaje a codificarse, eligiendo los conceptos relevantes y el lenguaje de expresión. Un concepto es visto como una red interconectada de huellas de memoria (*memory traces*). Esta posición permite que los conceptos entre la L1 y la L2 sean idénticos, compartidos o estén completamente separados. En un nivel bajo de competencia, la L1 y la L2 compartirán un mismo concepto, pero, avanzando, los alumnos enriquecerán sus representaciones conceptuales. También se etiquetan los conceptos con el lenguaje de formulación ya que, por ejemplo, quizás un estudiante no sepa una palabra y decida, en lugar de optar por un circunloquio, usar el término de su L1. Dado que el habla no siempre es creativa, existe la posibilidad de que se elijan ‘macro-conceptos’ que apunten a la elección de chunks (DESPEDIDA PARA ENFERMOS > «que te mejores»). El ‘plan pre-verbal’ constituye el resultado de dos fases de conceptualización. Primero se realiza una planificación a nivel macro estableciendo el acto de habla, seleccionando los conceptos y organizándolos en un orden serial. La segunda fase funciona a nivel micro y contiene indicaciones sobre el material lingüístico a seleccionarse en el formulador. Involucra factores como la perspectiva –p. ej., el uso de la voz pasiva o de relaciones espaciales: ¿digo «el libro está delante del ordenador» o «el ordenador está detrás del libro»?; la accesibilidad (si ya ha habido mención previa en el discurso, lo cual implica por ejemplo la selección posterior de «un niño» / «el niño» / «éste»); qué partes del mensaje constituyen información nueva o vieja; y otros tales como la estructura argumental –p. ej., complementos opcionales y obligatorios, roles temáticos (agente, paciente, etc.), modalidad (declarativa, interrogativa, etc.), registro (formal o informal) y, más generalmente, cualquier información que sea necesaria para convertir el significado en material lingüístico.

Formulación. El módulo del formulador selecciona los lemas (rasgos morfo-sintácticos) y los lexemas (palabras abstractas únicas o secuencias de ellas como en los chunks) que corresponden a las especificaciones del plan pre-verbal. Dicho plan envía activaciones a los lemas tanto de la L1 como de la L2 y ambos compiten, siendo el ganador aquel cuyos rasgos se adecuen a todas las especificaciones conceptuales. Los lexemas y los lemas están organizados también aquí en una red con conexiones intra L1/L2 o entre lexemas/lemas de L1 y los de la L2. Los ítems más frecuentes forman nodos centrales (*hubs*) con alta conectividad con otros ítems. En cambio, los menos frecuentes son periféricos en la red. El aprendizaje modifica la ‘estrechez’ en las conexiones: en los principiantes habrá más fuerza entre la conexión de las palabras de la L1 y la L2 que entre las de la L2 misma. Además, las conexiones pueden ser asimétricas, reflejando el vocabulario pasivo (comprender pero no poder seleccionar un ítem para producirlo). Dos mecanismos son importantes aquí. En una primera fase, la activación de la información morfo-sintáctica del lema (género⁴,

⁴El género se considera un rasgo del lema. En cambio, el número se computa online, por ejemplo, usando compatibilidad entre estructuras de rasgos (Jeschaniak y Levelt, 1994).

estructura argumental, otros rasgos como [+contable]). Luego, el uso de procedimientos de construcción morfo-sintáctica de sintagmas y oraciones. En la L1, el primer mecanismo utiliza la memoria declarativa, mientras que el segundo emplea las reglas de formación sintácticas y morfológicas de la memoria procedural. Los estudiantes avanzados utilizan lemas con información específica de la L2; en cambio, en los principiantes, el lema activa información de la L1 (por eso, en lugar de decir *enter the room*, los hispanohablantes dicen **enter into the room*, activando la estructura argumental del verbo entrar). Los estudiantes pueden ya haber automatizado reglas (que formen parte del formulador) o usar otras de modo consciente y menos automático (que están contenidas en el depósito declarativo de reglas), o bien no conocerlas del todo. En este último caso usarán alguna estrategia de compensación (usar la regla de la L1 o simple yuxtaposición léxica de conceptos).

Codificación fonológica y articulación. El formulador, a partir de los lexemas seleccionados, activa su forma fonológica y otros parámetros como la entonación, el acento y el volumen. Además, usa el silabario para identificar los programas de articulación de sílabas pertinentes. Se asume que el silabario es común a la L1 y la L2. Si bien los principiantes seleccionan programas de la L1, los más avanzados cuentan con programas de articulación propios de la L2. Los fonemas se representan en una única red en el nivel léxico con representaciones específicas para fonemas diferentes de la L2 (excepto en principiantes). El producto final del formulador es un plan fonético o ‘habla interior’. Finalmente, el módulo articulador ejecuta la información del plan fonético para producir sonidos (fonos).

El monitor. Por último, se postula un proceso de monitoreo que funciona igual para la L1 y para la L2 y que utiliza el sistema de comprensión del lenguaje para inspeccionar la propia producción. El monitor controla el producto de los tres módulos. A nivel del conceptualizador, chequea la adecuación del plan preverbal a las intenciones comunicativas (algún aspecto podría no adecuarse a la situación comunicativa, p. ej., el rasgo conceptual [-formal]). A nivel del formulador, controla aspectos de codificación lingüística (sintáctico-morfo-fonológica) del plan fonético (p. ej., la selección incorrecta de alguna palabra antes de ser pronunciada). Finalmente, a nivel del articulador, inspecciona el plan de articulación fónica (errores de pronunciación). Monitorear los productos de la L1 o la L2 siempre requiere atención. La diferencia radica en que los módulos procesan automáticamente en la L1 en cualquier nivel; en cambio, en la L2 los estudiantes dirigirán la atención a aspectos de un nivel dado de producción y esto quitará recursos al monitoreo de otros niveles. Por ello, deberán decidir qué priorizar: significado vs. forma, léxico vs. gramática, corrección vs. fluidez o pronunciación, etc. Kormos (2006) distingue tres tipos de movimientos de auto-corrección (*selfrepair*): (1) auto-corrección mediante nueva información (*D-repairs*): se corrigen errores en la etapa de conceptualización del mensaje, ya sea por una mala organización del contenido o por errores en el plan preverbal (p. ej., tiene ehh tenemos que cambiar todo); (2) auto-corrección de la ‘propiedad’ (*appropriacy*) del mensaje (*A-repairs*): el hablante adopta el plan preverbal original pero de un modo diferente, ya que en la codificación se cometieron errores de inexactitud (p. ej., tres ehh no cuatro libros), ambigüedad (p. ej., en la plaza hay muchos bancos ehh bancos para sacar dinero), terminología incorrecta (p. ej., hicieron un pedido ehhh una factura), adecuación pragmática a una situación (p. ej., [NEGOCIACIÓN COMERCIAL] entonces hay un trato ehh trato hecho), de estilo (p. ej., para no repetir palabras: hay algunos problemas; hay ehh existen también...); (3) auto-corrección de errores (léxicos, sintácticos, morfológicos, fonológicos) en la etapa de formulación (*E-repairs*) (p. ej., tú dice ehh dices).

El modelo de producción bilingüe delineado posee claras implicancias para el desarrollo de la L2: el aumento de la competencia lleva a una complejización paulatina del plan preverbal, lo cual implica la introducción y selección de información (léxica, morfológica y sintáctica) más elaborada en el formulador.

3.3. Teorías de adquisición de L2.

3.3.1. Gramática universal.

El debate en L2 dentro de la tradición generativa ha estado signado por la naturaleza de la representación cognitiva de la competencia en la segunda lengua y el rol del *transfer* (White, 2015; Rothman & Pascual y Cabo, 2014). Desde el punto de vista del nativismo, se supone que si los aprendientes poseen conocimiento de gramaticalidad y ambigüedad sobre estructura lingüística que no deriva del input que recibieron (o de su L1), entonces hay un conocimiento o competencia pre-existente (independiente de la experiencia). Dicho conocimiento guía la adquisición de la L2 y es *implícito*. Está representado en una gramática universal [GU], que es altamente abstracta y está codificada en término de rasgos, categorías y operadores (MERGE, MOVE, AGREE; ver cap. 1). Asimismo, la GU trata con elementos simbólicos discretos y se halla modularizada, siendo diferente de otros aspectos de la cognición. Se supone que la *inter-lengua*⁵ está limitada por la GU, lo cual quiere decir que los errores que se producen no van en contra de las delimitaciones que impone la gramática universal. Por ejemplo, supóngase el caso del orden *N-Adj* en inglés y en español: (a) Ing. *The (*soldiers) brave soldiers make large sacrifices*; (b) esp. *Los (soldados) valientes soldados sacrifican mucho*. En español existe la posición pre-nominal, inexistente en inglés. Esto se debe a la posibilidad de «mover hacia arriba» el sustantivo para chequear un rasgo en la categoría funcional *SNum*. En inglés no se puede realizar este movimiento. Esto sucede porque la GU contiene el rasgo *num*, que se parametriza como «fuerte» en español y como «débil» en inglés. En español la posición pre-nominal del adjetivo se interpreta como «para todos los miembros del conjunto»; en cambio, la posición post-nominal tiene la interpretación de «para algún miembro del conjunto». En inglés, al no recurrir al orden de palabras, dicha interpretación es ambigua. Si la GU guiara el aprendizaje, entonces los aprendientes podrían reparametrizar el rasgo de movimiento o bien parametrizar algún otro rasgo que no exista en la L1. Por otro lado, un error del tipo *soldados* los valientes sacrifican mucho* sería muy improbable porque la UG no permite el movimiento a posición pre-determinante (sin topicalizar). La *inter-lengua* es pues sistemática porque sigue las restricciones impuestas por la GU.

Ahora bien, en el caso de la lengua materna, el estado cognitivo inicial del niño previo a recibir input del lenguaje es la propia GU. Pero en el caso de la L2, los adultos ya han adquirido su L1 por completo. Entonces el debate giró en torno a la naturaleza del estado inicial en la L2: ¿existe todavía acceso a la GU para aprender una segunda lengua?, ¿Qué rasgos se pueden transferir de la L1 a la L2 (*transfer*)?. Una de las posturas consiste en plantear que no hay acceso alguno (Clahsen & Muysken, 1986). En consecuencia, no sería posible reparametrizar los valores de los rasgos de la GU y ésta tampoco podría guiar el aprendizaje; el cual se realizaría vía comprobación de hipótesis y mecanismos estadísticos. En este caso la GU sufriría un proceso de maduración, con un periodo crítico más allá del cual se volvería inexpugnable. Aquí el estado inicial es la L2. Una postura intermedia es sostener que la GU es accesible de forma parcial [*Failed Functional Features Hypothesis* (Hawkins & Chan, 1997)]. Allí se establece solamente acceso a los rasgos de la L1, pudiéndose reparametrizar solamente éstos y no valores de rasgos presentes en la L2 y no en la L1. La implicancia de ello es que los aprendientes anglo-parlantes de español no podrían adquirir la categoría funcional de género pero sí reparametrizar el rasgo de número de *débil* a *fuerte*. También se ha propuesto el acceso total a la GU [*Full Access / Full Transfer* (White, 1989)]. En este caso sería posible modificar valores de todos los rasgos de la GU, aunque no estén en la L1; con lo cual la GU no estaría sujeta a maduración. El estado inicial es pues la L1. Esto significa que el aprendiente va a transferir rasgos de la L1 a la L2 (*full transfer*), pero ya que goza de acceso a la GU (*full access*), podrá luego reparametrizarlos. En el caso de la concordancia de número, al principio no se producirían concordancias correctas con el determinante («lo niños» [*the boys*]) porque el rasgo está parametrizado como *débil*, y luego se lo podría reparametrizar como *fuerte*. Así el rasgo puede «subir» hasta *Snum* para chequearse y realizar la operación AGREE con el rasgo *num* en el especificador del *SD*. En el caso del género resultaría incluso más complicado porque es preciso adquirir una nueva categoría funcional. Con todo, la morfología resulta en general bastante difícil de adquirir. O sea que, a pesar de contar con el acceso irrestricto, los aprendientes siguen cometiendo errores,

⁵Sistema gramatical propio de los aprendientes de L2, que posee características conjuntas de su L1 y de la L2 que se aprende. Aunque la interlengua conlleva alta variabilidad, resulta sistemática.

inclusive en etapas tardías del aprendizaje. Prévost and White (2000) propusieron la hipótesis de *Missing Surface Inflection*. Se establece allí que el déficit no se encontraría en las representaciones morfo-sintácticas abstractas, adquiridas con éxito, sino en el mapeo entre la morfo-sintaxis y la fonología. Es decir que, aunque el aprendiente posea de hecho el rasgo *fuerte* de plural (y haya tenido lugar la operación AGREE y MOVE a nivel de la sintaxis), a veces no lo podría realizar fonológicamente y recurriría, por consiguiente, a formas de *default*. Por último, en Sorace (2011) se avanzó la *Interface Hypothesis*, la cual afirma que el déficit tampoco está en la sintaxis sino en la interfaz con el nivel del discurso debido a problemas de manejo de recursos cognitivos (memoria a corto plazo, atención, etc.).

3.3.2. Funcionalismo orientado a los conceptos.

En el funcionalismo se considera que el principal objetivo del lenguaje es la comunicación. El problema consiste en describir el mapeo entre funciones y sus exponentes formales. Dicho mapeo puede ir en dos direcciones: *función-forma* o *forma-función*. En el enfoque orientado a los conceptos [*concept oriented approach*] se pone el énfasis en la primera dirección. La investigación en segunda lengua busca establecer cuáles son las formas que los aprendientes usan para expresar conceptos, y cómo el mapeo de los conceptos con las formas cambia con el aprendizaje (Bardovi-Harlig, 2015). Por ejemplo, el concepto de «futuridad» puede expresarse mediante adverbios (*mañana estudio*), perífrasis (*mañana voy a estudiar*), morfología (*estudiar-é*), futuros léxicos (*necesito comprar un remedio*), verbos modales (*quiero dar la vuelta al mundo*). El concepto de «pluralidad» se puede realizar por medio de cuantificadores (*muchos*), numerales (*dos*) o morfología (*niño-s*). El análisis procede identificando las formas que tiene el aprendiente en su inventario para expresar el concepto en cuestión en cada etapa de adquisición. Se debe observar cómo van cambiando los elementos de dicho inventario y el balance entre estos elementos. A modo de ejemplo, en el caso del concepto «secuencia de eventos en el pasado», en una primera etapa el aprendiente apelaría a la pragmática, asumiendo que los eventos suceden en el orden en el cual se expresan [*Llega. Pide un café*]. Luego se agregarían adverbios y conectores [*Ayer llega y después pide un café*]. Por último se usaría morfología [*Llegó y pidió un café*]. En la segunda etapa, además de incorporarse marcadores léxicos al inventario, también cambia el balance. En efecto, en la segunda etapa la morfología puede usarse muy poco o mediante fórmulas, en contraste con el uso más extendido de las herramientas léxicas (el cual está desbalanceado entre los exponentes de la función). En cambio, en la tercera etapa la morfología puede volverse más productiva y el hablante recurriría a herramientas tanto léxicas como morfológicas, de modo más balanceado. Esta noción de balance está emparentada con la de «carga funcional» [*functional load*]. Se afirma que si la forma de expresar una función se realiza mediante un único exponente, entonces su carga funcional es alta. Por ejemplo, el adverbio *ayer* conllevaría carga alta si fuera el único modo de expresar que el evento sucedió en el pasado. Por otro lado, si se expresara también por medio de morfemas, entonces la carga funcional del adverbio y del morfema bajarían, resultando en una carga más balanceada entre ambos. Otros dos principios se aplican también para analizar el cambio de las relaciones función-forma: «uno-a-uno» y «multi-función». En las primeras etapas del aprendizaje los aprendientes usarían el principio de «asociar una función a una forma». En etapas posteriores se extendería a «asociar varias formas a una función o varias funciones a una forma» (Andersen, 1990). Por ejemplo, la noción de pluralidad podría expresarse en primera instancia con el cuantificador *mucho* [*mucho estudiante*] y luego extenderse al uso del morfema *-s* [*(mucho) estudiante-s*]. Téngase en cuenta que las formas que hacen de exponentes de las funciones no necesariamente coinciden con las del hablante nativo. La relación forma-función otorga sistematicidad a la *inter-lengua*, aunque dicha sistematicidad no se aproxime a la L2. O sea, si en una etapa de su *inter-lengua* un hablante utilizara *ayer estudió* [que también podría significar *estudi-é, estudiaste*] o bien *estudi-amos* [que también indicaría *estudiaron*] las formas «-ó / -amos» cuentan como exponentes de pasado para las personas del singular y del plural (un paradigma con dos elementos). Por lo tanto, a diferencia del enfoque generativo, la noción de *gramaticalidad* no resulta importante.

3.3.3. Teoría de la procesabilidad.

La teoría de la procesabilidad (PT) [Pienemann, 1998, 2005; Pienemann & Lenzing, 2015] afirma que en cada estadio de la adquisición el aprendiente podrá producir o comprender solamente aquellas formas que puedan ser manejadas por el estado corriente del módulo del formulador [ver apartado 3.2]. Aprender una lengua implica adquirir gradualmente una serie de mecanismos procedurales que permiten producir estructuras progresivamente más complejas. Es esta «capacidad de procesabilidad» lo que limita el aprendizaje: si no se ha desarrollado un determinado mecanismo procedural no se podrá producir una determinada estructura de modo sistemático y productivo. Por tanto, se trata de una teoría psicolingüística sobre cómo se adquieren los mecanismos del formulador en lengua segunda. Los mecanismos procedurales se basan en la unificación de rasgos en la estructura de constituyentes de una *gramática léxica funcional* [LFG] (ver capítulo 1). Cada entrada del léxico mental del aprendiente está anotada con información de rasgos gramaticales. Por ejemplo la entrada «MARIANA» conllevará los rasgos CLASE=nombre_propio; NÚMERO=singular, GÉNERO=femenino. Análogamente, la entrada «COMIÓ» , tendrá valores para CLASE=verbo, NÚMERO=singular, TIEMPO=pretérito, PERSONA=tercera, ASPECTO=aoristo. Por ejemplo, en la oración *mariana comió muchas manzanas*, para lograr la concordancia nominal habrá unificación de rasgos de número y género en el SN objeto. Para lograr la concordancia sujeto-verbo se requerirá que los valores de número y persona se transmitan al SN y al SV y de allí a Sintagma Oracional (S), donde se producirá la unificación. A partir de los sintagmas donde se lleve a cabo la unificación de rasgos se deduce una jerarquía de procesabilidad que sigue la escala: (i) falta de unificación de rasgos > (ii) unificación a nivel del sintagma nominal o verbal > (iii) unificación de rasgos a nivel oracional. La jerarquía general es la siguiente:

- *Sin procedimiento*: no hay unificación de rasgos. Se pueden producir formas aisladas como «perro» o fórmulas como «me gusta», «buenos días».
- *procedimiento categorial*: no hay unificación de rasgos. Se acoplan morfemas a bases léxicas, como *manzana-s*.
- *procedimiento del sintagma*: se unifican rasgos en el SN. Se pueden producir concordancias nominales: *much-a-s manzan-a-s*. También se hace posible la adjunción de categorías al SV. Por ejemplo en *por la tarde duermo la siesta*.
- *procedimiento oracional*: unificación de rasgos a nivel oracional. Por ejemplo, concordancia sujeto-verbo, o con el predicativo o clíticos: *los libr-o-s son azul-es; le dio los regalos a su hermano*.
- *procedimiento de oración subordinada*: unificación de información a nivel de oraciones subordinadas. Por ejemplo el requerimiento de verbos con complementos oracionales en subjuntivo: *quiero que me regalen un libro*.

Se trata de una jerarquía implicacional: el nivel superior supone que se ha desarrollado el inferior. No puede haber saltos de niveles. Por lo tanto, los estadios de desarrollo de la *inter-lengua* son resultado del estado de desarrollo de la jerarquía de procedimientos. Por ejemplo, si no se ha adquirido el procedimiento nominal, no se observará en los datos sistematicidad y productividad en el uso de concordancias nominales. Se considera que una regla se usa de modo sistemático si se la observa en un número suficiente de *tokens* y es productiva si es aplicada cuando el contexto lo requiere. Si se verifican ambos requisitos se considera que un dato *ha emergido* en la producción del aprendiente. Por ejemplo, para considerar como emergente a la estructura del tipo «me gusta» es preciso ver si se usa con diferentes verbos («me interesa», «me encanta») y con diferentes formas («me gusta(n)», «le gusta(n)», «nos gusta(n)»). PT adhiere al principio de *unmarked alignment*, el cual establece un mapeo «uno-a-uno» entre roles argumentales y categorías funcionales. Esto quiere decir que en el estado inicial de adquisición solamente se puede observar el orden canónico de la L1 [SVO]: sujeto (agente) - verbo - objeto (paciente), para el español. En etapas posteriores el mapeo se podrá modificar, como en el caso de la voz pasiva, que se deberá adquirir después de las oraciones activas de SVO. Otra hipótesis que caracteriza el estado inicial de la L2 es la llamada *topic hypothesis*, que versa sobre el mapeo entre estructura de constituyentes y funciones gramaticales. Se hipotetiza que al principio

los aprendientes no logran diferenciar en posición inicial de tópico a la función SUJETO de otras funciones, como por ejemplo ADJUNTO. En el desarrollo posterior, el ADJUNTO se podrá poner en posición de tópico. Por último, los aprendientes llegarán a diferenciar otras funciones gramaticales en posición inicial. Así por ejemplo, en la primera etapa de la adquisición se observarán datos emergentes como: *Mario/SUJETO mira una película*; en una etapa posterior: *Ayer/ADJUNTO Mario miró una película* y tardíamente: *una película/OBJETO fue a ver Mario ayer*. Cada mecanismo procedural de la jerarquía restringe el *espacio de las hipótesis* del aprendiente; es decir, el rango de variación de la *interlengua*. Por ejemplo, en el tercer procedimiento la información gramatical solamente se puede intercambiar vía unificación dentro del sintagma nominal. Todavía no puede haber intercambio de información entre sintagmas. Por ello, se esperará variación del tipo: *los niños va a la escuela* o bien *los niños son inteligente*, con uso no emergente de concordancias a larga distancia. En suma, los aprendientes no podrán producir (de modo sistemático y productivo) estructuras que se encuentren más allá del nivel de procesamiento alcanzado. La producción de nuevas estructuras refleja el nivel de procesabilidad del formulador pero no lo causa. Esto significa que la procesabilidad limita la instrucción formal: la enseñanza solamente puede promover la adquisición de aquellas estructuras que el alumno está preparado para procesar.

Bonilla (2014) aplicó la jerarquía de procesabilidad al español. La jerarquía permite la *emergencia* de las siguientes estructuras: (i) estadio I: producción de lemas o expresiones formulaicas; (ii) estadio II: orden SVX (X = cualquier función gramatical excepto SUJETO o VERBO); morfemas léxicos (por ejemplo, plural); (iii) estadio III: Adjunción de adverbios y sintagmas preposicionales en posición de tópico, concordancia en el entorno del SN; (iv) estadio IV: concordancia sujeto-verbo, con predicativos, con clíticos; orden oracional diferente del canónico (por ejemplo: *¿Con quién habló tu hermano hoy?* [XVS]); (v) estadio V: marcación de subjuntivo en oraciones subordinadas. La autora analizó datos provenientes de alumnos del *Spanish Learner Language Oral Corpus* (disponible en CHILDES [MacWhinney, 2010]). En lo que atañe a la morfología, confirmó la escala implicativa de emergencia: *plural > concordancia nominal en SN > concordancia a larga distancia*. Hubo algunos indicios de emergencia de marcación del subjuntivo para algunos alumnos.

3.3.4. Aprendizaje basado en el uso y en el conocimiento declarativo / procedural.

Nick Ellis (2001, 2003, 2005, 2015) sostiene que se aprende formando macro-estructuras de información. Los aprendientes, a través del uso y la exposición frecuente al input, asocian unidades recurrentes que sustentan patrones y construcciones de diferente nivel de abstracción. Buscan continuamente secuencias con significado y, al hacerlo, aprenden implícitamente aspectos de la serialidad de las lenguas (p. ej., [*¿qué tal?*], [*muy bien, gracias*], [*¡vení para acá!*], [*¿Está Pablo?*]). Luego, a partir de la sustitución paradigmática en una posición se generalizan construcciones más abstractas (p. ej., [*tengo*] [*un libro/un problema/dolor de cabeza/frío/...*] > [*tengo + SN*]). Adquirir es, pues, abstraer información probabilística y distribucional.

La unidad básica de representación es la *construcción* entendida de modo amplio como pares forma / significado-función (ver el enfoque de *gramática de las construcciones* en el capítulo 1). O sea que una construcción puede ser una palabra [”ardilla”] ↔ *ardilla*, un marco morfológico [”pluralidad”] ↔ *raíz#consonate + es / raíz#vocal + s*, sintáctico [*tener + SN*], etc. Las construcciones se aprenden por aprendizaje asociativo. El cerebro abstrae de modo *implícito* regularidades estadísticas y se va creando una red de asociaciones. Dicho aprendizaje puede depender de: (i) la frecuencia de *tokens*; (ii) la saliencia perceptiva (la medida en la cual una forma sobresale en su contexto), (iii) efectos de bloqueo de otras construcciones que sean relevantes para la interpretación de ésta o, a la inversa, otras interpretaciones que se relacionen con la interpretación que se quiera asignar a la construcción que se aprenda; (iv) regularidad (ej.: verbos irregulares serán más difíciles); (v) fiabilidad de las «pistas» [*cues*] que ofrece el input para la interpretación. Por ejemplo, en inglés la construcción «N+*s*» puede significar «plural» (*boys*) o «posesión» (*boy’s hat*), con lo cual la pista de la «*s*» no resulta muy fiable. El aprendizaje se basa en «ejemplares». Cada vez que el aprendiente encuentre un ejemplo de construcción la comparará con la traza de ésta en su memoria o de alguna construcción similar y así podrá generar una interpretación. La exposición a ejemplos de construcciones será suficiente para lograr la abstracción. Por ejemplo, escuchar «buenas tardes», «buenas noches», generará la construcción abstracta: [”saludo para parte del día”] ↔ *buenos/as + parte del día + s*.

Para este enfoque resulta pues fundamental favorecer el aprendizaje por ejemplares y el *chunking*. Por ejemplo, diseñando ejercicios que pongan el foco en fórmulas (p. ej., los *roleplay* con fórmulas ‘listas para usar’: *ya está, ¿Está listo?*); locuciones: ‘*estar en la luna /en babia*’; colocaciones: *estar equivocado, correr un rumor, desempeñar un cargo*, etc.; y además dirigir la atención del estudiante a casos en los que éstos no puedan abstraer información probabilística por sí solos porque resultan difíciles de detectar: (a) recurrencias gramaticales importantes: *estar + en* [lugar]; *estar + el/la/los/las* [existencia]: *allí donde está / *hay el monumento a San Martín*; (b) concordancias (dependencias de larga distancia); (c) evidencia negativa: el hecho de que algo no sucede en una lengua, pero al tener frecuencia *zero* no se lo puede notar (p. ej., en español es imposible colocar algo dentro de una frase verbal: **he mucho estudiado y trabajado*, pero en italiano sí: *ho molto studiato e lavorato*). Por último, es necesario señalar que este enfoque considera la lengua como un *sistema dinámico adaptativo* (ver abajo).

El conocimiento experto está automatizado. En la teoría de la adquisición de destrezas (*skill acquisition theory*) se reconocen tres etapas para llegar a la habilidad experta (DeKeyser, 2015). Primero se almacena conocimiento declarativo (“saber qué”) explícito, generalmente en la forma de una regla aprendida de forma deductiva (regla > ejemplos) o inductiva (ejemplos > regla). Luego se crea conocimiento procedimental (“saber cómo”): se pone en práctica la regla (primero con ayuda del análisis y luego sin recurrir a éste). Por último, se postula una etapa de automatización: se practica abundantemente para bajar el tiempo de ejecución, el número de errores y la cantidad de atención involucrada en la tarea. Por ejemplo: (a) declarativo: regla de formación del plural: “vocal + s / consonante + es”; (b) procedimental: al principio, ‘alemán’ [estímulo] > ¿plural? [respuesta] (análisis del dato: ¿hay una consonante? Sí, veo una «n»; entonces será “alemanes”); luego de la práctica, catalán > ¿plural? “catalanes” (sin análisis); (c) automatización: aunque no se use la regla para analizar, el aprendiente cometerá errores como “alemanos”; sin embargo, con la práctica se reducirán hasta una proporción no significativa. Anderson y sus colegas han propuesto varias versiones de una teoría computacional para explicar la adquisición de destrezas (Anderson et al., 2004; Taatgen & Anderson, 2008). En la teoría denominada ‘control adaptativo y racional de pensamiento’ [ACT-R, por sus siglas en inglés], la memoria declarativa modeliza la MT en términos de un subconjunto de ítems activados por su uso frecuente en el pasado y por su alta correlación con el contexto de uso actual. La práctica disminuye el tiempo que se tarda en seleccionar un ítem en la memoria, es decir, las latencias de selección. Por consiguiente, la memoria explícita estaría representada por un ítem seleccionado y la implícita, por el proceso sub-simbólico de activación que posibilita su selección. Un sistema central ejecutivo modeliza la memoria procedimental, que aplica reglas de producción del tipo “SI X ENTONCES Y” a determinada tarea. Supóngase, pues, teniendo como información en la memoria declarativa ESTAR EN [+LUGAR]: (a) regla 1: SI quiere expresar UBICACIÓN SIN MOVIMIENTO y se encuentra en un contexto de “verbo + preposición de lugar”, ENTONCES seleccione el verbo correspondiente asociado a la preposición; (b) SI el verbo ha sido seleccionado con respuesta *n*, ENTONCES imprima *n*. Así, si un sujeto quiere expresar ubicación usando la preposición «en», entonces recuerda la regla “¿Verbo de ubicación + en? = estar” y produce el verbo *estar* (y no el verbo *ser*). Un proceso de compilación unifica las dos reglas generales en una específica de uso concreto: (c) SI quiere expresar ubicación, ENTONCES imprima *estar + en*. En la compilación, pues, se combinan dos reglas existentes para crear una nueva de dominio específico que posea menos pasos. Una vez creada la nueva regla (c) entrará en competición con (a) y (b) por su valor de “utilidad” (calculada en función del costo de haber aplicado la regla anteriormente y el éxito obtenido en aquellas ocasiones para alcanzar un objetivo determinado).

También se ha avanzado sobre las bases neuronales de la distinción entre memoria declarativa y procedural [*declarative / procedural model*] (Paradis, 2009; Ullman, 2005; ver también: García, 2012). En el lenguaje, la memoria declarativa subsume, por ejemplo, las representaciones fonológico-conceptuales de las palabras, las fórmulas (p. ej., *quién sabe, haberlo pensado antes, de tal palo tal astilla*), los *chunks* (*tengo + N, cometer + error, ¿todo bien?*), las palabras sintácticas (*mesa redonda, boca de lobo*) y las formas irregulares (*mejor, fue*). Dicho brevemente, la memoria declarativa se especializa en la representación del léxico y la semántica. El sustrato neuronal está constituido por regiones medio-temporales y otras estructuras subcorticales y la codificación puede ser muy rápida –un único contacto con el estímulo puede bastar. La información declarativa es típicamente accesible de modo consciente o explícito (aunque también podría existir codificación implícita). En la memoria procedural se alojan mecanismos tales como la abstracción

probabilística, la serialización, la concatenación o el movimiento de larga distancia (todos relacionados con la sintaxis, la morfología y la producción fonológica); y se relaciona con estructuras frontobasales y el cerebelo. El desarrollo de representaciones en la memoria procedural es necesariamente gradual y los mecanismos de procesamiento son inconscientes e implícitos. Ambos sistemas funcionan en paralelo en el aprendizaje, o sea que no existe 'interfaz' entre ellas (cf. Ellis, 2005): el conocimiento declarativo no 'se convierte en' procedimental. También se ha observado que sendos tipos de memoria compiten entre sí: la actividad de una conspira bloqueando a la otra. De la evidencia empírica se desprende, además, que el aprendiente con alto nivel de automatización de su L2 procesa la sintaxis y la morfología regular en las regiones procedimentales; por otra parte, el aprendiente con bajo nivel de automatización lo hace en los sustratos involucrados en la memoria declarativa (García, 2012).

3.4. Sistemas Dinámicos Complejos (SDC).

Un *sistema dinámico* es aquel que cambia en el tiempo. El sistema posee un conjunto de componentes que interactúan de una forma determinada para generar un *estado* particular (Larsen Freeman & Cameron, 2008). Un *estado* se define como un patrón de actividad determinado del sistema en un punto determinado del tiempo. Los componentes pueden ser entidades, procesos e incluso otros (sub-)sistemas. Si un sistema se cataloga como *simple*, dichos componentes se hallan conectados de una manera que resulta en un comportamiento predecible y sin cambios abruptos. Ello quiere decir que una vez especificadas las reglas que gobiernan el sistema, es posible conocer sus estados futuros. Por otro lado un sistema que es *complejo* conlleva una serie de características que lo diferencian de sus contrapartes simples. A saber:

1. Heterogeneidad: Los componentes del sistema pertenecen a tipos diferentes y podrían ser incluso otros sistemas complejos.
2. Dinamismo: Los componentes y su forma de interactuar entre sí cambian con el tiempo.
3. Falta de anidamiento: las variables se relacionan con otras que se hallan en otros niveles. No obstante, dicha relación no puede pensarse como «jerárquica» ya que la influencia entre las variables pueden funcionar en cualquier dirección.
4. Proceso iterativo: El sistema es afectado por flujos provenientes de ambiente y por el estado previo del sistema.
5. Falta de linealidad: En un sistema lineal la relación entre los elementos está fija y el cambio es constante y proporcional al input recibido. Por otra parte, un escenario no lineal es aquel donde las relaciones entre elementos del sistema se mueven constantemente y el cambio no es proporcional.
6. Interacción con el ambiente: El ambiente o contexto es el origen de la información (recursos) necesaria para que el sistema lleve a cabo la dinámica. Sin embargo, dichos recursos son escasos. Existe un flujo constante de información del ambiente hacia el sistema y viceversa. Esta «apertura» quiere decir que el sistema permite la salida o entrada de energía; con lo cual, el contexto mismo forma parte del sistema y contribuye a su complejidad. Asimismo, el sistema ajusta de modo constante su respuesta a los cambios del ambiente; o sea, se adapta a transformaciones del contexto.
7. Auto-organización: Se produce un incremento de complejidad por medio de la creación de nuevas formas. Dichas formas no se hallaban codificadas en las condiciones iniciales. En consecuencia, el sistema se auto-organiza generando patrones más complejos. Esto se realiza al costo de consumir energía del ambiente (porque el sistema es abierto). Además se depende de la disponibilidad de información en el ambiente.
8. Fractalidad: Es un patrón que existe en un nivel de escala del sistema y también para cualquier otro nivel de escala. El sistema se vuelve similar a sí mismo.

Se pasa ahora a la descripción formal de un sistema dinámico (Kaplan and Glass, 1995; Kot, 2001; Strogatz, 1994, Sprott, 2003). Las reglas que gobiernan el sistema se especifican como ecuaciones lineales o no lineales. En el primer caso, cada término es o bien una constante o bien un producto de una constante y una variable (elevada a la primera potencia). Se denomina espacio de fase [*phase space*] a aquel espacio de m dimensiones donde se representan todos los posibles estados del sistema, correspondiendo cada estado a un único punto $x \in \mathbb{R}^m$ en dicho espacio. Para una sola dimensión se denomina «línea de fase» [*phase line*] y en dos dimensiones es un «plano de fase» [*phase plane*]. A medida que pasa el tiempo el vector de estado x cambia posición en el espacio de fase, creando así una trayectoria. Si el tiempo es una variable discreta, la dinámica está descrita por ecuaciones de diferencias en un mapa m dimensional: $x_{n+1} = f(x_n)$, $n \in \mathbb{Z}$. Por otro lado, si el tiempo es continuo, la dinámica se caracterizará con ecuaciones diferenciales ordinarias de primer orden⁶: $\frac{dx}{dt} = f(x, t)$, $t \in \mathbb{R}$. En cuyo caso, en lugar de tener un mapa, se tiene un *flujo*. Se restringirá la presentación a los flujos no lineales en una dimensión.

Un modelo famoso de ecología de poblaciones es el de crecimiento logístico, descrito como:

$$\frac{dN}{dt} = f(N) = rN \left(1 - \frac{N}{K} \right)$$

La variable de estado N es la abundancia de una especie⁷. Dicha abundancia crece a una tasa constante r positiva. Sin embargo, dicho crecimiento tiene el límite de la «capacidad de carga» del ambiente, la cantidad de individuos que el ambiente es capaz de soportar. La función $f(N)$ es cuadrática y tiene la forma de una parábola. A medida que $t \rightarrow \infty$ y para toda abundancia inicial $N_{t=0} > 0$, $N \rightarrow K$, y cuando $N = K$ la tasa de crecimiento es *ceró*. Es decir que dicha tasa de crecimiento disminuye con la abundancia poblacional de la especie. En el siguiente ejemplo, $K = 2$ y $r = 1$. En la gráfica de línea de fase siguiente se observa claramente que para cualquier condición inicial $N_{t=0} > 0$ (marcada en círculos), a medida que $t \rightarrow \infty$, el flujo va siempre hacia la asíntota $K = 2$.

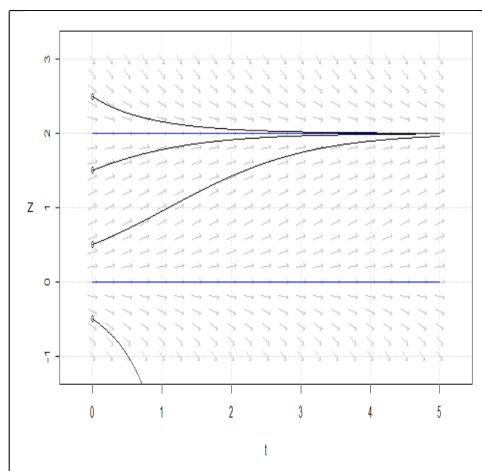


Figura 3.2: Línea de fase para la ecuación de crecimiento logístico.

Un punto fijo del sistema $N = N^*$ es aquel que hace que $f(N^*) = 0$. Por lo tanto:

$$f(N^*) = 0 \iff rN^* \left(1 - \frac{N^*}{K} \right) = 0 \Rightarrow Pf = \{N^* = 0, N^* = K\}$$

Es decir que cuando $N^* = \{0, K\}$, $f(N^*) = 0$, lo cual significa que la tasa de crecimiento es constante: el sistema se queda para siempre en dicho punto de equilibrio. Esto se ve en la gráfica: si la condición inicial fuera $N_{t=0} = \{0, K\}$ el sistema no se mueve (líneas azules).

⁶Son ecuaciones «diferenciales» porque involucran derivadas; de «primer orden» porque la derivada de nivel más alto es la primera; y son «ordinarias» porque no hay derivadas parciales.

⁷«especie» es un mero término para denotar cualquier entidad que se pueda contar. Pueden ser peces, pero también un tipo de virus o en el lenguaje, concordancias «erróneas».

En sistemas de flujos unidimensionales, un punto fijo puede ser *un atractor* (estable), *un repulsor* (inestable) o ser un «punto silla» (*saddle*). Un atractor es una región (en una dimensión es un número) del espacio de fase al cual el sistema es atraído. Con el objetivo de descubrir el tipo de punto fijo, se hace necesario calcular la derivada con respecto a N y evaluarla en el punto fijo en cuestión: $\lambda = \left[\frac{df(N)}{dN} \right]_{N^*}$. Si $\lambda < 0$, el punto fijo es atractor, si $\lambda > 0$, es repulsor. Si $\lambda = 0$ puede ser atractor, repulsor o saddle; y es necesario examinar el signo la segunda derivada. Si es un saddle, la segunda derivada tiene el *mismo* signo en un punto $x = x^* \pm \epsilon$, a ambos lados de x^* . En un saddle, el punto fijo atrae en una dirección y repele en la otra (o atrae a la izquierda y repele a la derecha o viceversa). Para el caso de un atractor con $\lambda = 0$, la segunda derivada cambia de signo cuando pasa por x^* : es positiva en un punto a la izquierda de x^* y negativa a la derecha de x^* ; y al revés con un repulsor: es negativa en un punto a la izquierda de x^* y positiva a su derecha. En el caso de la función logística:

$$\lambda = f'(N) = r - \frac{2rN}{K} \Rightarrow f'(N^*) = \begin{cases} r & N^* = 0 \\ -r & N^* = K \end{cases}$$

Con lo cual si $N^* = 0$, $\lambda > 0$ y es un repulsor; y si $N^* = K$, $\lambda < 0$ y es un atractor. El siguiente es el «retrato de fase» [*phase portrait*] del sistema. En el eje x se grafica la variable de estado N y en el eje y la derivada, o sea, la tasa de cambio de N . Las flechas indican la dirección del flujo. Como se ha mostrado, en $N = 0$ y $N = K$ se hallan los puntos fijos. Se observa que el flujo se aleja de $N^* = 0$ hacia la derecha (hacia la izquierda no tiene sentido porque las abundancias deben ser positivas). Es decir que cualquier condición inicial que comience arbitrariamente cerca del *cero* se aleja de éste, $N^* = 0$ repele. Por otra parte, el flujo se acerca a $N^* = K$ desde la izquierda y desde la derecha, con lo cual atrae el flujo. Nótese que la dinámica no depende del parámetro r .

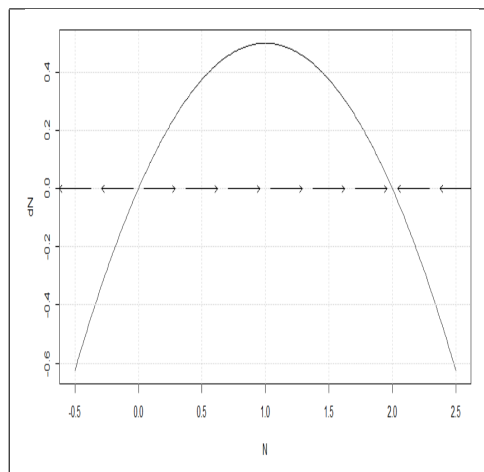


Figura 3.3: Retrato de fase para la ecuación de crecimiento logístico.

El conjunto de condiciones iniciales que llevan a un atractor se llama «basin of attraction» o «cuenca de atracción», aquí es $BA(N^* = K) : \{N \in \{\mathbb{R}_{>0} \setminus K\}\}$, cualquier abundancia positiva que no sea K va hacia el atractor.

En dos dimensiones es posible ver ciclos (un patrón que se repite con periodo k)⁸, así como también otros tipos de puntos fijos. Los flujos de tres dimensiones pueden exhibir comportamientos aperiódicos: la solución oscila pero el periodo es muy largo o infinito. Esto se llama *caos*. Dicho fenómeno posee las propiedades siguientes:

1. Es aperiódico: un valor no se repite nunca (ciclo infinito) o bien el ciclo es finito pero muy largo.
2. Está acotado [*bounded*]: en las sucesivas iteraciones del sistema los valores se mantienen en una región del espacio de fase; o sea que no escapan a infinito. Por eso se lo considera un atractor «extraño».

⁸En una dimensión no hay ciclos en los flujos en la línea, pero sí en los flujos en el círculo.

3. Es determinístico: En la ecuación no existen términos estocásticos. Para cada posible valor de $x(t)$ hay un *único* valor posible para el valor sucesivo $x(t + dt)$.
4. Es sensible a las condiciones iniciales: dos puntos que inicialmente están cerca en $t = 0$ van a divergir a medida que $t \rightarrow \infty$. Esto tiene la importante y fascinante consecuencia de que las predicciones tendrán validez solamente en el futuro cercano.

Con el objetivo de introducir el concepto de *bifurcación*, considérese el siguiente sistema de crecimiento logístico modificado.

$$\frac{dN}{dt} = f(N) = rN \left(1 - \frac{N}{K} \right) - qN$$

donde q es una constante de «depredación», que disminuye la tasa de crecimiento de la especie. Como en el ejemplo precedente, los puntos fijos son aquellos que hacen que $f(N^*) = 0$:

$$f(N^*) = 0 \Leftrightarrow rN^* \left(1 - \frac{N^*}{K} \right) - qN^* = 0 \Leftrightarrow rN^* \left(1 - \frac{N^*}{K} \right) = qN^*$$

Los puntos fijos son $N^* = 0$ y $N^* = K \left(1 - \frac{q}{r} \right)$. La Figura siguiente ilustra el retrato de fase. Los puntos fijos están representados por puntos negros. Se muestra la función logística en azul ($r = \frac{1}{2}$; $K = 100$) y el término de «depredación» qN en línea punteada. Obsérvese que para $q < r$ [Figura de la izquierda y del medio], $N^* = 0$ es repulsor (el flujo se aleja) y para $N^* = K \left(1 - \frac{q}{r} \right)$ es atractor, el flujo se dirige hacia el punto fijo. Notar que a medida que $q \rightarrow r$, el punto fijo atractor se va acercando al punto fijo repulsor. Ahora bien, cuando $q = r$, el punto fijo atractor «colisiona» con el punto fijo repulsor. Luego, con $q > r$, el punto atractor ha desaparecido del primer cuadrante ($N^* < 0$) y se ha convertido en repulsor. Por otra parte, el punto fijo $N^* = 0$ ahora es atractor. Esto se llama bifurcación *transcrítica*: dos puntos fijos colisionan e intercambian el estatus de estabilidad. En suma, una *bifurcación* es un cambio cualitativo en el comportamiento del sistema. Este resultado es previsible: cuando la tasa de depredación de un recurso sobrepasa a la tasa de crecimiento del mismo, la población tiende a la extinción. En el ejemplo, el parámetro dinámico es q : según los valores que tome el sistema tiene trayectorias con comportamiento diferente.

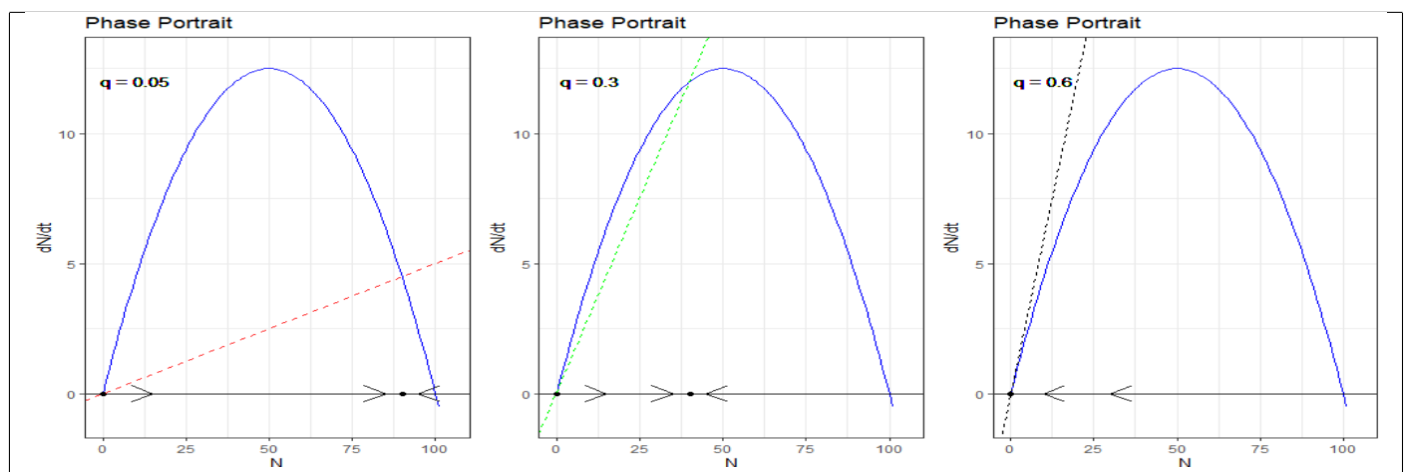


Figura 3.4: Bifurcación. Izquierda: $q = 0,05 < r$; Medio: $q = 0,3 < r$; Derecha: $q = 0,6 > r$.

Los tipos de bifurcación para flujos unidimensionales son los siguientes, expresados en sus *formas normales*:

- *Bifurcación nodo silla*: $\dot{x} = r - x^2$. Al principio no existen puntos fijos ($r < 0$). En $r = 0$ se crea un punto fijo, que es un *saddle node* (nodo silla). Cuando $r > 0$, se crean dos puntos fijos, uno estable y otro inestable.

- *Bifurcación transcritical:* $\dot{x} = rx - x^2$. Al principio, para $r < 0$, existe un punto fijo inestable en $x^* = r$ y un punto fijo estable en $x^* = 0$. En $r = 0$ los puntos fijos colisionan. Con $r > 0$, el punto fijo del origen es ahora inestable y $x^* = r$ se ha convertido en estable.
- *Bifurcación tridente super-crítica:* $\dot{x} = rx - x^3$. Se comienza con un punto fijo estable en $x^* = 0$ cuando $r < 0$. Para $r > 0$, aparecen dos puntos fijos estables en $x^* = \pm\sqrt{r}$ y $x^* = 0$ que era estable se vuelve inestable. El término cúbico es estabilizante.
- *Bifurcación tridente sub-crítica:* $\dot{x} = \mu x + x^3$. Se comienza con un punto fijo estable y otros dos puntos fijos inestables en $x^* = \pm\sqrt{-r} = \pm i\sqrt{r}$. Luego de $r = 0$, donde se produce la bifurcación, los dos puntos fijos inestables desaparecen y el punto fijo que era estable se vuelve inestable. A diferencia del caso anterior el término cúbico es desestabilizante.

Otro modo de visualizar la dinámica de un sistema unidimensional $\dot{x} = f(x)$ es mediante el *potencial* $V(x)$, definido por: $-\frac{dV}{dx} = f(x)$. Por ejemplo, para la bifurcación super-crítica $\dot{x} = rx - x^3$; el potencial (resolviendo la ecuación diferencial por separación de variables) resulta:

$$-\frac{dV}{dx} = f(x) \Leftrightarrow -\frac{dV}{dx} = rx - x^3 \Leftrightarrow \int dV = \int -rx + x^3 dx \Leftrightarrow V(x) = -\frac{1}{2}rx^2 + \frac{1}{4}x^4 + C$$

La gráfica que sigue lo ilustra (para $C = 0$). A la izquierda se grafica el potencial cuando $r < 0$. Allí el mínimo en $x^* = 0$ es un punto fijo estable. A la derecha se grafica el caso $r > 0$. Ahora el máximo local en $x = 0$ es inestable (representado por un círculo no relleno). En cambio los mínimos en $x^* = \pm\sqrt{r}$ son puntos fijos estables (representados por círculos rellenos). El potencial es como si fuera un «paisaje»: en las cimas se encuentran los repulsores; en el fondo de los valles, los atractores. Si una «canica» estuviera en el fondo del valle y fuera «perturbada» (movida para cualquier dirección) le sería muy difícil salir del «valle». Si, en la Figura de la izquierda, dicha «canica» estuviera en la «cima», cualquier movimiento la haría descender a uno de los valles. O sea: las condiciones iniciales arbitrariamente cerca de los puntos fijos repulsores (la cima) se alejan de ellos.

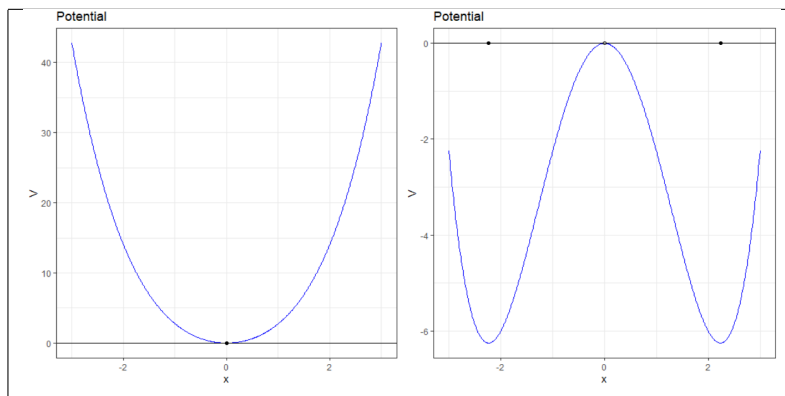


Figura 3.5: Gráfica de potencial para la bifurcación super-crítica $\dot{x} = rx - x^3$. Izquierda: $r = -5$; Derecha: $r = 5$.

3.5. Pensar la adquisición desde los SDC

3.5.1. Relación entre SDC y adquisición de segunda lengua.

Varios artículos han explorado la relación entre los sistemas dinámicos complejos y la adquisición de lenguas extranjeras: Larsen-Freeman (1997, 2008, 2011); Lowie (2013); De Bot et al. (2013, 2007). Esta sección ejemplifica los conceptos de las secciones anteriores con respecto al aprendizaje de español como lengua segunda y extranjera.

Un alumno, al aprender una lengua, crea variabilidad, o sea, produce nuevas formas que no están codificadas en el input: “errores”. La interacción entre los parámetros de un sistema dinámico eventualmente

hace que surja una coordinación entre ellos, emergiendo un patrón coherente que se perpetúa en el tiempo: el atractor; que se representa como puntos recurrentes en una región del espacio de fase. Es decir que los alumnos vuelven a cometer los mismos patrones de errores; incluso quedando atrapados en un estado de bimodalidad (por ejemplo, incurriendo siempre en faltas con la dupla *estar / haber*: *en la plaza donde hay / está el monumento*). O bien, si la adquisición avanza, se acercan a una región del estado de fase cuyos patrones son muy similares a los del hablante nativo. Existen también errores o construcciones que se producen raramente; o dicho de otro modo, constituyen “repulsores” en el espacio de fase. Sin embargo los atractores poseen diferentes grados de fuerza; de ahí que los alumnos persistan en la reproducción de determinado patrón más o menos tiempo antes de saltar a un patrón diferente.

Las condiciones iniciales de cada alumno son distintas, y, por lo tanto, los trayectos de aprendizaje también. Esto es así porque un estado del sistema depende del anterior, o sea, depende del conocimiento (habilidades) adquirido por el estudiante en particular. En otras palabras, las condiciones de los diferentes subsistemas (fonológico, morfológico, sintáctico, etc.) en un punto temporal determinan la producción en el siguiente.

El aprendizaje depende asimismo de los recursos del “ambiente cognitivo”: memoria de trabajo, atención, motivación, esfuerzo, etc; y el estado del sistema se adapta a cambios en dichos factores. El sistema puede ser “perturbado” en cualquier momento por transformaciones repentinas del “ambiente cognitivo”. Esto último implica que la predicción a largo plazo se torna imposible: razonamientos de la clase de “al ritmo que va, en tres semanas el alumno debería ya saber esto o aquello” carecen de sentido. Los alumnos adaptan, pues, su producción al contexto, que es cambiante; lo cual desencadena fluctuaciones y variabilidad. Ello equivale a decir que el aprendizaje no resulta lineal: un estudiante mejora y empeora, hay idas y vueltas, regresiones. En consecuencia, no es plausible pensar el aprendizaje desde la metáfora de una «escalera», en la cual los aprendientes pasan por estadios discretos de menor a mayor complejidad en sentido ascendente, partiendo de un “escalón inicial” y llegando a un “escalón final” o estado nativo (Larsen-Freeman, 2006). No existen estadios discretos sin variabilidad; estos son atractores en los que determinadas formas dominan en frecuencia sobre otras.

Además, se encuentran factores que son decisivos para el comportamiento del sistema. Estos se denominan variables controladoras, ya que, más allá de un cierto valor (punto crítico) el sistema sufre un cambio cualitativo (bifurcación) y se reorganiza hacia una nueva configuración de atractores. El atractor inicial se vuelve inestable y comienza a fluctuar cada vez más, llevando a saltos y discontinuidades. Como resultado *emergen* nuevos patrones que, eventualmente, se vuelven estables. Cerca del atractor hay linealidad y continuidad; en cambio, cerca de una transición se observan discontinuidad y saltos (variabilidad). En suma, opera un mecanismo acumulativo de competencia entre las formas hasta atravesar el punto crítico, lo cual dispara el cambio de dinámica. Por ejemplo, en el contexto de un locativo puede dominar la forma *estar*, y en el contexto de los adjetivos psico-físicos la forma de *ser*: *está en su casa porque *es enfermo*; en el cambio de fase observaríamos gran variabilidad entre *ser* y *estar* con dichos adjetivos; luego la dinámica se estabilizaría -con algunas perturbaciones que llevarán a errores, por supuesto- en una nueva configuración *estar + lugar / + adjetivos psico-físicos*. Otro contraste con la metáfora de la “escalera” es que una pregunta del tipo: “¿Se puede llegar al estado final del hablante nativo?” también carece de sentido en este marco teórico. A parte del hecho de que el supuesto estado final no coincide con una variedad homogénea y fija en la comunidad lingüística (lo cual constituye claramente una ficción de la normativa), los subsistemas no lograrían tampoco llegar a dicho estado simultáneamente. Recuérdese asimismo que los atractores no son irreversibles porque dependen de las condiciones de ambiente. Por ejemplo, luego de años de práctica, las configuraciones de *estar + lugar / + adjetivos psico-físicos / + adjetivos léxicos* (*el libro está ahí / mi amigo está contento; es un chico listo - la cena está lista*) pueden lograr precisión arriba del 85% (atractor más estable) pero las del tipo *está triste porque lo dejó la novia / es un niño triste* seguramente experimentarían mucha más variabilidad (atractor menos estable). Téngase en cuenta además que la variedad del nativo también es un sistema dinámico en flujo, las formas también cambian extendiendo usos: *el libro está ahí* (locativo: “cerca”) > *estás ahí de cobrar una buena paliza* (aspectual: “cerca” en el tiempo, “estar a punto de”). En suma, se trata de un mecanismo de coordinación entre dos sistemas (aprendiente / nativo) en flujo.

La auto - organización, es decir, el hecho de que el sistema avance hacia nuevas configuraciones estables y no presentes en las condiciones iniciales, implica una reorganización en la información. Aprender significa

que la información se vuelva más clara, que haya más certidumbre, ya que los atractores que emergen tienden a la estabilidad, hay menos períodos de intermitencia de error. Para usar el ejemplo anterior, en el contexto del locativo y del adjetivo psico-físico el aprendiente usará con más probabilidad *estar* que *ser*. Al auto-organizarse, el sistema del aprendiente reduce la incertidumbre (donde tenía más dudas sobre qué forma utilizar ahora tengo menos) y la complejidad de lo aprendido (lo que era difícil ahora es más fácil).

El aprendiente no es un mero receptor pasivo de input. En lugar de «input», es mejor hablar de *affordances* o percepción de oportunidades de aprendizaje. Se trata de relaciones aprendiente-entorno comunicativo, que el mismo aprendiente percibe como posibilidad o potencialidad de aprendizaje. Una vez detectada la oportunidad en la interacción comunicativa, el aprendiente pasará a la acción adaptando los recursos de expresión de los cuales dispone. Dichas potencialidades para la acción son percibidas por los aprendientes mientras están activos; o sea, participando de actividades y tareas comunicativas.

Por último, se ha observado evidencia de fractalidad: un patrón verificable en un nivel con una determinada escala también se verifica en otra escala, así el sistema se vuelve similar a sí mismo. En relación con ésto, se detectaron distribuciones de *Zipft* en la producción de las construcciones que se adquieren (Ellis & Ferreira, 2009; Ellis & Larsen-Freeman, 2009) o la evolución de la variabilidad en el aprendizaje hacia un patrón de «ruido rosa», que señala la coordinación óptima entre los subsistemas (Lowie *et Al.*, 2014).

3.5.2. Implicancias metodológicas e investigación en SDC.

Metodológicamente, se pone la cuestión de en cuál nivel estudiar el cambio lingüístico: ¿el del grupo o el del individuo? Por un lado, tomar promedios grupales permite «generalizar»; aunque, por otro, esconde la variabilidad que sucede realmente a nivel del individuo, reduciéndola a mero «error» estadístico.

Molenaar (2008; ver también: Fei Gu *et al.*, 2014) aporta argumentos sobre por qué el nivel del individuo parece ser el correcto. La equivalencia entre variación intra-individual e inter-individual (grupo) está asegurada únicamente para procesos *ergódicos*. Un proceso (una trayectoria en términos de una serie de tiempo) es ergódico si cumple dos requisitos: (1) es estacionario: la media no depende del tiempo; la covarianza depende solo de diferencias entre puntos temporales de la serie (varía sólo si varía el *lag*); (2) es homogéneo: cada elemento del grupo obedece a la misma dinámica y al mismo modelo estadístico (por ej., el supuesto de distribución gaussiana); haciendo que sean intercambiables. Sin embargo, la realidad está llena de procesos no estacionarios y heterogéneos, aun sin contar con la variación resultante de la auto-organización de todo sistema dinámico. Si bien a nivel grupal se ha encontrado evidencia de etapas ordenadas de adquisición, estas se basaban en promedios de precisión de grupos de aprendientes, y no en datos longitudinales de alumnos individuales. El análisis desde el punto de vista individual, en cambio, revelaba gran variabilidad y falta de ajuste al supuesto «orden universal» del promedio. Es decir que dichas etapas ordenadas de grupo no eran representativas de las trayectorias individuales (Lowie y Verspoor, 2015). Se desprende que las conclusiones obtenidas a nivel del grupo no son aplicables a nivel del individuo. Por eso, para «generalizar», en el presente marco teórico se adopta una estrategia «abajo hacia arriba», agrupando por similitud de trayectorias individuales.

Dörnyei (2014) distingue tres potenciales estrategias de investigación para abordar el aprendizaje como sistema dinámico: (a) identificación de fenómenos gobernados por atractores fuertes: ¿Qué atractores son más estables y cuáles menos? ¿Cambia la estabilidad de un atractor?; (b) identificación de «*cuencas de atracción*»: ¿Cuál es la óptima configuración de variables que lleva a que un atractor sea más estable? o dicho de otro modo: ¿Cuáles son las variables que se encuentran dentro del área de control de un atractor?; (c) identificar los patrones estables e investigar cómo la combinación e interacción de los elementos del sistema llevaron a su formación; es decir, identificar trayectorias (observando similitudes y diferencias entre los aprendientes). Recuérdese que la predicción a largo plazo no es posible en un sistema dinámico⁹; por eso, el método clásico de predecir hipótesis para luego probarlas no parece efectivo. Es necesario pues un método «retroactivo», a partir de la trayectoria que lleva a una configuración en particular, reconstruir los elementos del sistema y su interacción (Byrne, 2002).

⁹La falta de predictibilidad a largo plazo no es tan terrible como parece. Después de todo, como recuerda Larsen-Freeman (2011), se trata de una situación compartida con otras ciencias, por ejemplo: la meteorología, la geología, la biología evolucionista.

Un sistema dinámico comprende varias variables con diferente grado de acoplamiento entre ellas, lo que determina el comportamiento individual de cada individuo, o sea, su específica trayectoria de adquisición (Van Geert, 2014). Por ejemplo, Goldschneider y DeKeyser (2001), en un meta-análisis de los estudios transversales sobre el orden de adquisición de morfemas, observaron que los factores que daban cuenta de una gran porción de la variabilidad observada en la precisión eran: la saliencia perceptiva, la complejidad semántica, la regularidad morfológica y la frecuencia. Se presume que cada uno de estos afectará a cada individuo en diferente manera y medida en diferentes puntos de una particular trayectoria. Incluir la dimensión temporal es la manera de abordar dicha multi-causalidad: diseñar estudios longitudinales, con cantidad suficiente de puntos en la serie histórica. Dado que el cambio se produce por interacción de estos múltiples factores la estadística a usarse deberá ser necesariamente multivariada, con variables multidimensionales. Además resulta fundamental utilizar métodos sensibles a la no linealidad y a datos no estacionarios y heterogéneos.

Los artículos que han seguido este enfoque se han centrado en producción escrita, generalmente buscando evidencia de que la adquisición de lenguas segundas se rige efectivamente por los principios de los sistemas dinámicos arriba descritos. Varios de ellos han usado también el modelo ecológico de Paul Van Geert (1994, 1995, 2003); en el cual las variables interactúan entre sí por medio de varias relaciones funcionales y compiten por la disponibilidad de recursos limitados (memoria de trabajo, atención, motivación, tiempo invertido en el aprendizaje, conocimiento pre-existente, etc.). Estos recursos ejercen un efecto de curvatura en lo que de otra forma sería un crecimiento exponencial de dichas variables. Las relaciones entre éstas pueden ser: (a) de soporte: ambas se desarrollan al unísono porque se refuerzan mutuamente (correlación positiva); (b) de competición: cuando una crece la otra decrece debido a la competición entre ellas (correlación negativa); (c) condicional: se necesita que una variable llegue a un cierto nivel para que la otra empiece a crecer.

Verspoor, Lowie y Van Dijk (2008) estudiaron las producciones escritas a lo largo de dos años en un estudiante holandés de inglés L2 (nivel superior). La relación entre la longitud media de los SNs y el ratio palabras / verbos finitos (ambas medidas de complejidad) fue en general de apoyo mutuo; sin embargo al principio de la trayectoria se registró un poco de competición entre las variables. Esto significa que si bien en general la complejidad de los SNs aumenta junto a la oracional, al principio el aprendiente buscó complejizar produciendo SNs más largos a expensas de oraciones con verbos finitos, una estrategia común en Holandés.

Ellis y Ferreira (2009) encontraron, en estudiantes de inglés L2, evidencia de distribuciones de Zipft; las cuales siguen una ley de potencia, un posible índice de fractalidad. Examinaron la adquisición de tres tipos de construcciones: (a) Verbo + locativo: *come here!*; (b) Verbo + objeto + locativo: *put it over there*; (c) Verbo + objeto1 + objeto2 (distransitiva): *give the book to me / give me the book*. Hallaron evidencia de ley de potencia en la distribución de los tipos de verbos usados en cada construcción tanto en el input como en el output. Es decir que la frecuencia de cada tipo de verbo era inversamente proporcional a su posición en la tabla de frecuencias para cada construcción en particular. Además, comparando los verbos pertenecientes a cada construcción, el tipo de verbo de más frecuencia en el input era siempre mucho más frecuente que los otros verbos, resultaba ser además altamente prototípico y de significado genérico. O sea, un input que sigue la distribución de Zipft ayuda a optimizar el aprendizaje ofreciendo ejemplares prototípicos, frecuentes y ampliamente aplicables a varios contextos. Con dichos verbos el aprendiente se forma «una construcción marco», a partir de la cual puede reclutar los otros verbos para cada construcción.

Caspi (2010) siguió durante diez meses a cuatro estudiantes de inglés L2 de nivel avanzado. Consideró cuatro niveles jerarquizados de conocimiento de vocabulario de menor a mayor uso en producción: (1) reconocimiento; (2) recuperación de la memoria; (3) producción controlada; (4) producción libre. Halló que los niveles (1) y (2) más pasivos se apoyaban entre sí; en cambio, los más activos (3) y (4) competían entre sí. También estudió las variables de precisión y complejidad en la sintaxis y el léxico en producción escrita. Encontró que las variables de precisión y complejidad en el dominio léxico evidenciaban competición débil, mientras que en el dominio sintáctico se reforzaban entre sí. Por otro lado, la competición más fuerte se produjo entre la precisión léxica y la complejidad sintáctica.

Spoelman y Verspoor (2010) investigaron el desarrollo del sistema de casos en un estudiante holandés principiante de fines L2. A pesar de que el sistema de casos constituye un desafío de aprendizaje, la preci-

sión se mantuvo entre el 80-100%. La tasa de error superó el 10% solamente en el partitivo y el nominativo singular, notorios por su morfología no transparente y complejidad semántica. Hallaron dos etapas de adquisición, con aumento de variabilidad en la vecindad de una transición de fase de una etapa a la otra. Por un lado, la complejidad en el nivel de la palabra evidenció relaciones de mutuo refuerzo con la complejidad tanto a nivel del SN como de la oración. Por otro lado, la complejidad en la palabra estuvo en competición con la de la oración, estabilizándose al final.

Eskildsen (2012) estudió a dos aprendientes hispano hablantes de inglés L2 (Carlos y Valerio). Eran datos longitudinales de entre dos años y medio y tres años. Se focalizó en la adquisición de la negación. Las trayectorias de aprendizaje de ambos fueron diferentes. En el caso de Carlos, se caracterizó por: alta frecuencia de la fórmula «I don't know»; una competencia entre los patrones «SUJETO don't VERBO» (correcto) y «SUJETO no VERBO» (incorrecto). Valerio también exhibió alta frecuencia de «I don't know». No obstante ello y a diferencia de Carlos, al principio dominó el patrón «SUJETO don't VERBO por sobre «SUJETO no VERBO». Luego la dominancia se dio a la inversa; y al final el patrón «SUJETO no VERBO» bajó su frecuencia hasta casi desaparecer. En consecuencia, las trayectorias son específicas de los aprendientes y no se conforman a estadios «universales» de adquisición.

Baba y Nitta (2013) estudiaron la fluidez (cantidad de palabras) en producciones escritas con tiempo limitado (10 minutos). Siguió a dos estudiantes japoneses de inglés L2 durante un año académico. Hallaron evidencia de al menos una transición / bifurcación en cada estudiante según cuatro índices: (a) saltos repentinos de frecuencia en los datos; (b) aumento de la variabilidad cerca de la transición de fase; (c) presencia de un elemento de «catálisis» que empuja al sistema hacia un nuevo estadio; (d) cambios cualitativos en el modo de escribir. Los índices (a) y (b) se analizaron cuantitativamente y los otros cualitativamente. Por ejemplo, se observaron (a partir de comentarios de los alumnos para cada composición) como elementos catalizadores la auto-confianza y la facilidad del tópico elegido; en cuanto a cambios cualitativos, se notó, por ejemplo, que en la segunda etapa, se producían escritos con más párrafos y más sintéticos; o bien subdividiéndolos con subtópicos.

Lowie et al. (2014) también encontraron evidencia de fractalidad en el aprendizaje, detectando «ruido rosa». Por medio de un análisis espectral, la variabilidad fue analizada relacionando la magnitud de los cambios en el sistema con la frecuencia en la que éstos ocurren. Si todas las magnitudes de cambio ocurren con igual frecuencia, entonces el patrón es totalmente estocástico e impredecible, asociado al «ruido blanco». En el extremo opuesto, si la magnitud del cambio es completamente (inversamente) proporcional a la frecuencia de ocurrencia, se observa un comportamiento hiper-regulado o «ruido marrón». El «ruido rosa» es la situación óptima de organización u coordinación entre subsistemas (ya que permite la adaptación rápida), que se encuentra entre ambos extremos. Aquí *la relación de escala* de la magnitud del cambio alcanza una relación inversamente proporcional con la frecuencia del cambio: $\frac{1}{f}$. A lo largo de 6 años se estudiaron los tiempos de reacción de un individuo en una tarea para nombrar palabras en holandés (L1) e inglés (L2). Los resultados arrojaron, previsiblemente, más coordinación («ruido rosa») en L1 que en L2, pero también un aumento de la coordinación en L2 a medida que se incrementaba el nivel de competencia del aprendiente. Además luego de un período de práctica intensiva en la L2, los datos evidenciaban incluso una tendencia hacia el «ruido rosa», interpretado como incremento en la auto-organización del sistema de aprendizaje.

3.5.3. Relación con otros enfoques.

Siguiendo a Lourdes Ortega (2015) y a Larsen Freeman (2015) [ver también: VanPatten & Williams, 2015] se repasa cómo explica la teoría de Sistemas Dinámicos Complejos [SDC] la evidencia observada en segundas lenguas; y se lo compara con los enfoques descriptos en la sección 3.3.

3.5.3.1. La naturaleza del aprendizaje: incidental, implícito y que va más allá del input recibido.

Se ha reconocido que el aprendizaje en L2 es *incidental e implícito*. Los aprendientes extraen propiedades del input (morfología, sintaxis, léxico, etc.) mientras se hallan inmersos en actividades comunicativas focalizadas en el significado y no en la forma. Por otro lado, los aprendientes adquieren conocimiento

implícito de la lengua que no pudo provenir solamente de los datos del input, por ejemplo, sobre gramaticalidad, ambigüedad de oraciones, estructura de constituyentes (el input es lineal), patrones abstractos de construcciones.

La teoría de Sistemas Complejos en L2, al igual que el enfoque orientado al uso de construcciones, mantiene que el aprendizaje es implícito via mecanismos generales de asociación, inferencia estadística y extracción de regularidades distribucionales del input. Sin embargo, los aprendientes producen más de lo que estuvo presente en el input porque son creativos. Ellos adaptan los recursos de expresión de los que disponen según las oportunidades de acción [*affordances*] provistos por el contexto, mientras interactúan en tareas comunicativas. Por otro lado, al ser un proceso abierto, el aprendizaje no se dirige a un estadio «final» de conocimiento de «hablante nativo».

En *skill acquisition theory* también se sostiene que el aprendizaje procede utilizando mecanismos generales que se aplican también a otros tipos de conocimiento no lingüístico (ej. aprender a cocinar, aprender biología o aprender a andar en bicicleta). No obstante ello, se comienza necesariamente con conocimiento declarativo y explícito. Volverse un experto requerirá transformar el conocimiento del «que hacer» en conocimiento procedural de «cómo hacer», que sí es implícito.

El modelo declarativo / procedural concede roles diferentes para ambos tipos de sistema. El declarativo parece estar conectado con el aprendizaje en contexto formal, y especialmente en las primeras etapas. El procedural, en cambio, con el aprendizaje en contexto naturalístico y con etapas tardías de aprendizaje. Por otro lado, el hecho de que ambos sistemas compitan y se interfieran entre sí, implica que aprender *no* consiste en transformar conocimiento explícito en implícito.

En lo que respecta a la gramática universal, el conocimiento es implícito, innato y de naturaleza simbólica. Dicho conocimiento guía la adquisición de la L2. El aprendizaje es deductivo a partir de mecanismos y rasgos presentes en la GU.

3.5.3.2. La naturaleza de la inter-lengua: sistemática y con variabilidad dentro y entre subsistemas.

Se ha observado que la inter-lengua que producen los aprendientes es sistemática. Asimismo se siguen etapas de adquisición. Por ejemplo, en el caso de la negación en inglés se han evidenciado las etapas siguientes: (i) NO + SINTAGMA: *no want that*, (ii) SUJETO + NO + SINTAGMA: *He no want that*; (iii) SUJETO + NO/DON'T/CAN'T/NOT + SINTAGMA: *He can't/don't/not want that*; (iv) NEG + VERBO MODAL: *He can't do that*; (v) AUXILIAR + NEG: *He doesn't want that*. Asimismo, para ciertas estructuras se ha observado un orden particular de adquisición; por ejemplo, en morfología del inglés se constató el orden: presente progresivo [-ing] > pasado regular [-ed] > pasado irregular > tercera persona del presente [-s]. Por otro lado, la producción de los aprendientes resulta ser variable tanto respecto a un dominio o subsistema (p. e., morfología flexiva): *las casas / le casas / la casas / le casa / ...*; como entre subsistemas: se ha visto más variabilidad en morfología que en sintaxis.

Desde el punto de vista de los SDC, los aprendientes estarán más orientados a los atractores que ofrezcan más oportunidades de utilidad comunicativa. O sea que las formas de menos frecuencia, menos saliencia, más redundancia semántica y ambigüedad presumiblemente se adquirirán más tarde (por ejemplo, la tercera persona del presente en inglés). Sin embargo, las etapas de aprendizaje no se transitan linealmente como en una «escalera». Cada aprendiente persigue su propia trayectoria y percibe en el contexto diferentes oportunidades para el uso de construcciones. Por ello, no es esperable que se sigan órdenes fijos de adquisición. Por otra parte, la variabilidad de la *inter-lengua* permite notar diferencias entre la propia producción de un aprendiente y la del otro con quien se habla, promoviendo la co-adaptación de formas a la del otro. Asimismo, el aumento de la variabilidad es un indicador de una posible bifurcación. En consecuencia, es fundamental que no se la descarte como mero error de medición. Tanto en el enfoque basado en construcciones como en SDC, la sistematicidad *emerge* como producto de la auto-organización de los sistemas.

En lo que atañe al funcionalismo orientado a conceptos, el sendero *pragmática* > *léxico* > *morfología* ofrece en cada etapa mayor poder comunicativo. Cada etapa brinda mayor capacidad expresiva para comunicar conceptos y menos dependencia del interlocutor. El aprendiente busca eficiencia e independencia comunicativa. Por otra parte, las formas con mayor carga funcional se adquirirían antes. Tómese como

ejemplo el presente perfecto del inglés [*have gone*] contra el pluscuamperfecto [*had gone*]. Los datos muestran que el primero se adquiere antes. La teoría explicaría que es así porque no hay otra forma para expresar la misma función. En cambio, el significado de «evento anterior a otro» del pluscuamperfecto podría ser expresado echando mano de una forma de pasado junto con un adverbio [*Fue a hacer la compra y antes fue al banco*].

El objetivo mismo de la teoría de la procesabilidad radica en explicar etapas de aprendizaje. La sistematicidad de dichas etapas depende la capacidad del aprendiente para intercambiar información de rasgos en su memoria de trabajo, a partir de los mecanismos procedurales que logre manejar. El nivel de procesabilidad también delimita el espacio de hipótesis para la producción: las soluciones al problema de aprendizaje dependen del estadio procedural. Por ejemplo, si no se ha llegado a adquirir el procedimiento de Sintagma Oracional, no se podrán mover elementos entre sintagmas, y el hablante producirá variación como: *he has been where ?* [el-ha-estado-dónde ?], *where has been ?*, *where he has been?* y no: *where has he been ?*. Se nota que en los primeros tres casos el aprendiente evita mover el auxiliar.

En lo que respecta a la gramática universal, la sistematicidad constituye una propiedad del conocimiento lingüístico presente en la GU. Dado que la GU guía el aprendizaje en la L2, la inter-lengua también es sistemática y la variabilidad observada debe mantenerse dentro de las posibilidades permitidas por la GU misma.

3.5.3.3. El rol de la primera lengua.

Se ha observado que la influencia de la L1 varía con los individuos, los subsistemas y las etapas del aprendizaje. Si bien el *transfer* resulta más potente en las primeras etapas de la adquisición, hay subsistemas donde influye inclusive en periodos tardíos (por ejemplo, en la morfología).

En los SDC la L1 es sistema dinámico en sí mismo. El *transfer* implicaría que los rasgos de la L1 constituyen atractores fuertes. Para el enfoque orientado al uso de construcciones, los aprendientes tienen el foco de atención orientado por características de la L1. Esto redundaría en la capacidad de poder prestar atención a aspectos del input que son diferentes de los de la propia lengua materna. Las formas de mayor saliencia, frecuencia y suficientemente significativas lograrán escapar a este sesgo de atención. Otras quedarán fuera del foco para ser notadas por los aprendientes, quizás irreparablemente si no hay instrucción formal basada en la forma [*focus on form*]. Por ejemplo, en italiano la partícula «*ci*» conlleva una variedad de funciones. Entre ellas se encuentra la de expresar posesión para no repetir el objeto al que se hace referencia: A: *Hai una sigaretta ?* [¿tenés un cigarrillo?]; B: *no, non ce l'ho* [no, no tengo]. Si el objeto se topicaliza la partícula «*ci*» es obligatoria para referirse a éste: *una sigaretta, ce l'hai ?* [¿un cigarrillo, tenés ?]; pero no puede haber referencia «hacia adelante»: **ce l'hai una sigaretta ?*. Este error es muy común en hispano-hablantes que han aprendido italiano en contexto naturalístico. Aunque la partícula «*ci*» no exista en español, «*ce l'hai*» es una construcción muy frecuente y saliente para la percepción pero sus reglas de posición no lo son (porque no hay nada parecido en español). El foco de atención, orientado a la L1, se deposita en las «pistas» incorrectas.

La investigación en el marco de la gramática universal ha tratado de dilucidar si la GU guía la adquisición de la L2 como lo hace con la L1. Se atribuye a la L1 un rol fundamental como posible estado inicial de adquisición en las propuestas de acceso total o restringido a la GU. La cuestión radica en descubrir qué rasgos de la L1 se pueden transferir a la L2 y si sus valores son pasibles de reparametrización.

3.5.3.4. El rol del input y el output (producción).

La adquisición de una segunda lengua no tendrá lugar sin exposición a input. Con lo cual el input resulta necesario, sin embargo ¿es también suficiente para adquirirla?. La frecuencia constituye una característica del input. En muchos casos, se ha visto correlación entre mayor frecuencia y facilidad de adquisición. Sin embargo hay límites: a veces lo más frecuente se adquiere en etapas tardías. Por otra parte, existe evidencia a favor y en contra de que la producción de lenguaje en L2 logre promover la adquisición: la práctica no siempre hace al maestro.

El enfoque basado en las construcciones sostiene que la adquisición está guiada por el input. Cada vez que se encuentra un ejemplar en el input (oral o escrito) se refuerzan conexiones neuronales y se establecen

trazas en la memoria. De esta forma va emergiendo una red de asociaciones de construcciones abstractas. De modo inconsciente se computan las frecuencias relativas de las construcciones, los contextos en los que aparecen y los significados a los cuales se aparean. Sin embargo, el efecto de frecuencia está limitado. Se halla moderado por factores como la saliencia, efectos prototípicos del input y por la atención, cuyo foco está orientado hacia aspectos de la L1. En lo que respecta a la producción de lenguaje (output), éste aumenta la fluidez y refuerza conexiones en la red asociativa de construcciones, promoviendo la automatización.

La teoría de los SDC comparte lo apuntado pero no se considera que el input sea el «material de aprendizaje» puesto a disposición del aprendiente. El hablante no es un simple receptor pasivo de input. Larsen Freeman (2015) lo llama «ambient language». SDC se centra en el aprendiente. Por eso, es aquél el que debe percibir aspectos de su *ambient language* que considere como oportunidades de aprendizaje (*affordances*). Los hablantes son «hacedores de significado», ven en el ambiente del lenguaje lo que les sirve para expandir el potencial comunicativo de la *inter-lengua*. Lo que es crucial es su percepción del input como oportunidad para actuar, o sea, con el objetivo de producir involucrándose en tareas comunicativas. Esto conlleva implicancias fundamentales. No es que las características del input (frecuencia, saliencia, etc.) promuevan la adquisición. Resulta esencial que el aprendiente *note* que una forma frecuente del input lo ayudará a mejorar. Lo que *no* se perciba en el *ambient language* como oportunidad para la acción, aunque sea frecuente, constituye «ruido». En consecuencia, lo más frecuente, saliente, menos ambiguo, etc. no necesariamente se adquiere antes. Por otra parte, en los SDC los sistemas son iterativos: lo que resulta de una iteración se usa como punto de partida para la siguiente. Lo que promueve la adquisición es la repetición de los mismos eventos de uso del lenguaje a lo largo de tareas comunicativas diferentes. La repetición no es vista como replicación exacta sino como mecanismo de generación de variabilidad (Larsen Freeman, 2012). No repetir el mismo evento de lenguaje de la misma forma crea más opciones como recursos para expresión de significados. Cada actividad comunicativa conlleva sus propios objetivos. La repetición de construcciones en actividades diferentes permite que el aprendiente las adapte a los nuevos objetivos comunicativos. En suma, la iteración, al generar variabilidad, provee de opciones para adaptarse a nuevos contextos y satisfacer diferentes necesidades comunicativas.

En lo que atañe a la *skill acquisition theory*, el input es solamente un ingrediente más de la adquisición. Para promover la automatización, la exposición repetida y el input frecuente necesitan de la conjunción con la práctica deliberada (output), la cual pone en acción al conocimiento declarativo. Se adquirirían antes las formas más frecuentes del input que se practiquen más.

Respecto del modelo declarativo / procedural, la exposición repetida y la alta frecuencia son necesarias para la memoria procedural, pero con lento aprendizaje. En cambio, la baja frecuencia del input se asocia al aprendizaje declarativo, de rápido aprendizaje, inclusive con la exposición a una sola forma. Puesto de otro modo, el aprendizaje via memoria procedural es dependiente de la frecuencia pero el que se hace a través de la memoria declarativa no lo es.

Las propiedades estadísticas y distribucionales del input también se consideran de importancia en el funcionalismo orientado a conceptos. La frecuencia relativa de los pares forma-función predeciría la dirección en la que los aprendientes expenden su repertorio de formas lingüísticas. En Andersen & Shirai (1994) se afirma que los aprendientes siguen una hipótesis por sesgo distribucional. Por ejemplo, el progresivo *-ing* del inglés aparece en el input con más frecuencia asociado a eventos durativos atéllicos [*run-ing every day* (correr todos los días), *sing-ing at church* (cantar en la iglesia)]. Por consiguiente, dicha forma aparecerá primero en enunciados que contengan este tipo de eventos y en una segunda etapa se expandirá a eventos de duración télica [*run-ing a marathon* (correr una maratón); *sing-ing a lullaby* (cantar una canción de cuna)].

Desde el punto de vista de la teoría de la procesabilidad, no es el input *per se* o sus características lo crucial, sino el procesamiento que se hace de éste. Los mecanismos procedurales son los que permiten que se realice el intercambio de información en la memoria a corto plazo; es decir que se procesen allí rasgos del input. El output se considera un mero reflejo de la capacidad de procesabilidad: lo que mejor se procesa se genera de modo más sistemático. El procesamiento causa la producción y no al revés.

Por último, en la gramática universal el input activa el mecanismo deductivo de (re)parametrización de la GU. La frecuencia y el output no constituyen constructos teóricos.

3.5.3.5. El rol de la instrucción formal.

En lo que atañe al enfoque de la Gramática Universal, la instrucción no juega un papel importante. Si bien muchos aspectos del conocimiento del lenguaje se pueden enseñar (vocabulario, pragmática, preferencias de estilo), las propiedades morfo-sintácticas que forman parte de la GU no lo son, y se adquieren de modo implícito.

En la visión de la teoría de la procesabilidad se debería enseñar solamente aquello que el aprendiente está en condiciones de procesar dado el nivel procedural en el cual se encuentra. Es decir que la instrucción formal no causa la adquisición sino que la acompaña; incluso puede perjudicarla si se enseñan contenidos que van más allá de la capacidad de procesabilidad del alumno.

Para el enfoque funcionalista orientado a conceptos la adquisición es promovida por la necesidad de ser comunicativamente más eficiente. En la clase los estudiantes reciben un input “pre-procesado” y el profesor es fuente de “corrección”. Esto lleva a reducir la presión comunicativa de un contexto naturalista de adquisición, en el cual el aprendiente debe modificar continuamente su repertorio lingüístico para hacer frente a dicha presión. Son las fuerzas comunicativas lo importante.

En lo concerniente al modelo declarativo-procedural, el aprendizaje vía la memoria declarativa (con instrucción formal explícita) es fundamentalmente diferente del aprendizaje vía memoria procedural (con instrucción formal implícita) y se interfieren mutuamente sin existir interfaz entre ellos (lo declarativo no se automatiza con la práctica).

Desde la perspectiva de *Skill Acquisition Theory* la instrucción debe consistir en ciclos de enseñanza declarativa explícita de reglas, acompañadas con abundantes ejemplos; y su posterior puesta en práctica para lograr la automatización de la habilidad a adquirir. A diferencia del enfoque anterior, lo declarativo se convierte en procedural a través de dichos ciclos de instrucción debidamente implementados.

Con respecto al enfoque orientado al uso de las construcciones, el aprendizaje procede implícitamente por inducción de patrones distribucionales y estadísticos del input. Con lo cual, resulta fundamental que en la clase dicho input sea lo más auténtico posible, rico y abundante. Además, la instrucción debe utilizar *feedback* implícito y/o explícito [ver capítulo 12] para que los alumnos logren notar patrones poco frecuentes o de reducida saliencia. Se considera imprescindible modificar el foco de atención que los alumnos traen con su L1, que les impide darse cuenta de determinados aspectos del input de la L2.

La teoría de los sistemas dinámicos complejos acuerda con el enfoque precedente pero poniendo siempre el foco en el alumno y su particular trayectoria de aprendizaje. Lo fundamental son las oportunidades de aprendizaje [*affordances*] que el alumno percibe en el ambiente (contexto comunicativo), incluido el de la instrucción formal. Ésta conllevará beneficios si: (i) se centra en la agentividad y creatividad del alumno; (ii) considera que la motivación central es la creación de significados; (iii) pone el foco en los objetivos de los propios alumnos y sus propias percepciones de oportunidades de aprendizaje; (iv) favorece que los alumnos noten las diferencias entre sus propias interlenguas; (v) promueve la repetición para generar variabilidad y construir de este modo opciones para adaptarse a nuevos contextos.

El Cuadro siguiente resume las características de la teoría de los SDC en L2.

Conceptos fundamentales	heterogeneidad; dinamismo; falta de anidamiento; proceso iterativo; no linealidad; interacción con el ambiente; fractalidad
Metodología	estudios de caso; diseño longitudinal; series no estacionarias; análisis de trayectorias; métodos no lineales
Naturaleza del Aprendizaje	implícito: mecanismos generales de asociación, inferencia estadística y extracción de regularidades distribucionales del input. creatividad. adaptación de los recursos expresivos según oportunidades de aprendizaje Sensibilidad a las condiciones iniciales. Sin estadio final «nativo».
Interlengua	aumento de variabilidad individual marca bifurcaciones; permite notar diferencias; sistematicidad emerge por auto-organización
L1 (transfer)	rasgos de la L1 son atractores fuertes; atención está «afinada» por la L1 (bloqueo)
Input	aprendizaje asociativo guiado por el input pero también por las percepciones de los aprendientes en el «ambient language» [<i>affordances</i>]
Frecuencia	efectos moderados por saliencia, efectos prototípicos y atención afinada por L1 importante pero mediada por la agentividad del alumno y co-adaptación
Output	promueve fluidez y automatismo; importancia de la repetición como mecanismo generador de variabilidad; para adaptarse en actividades comunicativas
Instrucción formal	focus on form; <i>feedback</i> para «notar» diferencias; centrado en la agentividad del alumno; en sus propios objetivos y percepciones de <i>affordances</i> uso iterativo del lenguaje en diferentes actividades para generar variabilidad

Table 3.1: La teoría de los Sistemas Dinámicos Complejos en L2.

3.6. La concordancia como sistema dinámico.

3.6.1. Un modelo para L1.

Smith, Frank & Tabor (2018) propusieron un modelo para el fenómeno de atracción en la concordancia sujeto-verbo en L1, en inglés. Teniendo un sujeto del tipo «N1 de N2», siendo N1 su núcleo, el verbo a veces concuerda en número con N2 en lugar de N1. Se focalizaron en las construcciones pseudo-partitivas del tipo *una caja [N1] de chocolates [N2]*. Se clasificó dichas construcciones según el tipo de controlador de N1: (i) contenedores: *un frasco de píldoras, una caja de bombones*; (ii) colecciones: *un mazo de cartas, un racimo de flores*; (iii) medidas: *un kilo de manzanas, un montón de libros*. También se caracterizó a los controladores según los rasgos: (a) \pm *contendor*; (b) \pm *configuración espacial*; (c) \pm *N2 abstracto*. Los N1 contenedores conllevan referencia autónoma: no es necesario saber qué contienen para identificar su referente. En cambio, las colecciones indican un grupo de elementos denotados por el N2, organizados según una determinada configuración espacial. Por lo tanto, el N2 es necesario para la referencia. Las cantidades ni siquiera indican un grupo, sino una cantidad indeterminada. Se trata de operadores abstractos del significado de la N2; es decir, tienen una dependencia máxima de los N2. Este estatus de mayor abstracción de las medidas hace que puedan tener como N2 tanto sustantivos concretos como abstractos (*un montón de diarios / ideas; un mazo de flores / *ideas; una caja de herramientas / *ideas*). La combinación de estos rasgos determina una escala de «buenos sujetos» a «malos sujetos» para la concordancia con N1, según: contenedores > colecciones > medidas. El modelo intentó simular los resultados del experimento siguiente, donde se utilizó una tarea de completamiento. Se le presentaban al sujeto palabras de una oración en una pantalla de modo serial, previas a la presentación del verbo. Luego aparecía el verbo en forma tanto singular como plural y el participante debía elegir entre ambas opciones presionando un botón. Las condiciones fueron: (i) tipo de N1: contenedores, colecciones, medidas y cuantificadores (este último se agregó como control) y (ii) presencia o ausencia de N2. Se llevó a cabo un diseño de factores cruzados 2×4 . Una instancia típica fue: (a) condición: $+N2 \times contenedor$:

Do they have anything to light these candles with?

[¿tienen ellos algo para encender estas velas?]
 A pack / of / matches / [is / are] / in the / game room
 [la cajita de fósforos [está/n] en la sala de juegos]

(b) condición: $-N2 \times \text{contenedor}$:

Do they have anything, like matches, to light these candles with?
 [¿tienen ellos algo para encender estas velas, por ejemplo fósforos?]
 A pack / of / matches / [is / are] / in the / game room
 [la cajita [está/n] en la sala de juegos]

Las unidades de representación del modelo son *treelets* sintácticos organizados en una gramática de dependencias. Cada *treelet* está anclado a una palabra y posee un vector de rasgos para la palabra y otro para la dependencia esperada de ésta. Asimismo, cada *treelet* puede tomar como dependencia a otro *treelet* o ser dependiente de otro. Las dependencias se representan mediante enlaces. Los enlaces compiten unos con otros respecto de su nivel de activación. Si una medida de distancia entre los vectores de rasgos entre el origen del enlace y su sitio de llegada es alta, la tasa de crecimiento de la activación será más fuerte que si la distancia fuera más baja. Además, el nivel de activación se halla sujeto a ruido (estocasticidad). Si dos sitios de llegada de un enlace arrojaran distancia parecida, el sistema se estabilizaría en una determinada configuración que involucraría a alguno de los dos sitios de llegada, debido a la influencia del ruido. El siguiente Cuadro muestra los vectores de rasgos asignados a cada sitio de llegada del enlace:

		rasgos:				
	tipo de N1:	N	a	b	c	N2 presente
N1	Contenedores	1	1	1	0	1
	Colecciones	1	0	1	0	1
	Medidas	1	0	0	1	1
«de»	dependencias	1	NA	NA	NA	1= +N2 ; 0 = -N2
N2	núcleo	1	1	1	0	1= +N2 ; 0 = -N2
Verbo	dependencias	1	1	1	0	1

Cuadro 3.2: Vectores de rasgos para cada sitio de llegada de un enlace. Rasgos: N = «sustantivo»; a = «contenedor»; b = «configuración espacial»; c = «N2 abstracto»; NA = dato no disponible (no se usó para calcular la distancia).

La Figura 6 ilustra el modelo. Los atractores son N1 o N2. Visto que el N1 es siempre singular; y el N2, siempre plural; ambos atractores representan la concordancia singular o plural del verbo, respectivamente. En lo que refiera a la configuración con N1 como atractor, «de» es el núcleo de un SP, que es dependiente del núcleo de la NP «caja» [N1] y que a su vez tiene como dependiente a «manzanas» como núcleo [N2] de la NP. Además, el NP del S tiene un enlace cuyo sitio de llegada es el NP de «caja». Por otro lado, en la configuración con N2 como atractor, «de» posee como dependiente a «caja» y depende a su vez de «manzanas», cuyo NP es sitio de llegada del NP del S. Los autores utilizaron una dinámica guiada por un modelo de *Lokta-volterra* generalizado (ver cap. 9). El modelo es una generalización del modelo logístico incluyendo competencia entre especies. Aquí las «especies» son los seis enlaces que compiten entre sí para terminar estabilizándose en una configuración compatible con N1 como atractor o como N2 como atractor. Se describe mediante la siguiente ecuación diferencial estocástica, donde r_i es la tasa de crecimiento del i -ésimo enlace; x_i su nivel de activación; W_{ij} es una matriz con pesos que modelan la fuerza de la competencia entre los enlaces y η es el término estocástico («ruido»). La tasa de crecimiento de cada enlace se determina mediante la distancia de Hamming, que cuenta la cantidad de rasgos que difieren entre dos vectores: $r_i = e^{-\text{distancia}}$.

$$\frac{dx_i}{dt} = r_i x_i \left(1 - \sum_{j=1}^{n=6} W_{ij} x_j \right) + \eta$$

En la Figura 6 se ilustra el estado del sistema próximo al atractor N1, donde todos los enlaces que pertenecen a la primera configuración alcanzan máxima activación (en negro). En contraste, los enlaces en rojo punteado tienen activación mínima (configuración con N2). El modelo logró resultados similares a las proporciones de concordancia plural y singular hallados en el experimento. Como en éste, la simulación produjo más configuraciones con N2 con la condición $+N2$ que con la de $-N2$. Además se replicó el hallazgo de que la probabilidad de configuraciones con N2 crecía según los tipos de N1: contenedores > colecciones > medidas. Vale la pena resaltar que se trata de un modelo continuo cuyo resultado es un estado simbólico discreto: una concordancia plural o singular. Los modelos de sistemas dinámicos no procesan manipulando símbolos sino mediante *configuraciones de estados continuos* (de activación «mental») que se estabilizan a medida que pasa el tiempo. El estado final «concordancia singular» o «concordancia plural» es el producto de la auto-organización del sistema. Dicho de otro modo, la concordancia no se produce por aplicación de reglas simbólicas pre-existentes (o por unificación de rasgos), sino que constituye un fenómeno que *emerge* de la dinámica misma.

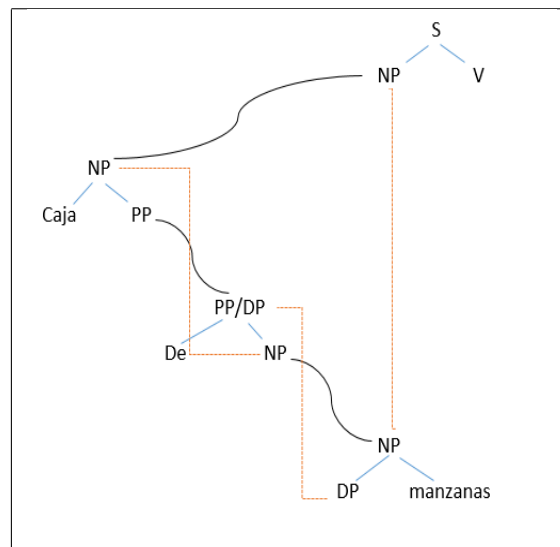


Figura 3.6: Representación del modelo de *treelets* [adaptado de Smith, Frank & Tabor (2018)]. En rojo (punteado): enlaces de la configuración con atractor N2 con activación mínima. En negro (liso): enlaces de la configuración con atractor N1 con activación máxima.

3.6.2. El error de concordancia en L2 como atractor. Estrategia de análisis.

En esta tesis se propone ver a las concordancias «correctas» e «incorrectas» como dos atractores. Se analizará la dinámica a nivel microscópico y macroscópico. En el primer caso se intentará modelar la dinámica a nivel de trayectorias de «activación mental» que lleven el flujo hacia alguno de los atractores mencionados. Cada punto en el estado de fase es un vector que describe la configuración de activación «mental». Se considerará que inicialmente el aprendiz se encuentra en un estado mental «mixto» entre equivocarse y no equivocarse; o sea, en un atractor inestable del cual se debe alejar para entrar en el *basin of attraction* de alguno de los atractores mencionados. En términos de la representación de potenciales de la sección 4, se comenzaría en la «cima de la montaña», pudiendo descender hacia alguno de los dos «valles». Diversas características de las concordancias (el tipo de «objetivo», características del «controlador», si es a larga distancia, etc.) determinarán hacia cuál *basin of attraction* habrá inclinación, lo cual indicará a su vez qué atractor resultará ganador. Esta dinámica se producirá para *cada* instancia de concordancia producida por el aprendiz. Análogamente al modelo del apartado anterior, «concordancia correcta» y «concordancia incorrecta» son dos símbolos discretos que emergen de una dinámica continua. Si en cada instancia de concordancia se aplica la dinámica, colectivamente emergerá un fenómeno de regularización: la extinción paulatina de los errores si hay aprendizaje y la difusión de las instancias correctas. Dicho fenómeno se visualizará en dos patrones: (i) proporción global de error y (ii) serie de tiempo de proporción de errores a través de las sesiones. Con el propósito de identificar las variables pertinentes que permitan

identificar la pertenencia al *basin of attraction* de un determinado atractor, se aplicarán modelos estadísticos utilizando como variable respuesta: (i) estatus de error (si / no) ; (ii) tipo de error ; (iii) tiempo hasta que se produce el error. Se propondrán tres modelos simples para la dinámica del error. Las variables identificadas formarán parte de un mecanismo de distancia a configuraciones de variables con chance de error baja, que permitirá «romper» el estado mixto inicial e inclinar el flujo hacia alguno de los atractores. Se realizarán simulaciones para aproximar los patrones observados de error global y por sesiones.

En el nivel macroscópico, la trayectoria estará compuesta por la frecuencia de error por sesión. En este nivel el objetivo consistirá en detectar posibles transiciones entre regímenes de errores, sesiones de dinámica de error similar y aquellas con dinámica diferente. Se aplicará una serie de medidas para determinar si hay evidencia de cambio de dinámica; en una palabra, si existe promoción de la adquisición de la concordancia.

Otra forma de estudiar el fenómeno macroscópicamente es considerar una red de concordancias que evoluciona en el tiempo, con cada sesión. Cada nodo de la red es una palabra, con determinados atributos a definir, y cada enlace indica si se establece una relación de concordancia (correcta o no) con otra palabra. Un modo de ver cambios en la dinámica consiste en estudiar la forma en la cual las características globales de la red cambian con paso de las sesiones. Otra estrategia se enfoca en analizar la evolución de la redes de crecimiento de error (sin considerar las concordancias correctas). Se aplicará una técnica de análisis estadístico longitudinal de redes. Se examinará si el aumento de enlaces de error está asociado a alguno de los nodos que participan del enlace o a ambos, según un determinado atributo de los nodos. O bien se verificará si el crecimiento del error depende globalmente de algún atributo determinado.

Por último, se tomarán atributos de las concordancias definidas para el análisis microscópico, atributos dinámicos del análisis de trayectoria y atributos de las redes complejas. Se aplicarán métodos de aprendizaje automático para intentar predecir el error en cada instancia de las últimas cuatro sesiones de cada alumno.

Para finalizar, se comentarán ciertos aspectos didácticos respecto de la concordancia desde el enfoque de los SDC. Luego se pasará a la discusión con lo hallado en la literatura (ver cap. 1) y a las conclusiones.

El Cuadro que sigue muestra la estrategia de análisis para los capítulos técnicos que vendrán.

Nivel de análisis	Capítulo	Objetivo	Técnicas empleadas
	cap 4	Codificación del corpus y creación de variables	Técnicas para datos faltantes; colinealidad; discretización; transformaciones
	cap 5	Análisis descriptivo	Estadística descriptiva, Análisis de correspondencias; técnicas de selección de atributos
Microscópico	cap 6	Variables asociadas al estatus de error	Selección de modelos. Regresión logística mixta.
	cap 7	Variables asociadas al tipo de error	Regresión multinomial bayesiana.
	cap 8	Variables asociadas al tiempo hasta el error	Modelos de tiempo hasta que se produce un evento
	cap 9	Simulación con sistemas dinámicos	Modelo de Lokta-Volterra; modelo de teoría de los juegos evolutiva; modelo de computación simbólica.
Macroscópico	cap 10	Detección de transiciones	Plano entropía / complejidad, LDA; <i>Change point</i> ; Recurrence Quantification analysis (RQA); Indicadores de cambio de régimen
	cap 11	Cambio de dinámica en redes de concordancia	Redes complejas; Actor-Based Stochastic Model
		Predicción del error	Técnicas de minería de datos (clasificadores); Clustering con distribuciones; Lasso mixto y GLM mixto

Cuadro 3.3: Estrategia de análisis.

3.7. Apéndice de capítulo.

3.7.1. Análisis del modelo logístico con «depredación».

El modelo logístico con depredación es: $\dot{x} = rx \left(1 - \frac{x}{K}\right) - qx \Leftrightarrow \dot{x} = f(x)$.

Se reconocen las funciones $g_1(x) = qx$ y $g_2(x) = rx \left(1 - \frac{x}{K}\right)$. La condición de bifurcación es dichas funciones sean tangentes y se toquen en un solo punto. Esto es lo mismo que pedir que:

$$\frac{dg_1(x)}{dx} = \frac{dg_2(x)}{dx} \Leftrightarrow \frac{dg_1(x)}{dx} - \frac{dg_2(x)}{dx} = 0 \Leftrightarrow \frac{d}{dx} [g_1(x) - g_2(x)] \Leftrightarrow \frac{df(x)}{dx}$$

Ahora bien: $g_1'(x) = q$ y $g_2'(x) = r - \frac{2rx}{K}$. Por lo tanto:

$$g_1'(x) = g_2'(x) \Leftrightarrow g_1'(x) - g_2'(x) = 0 \Leftrightarrow q - \left(r - \frac{2rx}{K}\right) = 0$$

Es necesario despejar q_c , el valor crítico del parámetro dinámico en el cual sucede la bifurcación. Con lo cual:

$q_c - r + \frac{2rx}{K} = 0 \Leftrightarrow q_c = r - \frac{2rx}{K} = r \left(1 - \frac{2x}{K}\right)$. Ahora se reemplaza $q = q_c$ en $\dot{x} = 0$. Con lo cual se tiene:

$$rx \left(1 - \frac{x}{K}\right) - q_c x = 0 \Leftrightarrow rx \left(1 - \frac{x}{K}\right) - rx \left(1 - \frac{2x}{K}\right) = 0 \Leftrightarrow rx \left[1 - \frac{x}{K} - 1 + \frac{2x}{K}\right] = rx \left(\frac{x}{K}\right) = \frac{rx^2}{K} = 0$$

Por lo tanto, $x = 0$ y sustituyendo en q_c : $q_c = r \left(1 - \frac{2(x=0)}{K}\right) \Leftrightarrow q_c = r$. El valor del punto crítico es r , donde sucede la bifurcación. Ahora es preciso realizar el análisis de estabilidad para $q < r$ y $q > r$. Los puntos fijos son los que hacen que $f(x^*) = 0$. Y son: $x^* = 0$ y $x^* = K \left(1 - \frac{q}{r}\right)$.

En efecto, la derivada de la función es $f'(x) = r - \frac{2rx}{K} - q$. Evaluada en $x^* = 0$, es $f'(0) = r - q$. Con $q < r$, $f'(0) > 0$, y por lo tanto es repulsor. En cambio, con $q > r$, $f'(0) < 0$, y por lo tanto es un atractor.

Evaluando la derivada en $x^* = K \left(1 - \frac{q}{r}\right)$, se tiene: $f' \left(K \left(1 - \frac{q}{r}\right)\right) = r - 2r \left(1 - \frac{q}{r}\right) - q = r - 2r + 2q - q = -r + q$.

Con $q < r$, $f' \left(K \left(1 - \frac{q}{r}\right)\right) < 0$, y por lo tanto es un atractor. En cambio, con $q > r$, $f' \left(K \left(1 - \frac{q}{r}\right)\right) > 0$, y por lo tanto es un repulsor.

En resumen, con $q < r$, hay dos puntos fijos, uno repulsor ($x^* = 0$) y el otro atractor. Cuando $q = r$ se produce la bifurcación. Luego, con $q > r$, los puntos fijos intercambian el estatus de estabilidad: el origen pasa a ser atractor y el otro pasa a ser repulsor. Ha ocurrido una bifurcación transcítica del tipo: $rx - x^2$. ¿Cómo es posible ver que la ecuación inicial efectivamente pertenece a esta forma canónica? Una posibilidad es adimensionalizar la ecuación original para llevarla a otra de dinámica equivalente pero con menos parámetros. Para hacer eso, sea: $\tau = rt \Leftrightarrow t = \frac{\tau}{r}$ y $u = \frac{x}{K} \rightarrow x = Ku$ y $\alpha = \frac{q}{r}$. Con lo cual:

$$\begin{aligned} \frac{dx}{dt} = rx \left(1 - \frac{x}{K}\right) - qx &\Leftrightarrow \frac{d(Ku)}{d\left(\frac{\tau}{r}\right)} = rx(1-u) - q(Ku) \Leftrightarrow \\ \frac{rKd(u)}{d(\tau)} = rx(1-u) - q(Ku) &\Leftrightarrow \frac{d(u)}{d(\tau)} = \frac{r}{r} \frac{x}{K} (1-u) - \frac{q}{rK} (Ku) \end{aligned}$$

En consecuencia, se llega a la forma equivalente:

$$\frac{du}{d\tau} = u(1-u) - \alpha u = u - u^2 - \alpha u = (1-\alpha)u - u^2$$

Llamando $\rho = 1 - \alpha$. Se obtiene la forma canónica $\rho u - u^2$, equivalente a la de la bifurcación transcítica.

Capítulo 4

Recolección de datos y creación de variables.

El capítulo describe el proceso de recolección de datos; la posterior creación de atributos y de variables respuesta, así como su posterior preprocesamiento: datos faltantes, problemas de colinealidad, discretización y transformaciones de escala.

4.1. Recolección de datos.

La presente tesis analiza cuatro casos de estudiantes de español como lengua extranjera. Se trató de cuatro alumnos adultos, de lengua nativa italiana, estudiantes del Instituto Cervantes de Milano en el año académico 2008/09. Cada alumno poseía un nivel distinto de competencia lingüística (según el Marco Común Europeo de Referencia). El siguiente Cuadro ilustra las características de los sujetos de estudio (los nombres son ficticios para preservar la identidad de los sujetos).

Alumno	Nivel (código)	Profesión	L1	Horas de español previas	Otras L2
SONIA	A1/A2 (1)	Investigadora	italiano	0 horas	alemán – inglés
NATI	B1 (2)	Agente de seguros	italiano	120 horas	inglés – francés
JAKO	B2 (3)	Estudiante de Sociología	italiano	240 horas	inglés
MIRKA	C1 (4)	Programadora	italiano	360 horas	inglés - japonés

Cuadro 4.1: Características de los sujetos del estudio.

Se hicieron entrevistas de 30 minutos entre el alumno y el investigador (autor de esta tesis). La tarea consistió en una conversación no estructurada, sobre temas acordes al nivel de competencia del sujeto. Dichas entrevistas tuvieron lugar aproximadamente cada 20 días, según la disponibilidad de los alumnos. Cada alumno realizaba simultáneamente el curso de español del nivel descrito en el Cuadro 1. Hubo entre doce y catorce entrevistas por alumno. El curso era anual; por ende, se realizaron al menos seis entrevistas por semestre. Los Cuadros que siguen muestran las fechas de las entrevistas. Éstas se nombran con el código asignado de transcripción.

Sonia1	25-10-08	Sonia7	17-02-09	Mirka1	27-10-08	Mirka7	29-01-09
Sonia2	7-11-08	Sonia8	3-03-09	Mirka2	10-11-08	Mirka8	12-02-09
Sonia3	20-12-08	Sonia9	20-03-09	Mirka3	24-11-08	Mirka9	26-02-09
Sonia4	2-1-09	Sonia10	10-04-09	Mirka4	4-12-08	Mirka10	13-03-09
Sonia5	17-1-09	Sonia11	25-04-09	Mirka5	19-12-08	Mirka11	27-03-09
Sonia6	2-2-09	Sonia12	15-05-09	Mirka6	15-1-09	Mirka12	16-04-09

Cuadro 4.2: Fechas de las sesiones: SONIA, MIRKA

Nati1	21-10-08	Nati8	10-02-09	Jako1	22-10-08	Jako8	11-02-09
Nati2	4-11-08	Nati9	24-02-09	Jako2	5-11-08	Jako9	25-02-09
Nati3	18-11-08	Nati10	11-03-09	Jako3	19-11-08	Jako10	12-03-09
Nati4	2-12-08	Nati11	25-03-09	Jako4	3-12-08	Jako11	26-03-09
Nati5	17-12-08	Nati12	14-04-09	Jako5	18-12-08	Jako12	15-04-09
Nati6	13-1-09	Nati13	28-04-09	Jako6	14-1-09	Jako13	29-04-09
Nati7	27-01-09	Nati14	12-05-09	Jako7	28-01-09	Jako14	14-05-09

Cuadro 4.3: Fechas de las sesiones: NATI, JAKO

4.2. Transcripción y codificación del corpus.

El *corpus* está constituido por los siguientes conjuntos de transcripciones : SONIA (nivel A1/A2): 12 transcripciones; NATI (nivel B1): 14 transcripciones; JAKO (nivel B2): 14 transcripciones; MIRKA (nivel C1): 12 transcripciones.

La codificación y transcripción de los datos se hizo mediante el formato CHAT, siguiendo a Mac Whinney (2010); y el conteo, usando el paquete CLAN. A continuación, el Cuadro detalla las convenciones de transcripción utilizadas. El signo «\» solamente separa ejemplos y carece de significado en la codificación. En negrita se resalta el uso de cada marcador.

Marcador	Significado	Ejemplo
(.) (?) (!)	Fin de enunciado	cómo es la gente? \ es gente de montaña. \ no!
www	Diálogo no importante para la investigación	*STU: www . %sit: el estudiante se levanta un momento y habla con otra persona. Vuelve luego de 1 minuto.
()	Parte de palabra no producida pero adivinable	alema(nos) [//] alemanos +//.
[//]	Reformulación	cada barrio dovrebbe@s:it [//] debería [//] tiene que +...
[/]	Repetición	hablan más o menos sobre un libro [/] un libro.
(.) (.)	Pausa	no, siempre en Milán. \ problemático (.) creo que está [*] problemático .
@s:'lengua'	Palabra en otra lengua	und@s:de [//] e@s:it Berlino@s:it +...
[*]	Error pero no de concordancia	*STU: las primeras veces yo tengo lecciones@n [*]. %err: las primeras veces yo imparto las lecciones
['*código']	Error de concordancia	hay muchas calles grandas [*3].
xxx	segmento inaudible	la belleza@s:it del xxx .
@i	Exclamación	ahh@i, es muy diferente que Benisol.
+...	Continuación de enunciado	tu hai@s:it [//] tu tienes el mar +...
+//.	Interrupción de enunciado	muchos ehhh@fp +//.
@n	Neologismo (cognado entre L1 y L2)	pues yo creo de recordar@n la Liguria
ehh@fp	Pausa del tipo «Filler»	no es que los italianos ehh@fp al contrario +//.
+/.	Interrupción por otro hablante	*STU: Milano significa Torino, significa +/.
+	Enunciado continua luego de interrupción	*INV: Venecia. *STU: +, Venecia, significa Génova.

Cuadro 4.4: Convenciones de transcripción.

Un archivo estándar en formato CHAT tiene tres secciones principales:

(I) Sección de encabezados: comienzan siempre con el símbolo «@» (el cual señala al programa CLAN que la línea no contiene datos lingüísticos sino meta-información), y son porciones de texto que dan información general sobre la sesión transcrita; por ejemplo: los participantes, el contexto de la entrevista, etc. Los encabezados más usados son:

(a) *@Begin*: se pone siempre al principio de la transcripción y no le sigue ninguna información adicional;

(b) *@Languages*: la(s) lengua(s) principal(es) del texto (en código ISO), aquí: «spa»;

(c) *Participants*: un código de tres letras mayúsculas que especifica las personas que participarán de la interacción. A dichos actores se los identifica mediante:

(1) *ID: nickname* que identificará a los hablantes; debe ser único puesto que se repetirá en cada línea de la transcripción (y es útil para hacer búsquedas en los archivos), los dos participantes son *STU* e *INV*; (2) Nombre del hablante: puede omitirse o usarse un *nickname*; aquí se usarán SONIA / NATI / JAKO / MIRKA para los sujetos y PABLO, el nombre real para el investigador; (3) Roles de los participantes: aquí «Participant» e «Investigator».

(d) *@ID*: especifica información adicional sobre los participantes (debe haber un encabezado *ID* para cada uno), en las transcripciones se usaron los campos siguientes: (1) lengua: repite el código de (b), (2) Nombre del Corpus (Conjunto de transcripciones de cada sujeto) en minúscula, por ejemplo «Sonia», (3) Sexo del participante, (4) Rol del participante: como en (c).

(e) Por último, el encabezado *@End* indica el final de la transcripción en la última línea.

(II) Sección de enunciados: identifica la persona que toma la palabra y contiene los datos lingüísticos. Cada línea inicia con un asterisco (*) seguido del código de tres letras (mayúsculas) que identifica al hablante (ya especificado en la sección de encabezados), luego le siguen dos puntos y el tabulador (sin espacios en blanco para teclear). Cada línea *principal* contiene palabras, espacios y los marcadores detallados más arriba. Además dichas líneas nunca comienzan en mayúscula, aun si se usa un nombre propio (las mayúsculas solamente se usan cuando se debe identificar a un nombre propio, por ejemplo «EEUU»). La línea principal contiene también los marcadores de tipo de error de concordancia «[*'código']» y de error de otro tipo «[*]» detallados en el Cuadro de más arriba.

(III) : Sección de dependencias: contiene información adicional que el investigador desea resaltar. Comienza con el símbolo «%» seguido de un código de tres letras (en minúscula) que indica la información que se encontrará allí. Aquí se usarán los códigos: (a) «%err»: para especificar la forma correcta de los errores (de concordancia y de cualquier otro tipo) marcados en la línea principal; (b) «%sit»: para comentar información relevante sobre el contexto de la conversación, por ejemplo: «el teléfono suena y el participante atiende».

El siguiente constituye un ejemplo de transcripción del aprendiente del nivel A1/A2 (SONIA):

```

1      @Begin
2      @Languages: spa
3      @Participants: STU Sonia Participant, INV Pablo Investigator
4      @ID: spa | sonia | |lfemale| |Participant| |
5      @ID: spa | sonia | |lmales| |Investigator| |
6      *INV: hálame de tus amigos .
7      *STU: yo tengo muchos amigos [*0] .
8      *STU: Marina es fotógrafa .
9      *STU: ella quiere hacer fotos .
10     *STU: y para hacer fotos ella va a Londres (.) a París (.) a Madrid (.) a Berlino@s:ita a Roma .

```

11 *STU: por le@s:ita grande ciudades [*3] de Europa .

[...]

240 @End

[sonia2.cha]

Es importante aclarar que una línea no necesariamente es igual a un enunciado. Éste podría durar varias líneas si está ausente el marcador de fin de enunciado: (.), (?), (!) y presente uno de continuación: (+...). En el caso que sigue el enunciado empieza en la línea principal 252 y finaliza en la 254 (las líneas 253 y 255 son comentarios de error secundarios).

252 *STU: bastería [*] que todos <paga> [/] pagan los impuestos [*0] +...

253 %err: bastaría que todos pagaran

254 *STU: y partes de los problemas [*0] podría ser reducidos [*] .

255 %err: y se podría reducir parte de los problemas .

[jako5.cha]

Se siguió a Pallotti (2007) para eliminar los siguientes casos de concordancia del corpus. A saber: (1) un ‘eco’ o repetición exacta dentro de los 7 enunciados (secuencia de caracteres terminados en: (.), (?), (!)) a partir del caso en consideración; (2) repeticiones inmediatas en el mismo enunciado del *token*; (3) palabras cuyas desinencias no pueden identificarse claramente debido a problemas de sonido; (4) casos de concordancia con palabras pertenecientes a la L1 (por ej.: *STU: y conozco muy bien le@s:it calles.); (5) casos de formas congeladas (‘chunks’): *primeros auxilios*; (6) casos en los que el plural se expresa en ítems léxicos, como números o pronombres (*STU: tu puedes beber tres cuatros vasetto@s:it vasos.), con excepción de los números de 200 a 900 que en español tienen concordancia; (7) sobre-extensiones: *un autos, muchos dineros*. Por otra parte, no se considera como error a los neologismos, ya que la raíz léxica no es conocida por el aprendiente:

190 *STU: la crema por los rugos [*0]. → neologismo: para las arrugas. [it.: *le rughe*]

[sonia6.cha]

Si una reformulación es correcta, no se computa como error. En el siguiente ejemplo «alemanos» se reformula correctamente:

192 *STU: ehh@fp yo creo que los [*0] estudiantes <alemanos>

193 [/] <alemanes> [/] alemanes [*0] .

194 *STU: leen menos .

[sonia4.cha]

Se anotaron a continuación marcadores [‘tags’] en el corpus para realizar el conteo posterior. Son los siguientes:

- [*0] = ausencia de error;
- [*1] = errores en el género;
- [*2] = errores debidos al uso de la terminación «(-e-)s»: (i) por no tomar en cuenta la última consonante de la raíz léxica, que exige un plural con “e” epentética en *-(e)-s*; (ii) por uso en contexto incorrecto u omisión en correcto;
- [*3] = errores de plural, o sea ausencia de *-s*;
- [*4] = errores mixtos por acumulación de los anteriores.

Por ejemplo:

56 *STU: leer o hablar con muchos personas [*1]. → error de género: muchas personas.

[sonia1.cha]

148 *STU: después les@s:ita ehh@fp después las doce hay muchos trenos [*2] → error de -e-epentética: muchos trenes.

[sonia7.cha]

293 *STU: si pero los veneciano [*3] conocen donde ir por comprar mejor. → error de plural: los venecianos.

[sonia5.cha]

144 *STU: por les joven [*4]. → error por acumulación: para los jóvenes.

[sonia2.cha]

El conteo se hizo con el programa CLAN (Mac Whinney, 2010). El siguiente constituye un ejemplo de búsqueda en el *Corpus* ya codificado:

```
kwat +z94u +t*STU +s"+[*]" sonia7.cha
```

En el ejemplo: KWAL encuentra los enunciados que satisfacen la búsqueda. En este caso en particular, se busca en la transcripción (archivo) «sonia7.cha» cualquier tipo de «tag» de error sin importar su contenido; en las líneas del sujeto (STU) y solamente hasta la línea 94.

Por último, es necesario aclarar que se codificaron siempre concordancias de dos términos, entre el controlador y el objetivo. Es decir que en el caso que sigue se codifica por un lado: *los monumentos*, y por otro: *monumentos históricos*; no obstante haya dos «objetivos» que concuerdan con el «controlador».

*STU: ehh@fp porque me gustan los monumentos [*0] históricos [*0] .

[mirka5.cha]

4.3. Variables como respuesta y atributos.

Esta sección explica la creación de las variables respuesta y los atributos, o sea características de cada instancia de concordancia producida por los alumnos. Se utilizó el lenguaje *R* (diferentes librerías) .

4.3.1. Variables «respuesta».

Se definieron las siguientes variables respuesta:

1. Binaria (RES_BIN). Según los niveles: 1 = error; 0 = correcto.
2. Categorica (RES_CAT). Según los niveles definidos por los tipos de errores: 0 = sin error; 1 = error de género; 2 = error de «(-e)s»; 3 = error de plural; 4 = errores acumulados.
3. Tiempo hasta el evento «error» (TEV): Es el tiempo, medido en cantidad de instancias correctas de concordancia sucesivas observadas, hasta que se produce un error.
4. Serie simbólica (S): Conformada por símbolos para cada tipo de *tag*, según el alfabeto $\mathcal{A} = \{0, 1, 2, 3, 4\}$, $|\mathcal{A}| = 5$. Cada concordancia es un conjunto de dos símbolos.

Por ejemplo, el siguiente error, según las variables definidas se clasifica como: (1) Binaria = 1; (2) Categórica = 3 (error de plural); (3) TEV = 3 (tres instancias correctas pasaron hasta que se produce el error); (4) S = {0,3} (sin error [0] + error de plural [3]).

Transcripción	Instancia producida	español	Binaria	Categorica	TEV	S
nati1.cha; línea 29	<i>los nombre</i>	los nombres	1	3	3	03

4.3.2. Variables «atributos».

Se crearon los siguientes atributos para describir el contexto en el que sucede cada instancia de concordancia. En primer lugar se codificó la sesión y la línea de transcripción en el corpus de cada instancia; y el nivel de competencia del alumno. Luego, se registró en un atributo binario la presencia (= 1) o ausencia (= 0) de un modificador que *no* sea el artículo definido [ADJ]. Dicho atributo se especificó todavía más según el tipo de modificador del controlador [MOD], según: $0 = l = \textit{artículo definido}$; $1 = l_i = \textit{artículo indefinido}$; $2 = d = \textit{determinante}$ (adjetivos posesivos, indefinidos, demostrativos, interrogativos, exclamativos); $3 = j = \textit{adjetivos}$ (calificativos, numerales, ordinales). Para este atributo se hizo una anotación manual de *POS tagging* y se la controló mediante anotación automática (y luego corregida por el autor de esta tesis) hecha por el *Stanford CoreNLP POS Tagger* (Manning et Al., 2014). Por otro lado, ya que cada concordancia se codifica con dos términos pero puede haber más términos «objetivo», se codificó si se trataba o no de una concordancia de más de dos términos [GRAM]: $0 = \textit{dos términos}$; $1 = \textit{más de dos términos}$.

Además, se especificó si en el el controlador, en el objetivo, o en ambos, había una desinencia que requiera la inserción de «e» epentética [-(*e*)s]. El razonamiento fue que realizar concordancia con dos operaciones de este tipo resulta más complicado que con una o con ninguna; según [ES]: $0 = \textit{sin «e» epentética}$; $1 = \textit{con «e» epentética en un término}$; $2 = \textit{con «e» epentética en ambos términos}$. Este atributo fue motivado por la dificultad que se reporta en la literatura para el plural en *-es* frente al plural en *-s*.

A continuación se registró si se trataba o no de una concordancia a larga distancia [LDA, por sus siglas en inglés], según: $0 = \textit{no}$; $1 = \textit{sí}$. En el caso de haber concordancia a larga distancia, se codificó el contexto en el que sucedía [LDA_C], utilizando los siguientes marcadores:

- Marcadores de clase de palabra en el contexto del término «objetivo»: <V> = verbo; <REL> = pronombre relativo.
- Marcadores de lema correspondientes a dicha clase de palabras: <SER>, <QUE>, <QUERER>, <PARECER>, <TENER_QUE>.
- Marcadores de modificadores del término «objetivo»: <tan>, <menos>, <más>, <muy>, <bastante>, <un poco>.
- Marcadores de distancia lineal (en palabras) hasta el controlador según <nL> donde n = número de palabras; o bien <E-c> donde c = cantidad de enunciados hacia atrás hasta encontrar aquel donde está el controlador. Por ejemplo: <2L>, <E-1>.
- Marcadores sobre sintagmas nominales en plural que pueden contener un «distractor» para la concordancia: <DE+NP.PL>, <POR+NP.PL>.

En total hubo 78 concordancias a larga distancia (TOKENS) y 39 configuraciones de codificación (TYPES). Los ejemplos siguientes ilustran la codificación:

333 *STU: y es normal que ehh@fp asiñatura [*] así que <no te> [/] no te parecen &u útiles [*3].

334 %err: asignaturas así que no te parecen útiles .
[mirka9.cha]

Este ejemplo se codificó como: <REL><QUE><V><PARECER><5L>. O sea que el objetivo «útiles» concuerda a larga distancia con el controlador «asignaturas» que se encuentra cinco palabras [<5L>] (sin contar reformulaciones o falsos comienzos); es predicativo del verbo «parecer» [<V><PARECER>] y está embebido en una oración subordinada relativa con pronombre «que» [<REL><QUE>]. Si se considera:

128 *STU: <en Italian no era> [/] en Italia non era &como ehh@fp comune .

129 *STU: <la> [/] las mujeras [*2] de los políticos [*0] no eran famosas [*2].
[nati13.cha]

En este caso: <DE+NP.PL><V><SER><5L>; significa que «famosas» está en posición predicativa del verbo «ser» [<V><SER>], a cinco palabras del controlador «mujeres»; el cual es un sintagma nominal complejo que contiene un sintagma preposicional con «de» [DE+NP.PL] cuyo núcleo «políticos» podría interferir con el género en la concordancia.

Ahora bien, de dichas configuraciones solamente se extrajo el marcador de la clase de palabra (POS) y el de la distancia lineal al controlador (DIS). Se crearon pues dos atributos cualitativos, con las categorías: (i) <NA> = 0 = *sin larga distancia*, 1 = <V> = *sin subordinación*, 2 = <REL><V> = *con subordinación*; (ii) <nL>: <1L>, <2L>, <3L>, <4L>, <5L>, <6L>, <9L> y <E-c>; agrupadas como: <NA> = 0 = *sin larga distancia*; <1L>, <2L>, <3L> = 1 = *distancia corta*; <4L>, <5L>, <6L>, <7L>, <9L> = 2 = *distancia larga*; <E-c> = 3 = *distancia a otros enunciados*.

También se registró si el controlador era o no animado, según: 0 = *inanimado*, 1 = *animado*.

Asimismo, para reflejar el hecho de que la producción de las instancias de concordancia es temporal en la cadena del habla, se creó una variable TIEMPO, para controlar por posibles efectos de correlación temporal entre las observaciones. Dicha variable codifica con números enteros positivos sucesivos a las instancias dentro de *una misma sesión*; o sea, que el conteo temporal vuelve a comenzar cuando se cambia de sesión.

A modo de ilustración de los atributos descriptos hasta ahora, considérese el siguiente fragmento de transcripción de MIRKA:

46 *STU: entonces <lo que> [/] ehh@fp yo creo que ehh@fp los animales ehh@fp tienen derechos .
 47 *STU: pero <no> [/] no son **lo** [*3] **mismos derechos** [*0] ehh@fp que ehh@fp
 48 +...
 49 *STU: lo <que deben tener las> [/] que tienen **las personas** [*0] .
 50 %err: los mismos derechos
 51 *STU: no se si me explico .
 52 *STU: **los** [*0] **seres humanos** [*0] ehh@fp tenemos +...
 53 *STU: es algo un poco malo que decir .
 54 *STU: pero tenemos [*] más derechos que los animales .
 55 %err: tenemos
 56 *STU: en el sentido que [*] ehh@fp <me>[/] me doy cuenta que a veces **los animales** [*0] son muy **sensibles** [*0] .
 [mirka6.cha]

Se registraron las variables como sigue:

LINEA	SESIÓN	TIEMPO	INSTANCIA	LDA_C	POS	DIS	MOD	LDA	ES	GRAM	ANIM
47	6	5	lo derechos (1)	<NA>	0	0	l	0	0	1	0
47	6	6	mismos derechos (2)	<NA>	0	0	d	0	0	1	0
49	6	7	las personas	<NA>	0	0	l	0	0	0	1
52	6	8	los seres (1)	<NA>	0	0	l	0	1	1	1
52	6	9	seres humanos (2)	<NA>	0	0	j	0	1	1	1
56	6	10	los animales	<NA>	0	0	l	0	1	0	1
56	6	11	[animales] <muy> sensibles	<V><SER><2L><muy>	1	1	j	1	1	0	1

Cuadro 4.5: Ejemplo ilustrativo de registro de atributos. Leyenda: POS: clase de palabra (contexto no local); DIS: distancia en palabras (contexto no local); MOD (tipo de modificador («objetivo») del controlador); LDA: concordancia a larga distancia; ES: «e» epentética; GRAM: concordancia de más de dos términos; ANIM: animicidad del controlador; LDA_C: contexto no local.

La primera instancia de concordancia «los mismos derechos» se registra en dos partes (esto se indica en la columna «instancia» con números entre paréntesis): (1) «los derechos»; (2) «mismos derechos». La primera posee un artículo definido como modificador [*MOD = l*], y la segunda, un determinante [*MOD = d*].

Ambos forman parte de una concordancia de más de dos términos [$GRAM = I$]. El controlador es inanimado [$ANIM = 0$]. La que sigue, «las personas», registra artículo definido [$MOD = I$] y controlador inanimado [$ANIM = I$]. A continuación hay otra concordancia de dos términos [$GRAM = I$]. En la primera concordancia («los seres») se encuentra un artículo definido [$MOD = I$]; y en la segunda, un adjetivo calificativo [$MOD = j$]. El controlador es animado [$ANIM = I$]. Además, «seres» posee una desinencia con «e» epentética: *ser-e-s* [$ES = I$]. Le sigue «los animales» con un artículo definido [$MOD = I$]; controlador animado [$ANIM = I$] y con desinencia con «e» epentética: *animal-e-s* [$ES = I$]. Por último se halla una concordancia a larga distancia «animales sensibles» [$LDA = I$], el «objeto» está en posición predicativa del verbo «ser», modificado por «muy» y a dos palabras de distancia del controlador [$<V><SER><2L><muy>$]. La clase de palabra (POS) extraída es [V] = I ; o sea que no hay un contexto de subordinación; y la distancia (DIS) es $<2L> = I$; o sea corta distancia. El modificador es un adjetivo calificativo [$MOD = j$]; y el controlador a larga distancia es animado [$ANIM = I$] y con desinencia con «-e-» epentética: *animal-e-s* [$ES = I$]. Por último, las observaciones se encuentran entre los tiempos 9 y 11.

Los siguientes atributos, sobre rasgos del controlador, se extrajeron de la base de datos «BuscaPalabras» (Davis & Perea, 2005):

- Concretud (CONC): índice subjetivo en escala de 1 a 7 que indica cuán concreta es una palabra de menos (+ abstracta) a más (+ concreta).
- Familiaridad (FAM): índice subjetivo en escala de 1 a 7, que indica cuán frecuentemente una palabra es oída, leída o producida diariamente.
- Imaginabilidad (IMA): índice subjetivo en escala de 1 a 7 que indica la intensidad con la que una palabra evoca imágenes.
- Frecuencia (LEXESP): frecuencia de la palabra en el corpus «BuscaPalabras», en escala por mil.

Por ejemplo, para los ejemplos del Cuadro anterior el registro fue:

LÍNEA	SESIÓN	TIEMPO	INSTANCIA	FAM	IMA	CONC	LEXESP
47	6	5	lo derechos (1)	6,17	3,71	3,62	130
47	6	6	mismos derechos (2)	6,17	3,71	3,62	130
49	6	7	las personas	7	6,22	5,49	171,79
52	6	8	los seres (1)	5,29	4,23	2,37	82,5
52	6	9	seres humanos (2)	5,29	4,23	2,37	82,5
56	6	10	los animales	6,63	6,31	3,54	73,04
56	6	11	[animales] <muy> sensibles	6,63	6,31	3,54	73,04

Cuadro 4.6: Ejemplo ilustrativo de registro de atributos. Leyenda: CONC = Concretud; FAM = Familiaridad; IMA = Imaginabilidad; LEXESP = Frecuencia.

Se registró también lo siguiente: (1) ESP: el par de términos en español; (2) ITA: el par de términos en italiano; (3) ESP1: la vocal final de la palabra + la desinencia de plural en español del primer término; (4) ESP2: la vocal final de la palabra + la desinencia de plural en español del segundo término; (5) ITA1: la vocal final de la palabra italiana, que coincide con el morfema de plural, del primer término; (6) ITA2: la vocal final de la palabra italiana, que coincide con el morfema de plural, del segundo término; (6) O1: la vocal final de la palabra + la desinencia de plural en español del primer término, *efectivamente observado*; (7) O2: la vocal final de la palabra + la desinencia de plural en español del segundo término, *efectivamente observado*. La tabla que sigue muestra las combinaciones posibles en el Corpus para ITA1/ITA2 hacia ESP1/ESP2:

	(V)AS	(V)ES	IS	(V)OS	US
A	-	cinema [MP] / cines	-	ouva [MP] / huevos	-
E	case [FP] / casas	strade [FP] / calles	-	scarpe [FP] / zapatos	-
EI	-	-	(i) miei [MP] / mis	miei [MP] / míos	-
I	turisti [MP] / turistas	tedeschi [MP] / alemanes	-	libri [MP] / libros	-
IE	-	-	(le) mie [FP] / mis	-	-
O	duecento [I] / doscientas	-	-	quattrocento [I] / cuatrocientas (le) radio [FP] / radios	-
OI	-	-	-	-	(i) suoi [MP] / sus
UE	-	-	-	-	(le) sue [FP] / sus
V	film [I] / películas	cittá [I] / ciudades	-	jean [I] / vaqueros	-

Cuadro 4.7: Posibles combinaciones entre español e italiano en el Corpus. Leyenda: M = masculino; F = femenino; P = plural; I=invariable.

Se requieren algunas observaciones respecto de la morfología del italiano:

1. En italiano algunas palabras son invariantes: (i) nombres terminados en vocal acentuada: (*la / le*) *cittá* / (*la / las*) *ciudad(-es)*; (ii) monosílabos: *il re / i re* (*el rey / los reyes*); (iii) nombres de origen extranjero, terminados generalmente en consonante: *il computer / i computer* (*el / los ordenador(es)*); *la / las computadora(s)*). Estos se registraron con «v» en ITA1/ITA2 y con «vas / ves / vos» en ESP1/ESP2. También son invariantes los terminados en -o (todos femeninos): (*la / le*) *radio* / **las radios**; (*la / le*) *foto* / *las fotos*; estos se registraron como «o» [ITA1/ITA2] > «as» [ESP1/ESP2].
2. Los números de 200 a 900 carecen de género, como en español: *duecento case - monumenti / doscientas(-os) casas - monumentos*; en este caso se registró «o» en ITA1/ITA2 y «os / as» en ESP1/ESP2.
3. Los posesivos de 1ra. a 3ra. persona (singular / plural) tienen diferentes morfemas, marcando género y número, donde en español solamente se marca número: (1) ***i miei libri - le mie case / mis libros - casas***; (2) ***i tuoi libri - le tue case / tus libros - casas***; (3) ***i suoi libri - le sue case / sus libros - casas***. Estos casos se marcaron como: «ei» / «ie» > «is» (*miei / mie > mis*); «oi» / «ue» > «us» (*tuoi / suoi / tue / sue > tus / sus*).
4. Algunas palabras tienen plural en «a»: (*il dito - (le) dita / el dedo - los dedos*); estos casos se registraron como: «a» [ITA1/ITA2] > «os» [ESP1/ESP2]. Unos pocos nombres masculinos son invariantes en -a también: *il cinema / i cinema* (*el / los cine(s)*); *il delta / i delta* (*el / los delta(s)*); el único caso fue «cinema», que se registró como: «a» [ITA1/ITA2] > «es» [ESP1/ESP2].
5. El morfema de masculino plural de más frecuencia es «-i»: (*il libro - (i) libri / el libro - los libros*); sin embargo, no siempre coincide con el masculino plural «os» en español: *i turisti (it.) / los turistas (esp.)*; *i tedeschi (it.) / los alemanes (esp.)*.
6. El morfema de femenino plural de más frecuencia es «-e»: (*la casa - (le) case / la casa - las casas*); sin embargo, no siempre coincide con el femenino plural «as» en español: *le strade / las calles*; ***le scarpe [Fem. Pl.] / los zapatos [Masc. Pl.]***.

El Cuadro siguiente ilustra con ejemplos del corpus:

SUJETO	LINEA	SESIÓN	INSTANCIA	ESP	ITA	ITA1	ESP1	ITA2	ESP2	O1	O2
SONIA	160	3	las comoditas	las comodidades	le comodità	e	as	v	ves	as	vas
SONIA	19	3	mis estudios	mis estudios	i miei studi	ei	is	i	os	is	os
SONIA	15	4	las playas	las playas	le spiagge	e	as	e	as	as	as
MIRKA	255	2	radio especializadas	radios especializadas	radio specializzate	o	os	e	as	o	as
NARTI	322	2	tu pies	tus pies	i tuoi piedi	oi	us	i	es	u	es
JAKO	11	1	lo olores	los olores	gli odori	i	os	i	es	o	es

Cuadro 4.8: Ejemplos de: ESP, ITA, ESP1, ESP2, ITA1, ITA2, O1, O2.

Además, se crearon dos variables basadas en la distancia de *Levenstein* (Nerbonne et. A., 2013; Oakes, 1998), con el objetivo de medir la similitud entre las raíces léxicas entre el español y el italiano; y entre los morfemas de género y número plural. El algoritmo de *Levenstein* calcula la distancia entre dos secuencias de caracteres como el número mínimo de operaciones necesarias para transformar una secuencia en la otra. Estas operaciones son: DELETE (borrar); SUBSTITUTE (sustituir); INSERT (insertar). Si se denota con m al número de caracteres de la primera secuencia; y con n al número de caracteres de la segunda; se puede crear una matriz $(m + 1) \times (n + 1)$ donde las filas (i) y las columnas (j) están numeradas de 0 a n . La entrada en la celda (i, j) de la matriz denota la distancia entre el primer carácter i en la palabra (secuencia) n y el primer carácter j de la palabra (secuencia) m . Se comienza calculando la distancia en la celda $(0, 0)$ [en el extremo superior izquierdo de la matriz], o sea la distancia entre «NULL» / «NULL» y luego se mueve progresivamente, calculando todas las distancias intermedias, hasta alcanzar la celda (m, n) en el extremo inferior derecho de dicha matriz. Los valores en las otras celdas (i, j) siempre tienen en cuenta los valores de las tres celdas precedentes: $(i - 1, j)$, $(i - 1, j - 1)$, $(i, j - 1)$; sin embargo, si $i = 0$ o $j = 0$ los valores de las celdas precedentes no se usan. Para cada celda se realizan los siguientes cálculos:

1. El valor en la celda $(i - 1, j)$ más el costo de la operación DELETE sobre el carácter i en consideración en el segmento a (o sea: $a[i]$).
2. El valor en la celda $(i - 1, j - 1)$ más el costo de la operación SUBSTITUTE sobre el carácter i en consideración en el segmento a ; sustituyendo con el carácter j en consideración en el segmento b . O sea: sustituir $a[i]$ por $b[j]$. Pero si $a[i] = b[j]$ dicho costo es cero.
3. El valor en la celda $(i, j - 1)$ más el costo de la operación INSERT sobre el carácter i en consideración en el segmento a .

El valor resultante de la celda (i, j) es el mínimo de estos tres. Se asignaron los siguientes pesos para las operaciones : (i) DELETE = 0.3; (ii) SUBSTITUTE = 0.6; INSERT = 1 . Supóngase el siguiente ejemplo:

7 *STU: yo tengo muchos amigos [*0] .

[sonia1.cha]

La similitud entre las terminaciones de ambos términos de la concordancia [usando las variables: ITA1 > ESP1; ITA2 > ESP2] se calculó como: $(i > os) + (i > os) = 1,6 + 1,6 = 3,2$. Ya que, según la matriz descripta para la distancia $(i > os)$:

-	(n)	0	1	2
(m)	-	«NULL»	«O»	«S»
0	«NULL»	0	1	2
1	«I»	0.3	0.6	1.6

Por ejemplo, para las últimas dos celdas $(1, 1)$ y $(1, 2)$ los cálculos resultan:

	(1, 1)	(1, 2)
DELETE	$[(0, 1) > (1, 1)] = 1 + 0,3 = 1,3$	$[(0, 2) > (1, 2)] = 2 + 0,3 = 2,3$
SUBSTITUTE	$[(0, 0) > (1, 1)] = 0 + 0,6 = \mathbf{0,6}$	$[(0, 1) > (1, 2)] = 1 + 0,6 = \mathbf{1,6}$
INSERT	$[(1, 0) > (1, 1)] = 0,3 + 1 = 1,3$	$[(1, 1) > (1, 2)] = 0,6 + 1 = \mathbf{1,6}$

También se practicó una corrección sumando 0.2 al valor arrojado por el algoritmo en los casos en los que interviene una *-e-* epentética. Esto es así porque el algoritmo no distingue entre una terminación «es» donde la «e» es epentética y otra en la que no sucede ésto. A modo de ejemplo, en el caso siguiente («las lecciones») donde el segundo término tiene *-e-* epentética, se calculó:

369 *STU: creo que <no es bueno> [//] no está bueno que todo [*3] los días [*0] +...

370 están ocupado [*3] da@s:ita *las lecciones* [*0] .

[sonia8.cha]

$$(e > as) + (i > es) = 1,6 + (1,6 + \mathbf{0,2}) = 3,4.$$

Además de la similaridad entre las desinencias, se calculó la similaridad entre las raíces léxicas de ambos términos. Para ello, se extrajo la raíz de cada uno de los términos (a partir de las variables «ESP» e «ITA», detalladas más arriba) y se aplicó la distancia de *Levenstein* como sigue:

56 *STU: leer o hablar con **muchos personas** [*1] . (it.: **molt-e person-e**)

[sonia1.cha]

$$(molt > much) + (person > person) = 1,8 + 0 = 1,8$$

Otro atributo que se creó fueron los errores acumulados [ACUM]: cantidad de errores hechos hasta la instancia anterior a la actual dentro de una misma sesión (o sea que el conteo es desde cero cuando cambia la sesión). Por ejemplo, para las primeras seis concordancias observadas de la primera sesión de NATI, el error se produce en la tercera concordancia; por lo tanto, en la cuarta concordancia los errores acumulados hasta la instancia anterior son iguales a *uno*, y continuará así hasta que se produzca un nuevo error y el conteo crezca.

LINEA	SESIÓN	TIEMPO	INSTANCIA	ESP	ITA	ACUM
27	1	1	las iglisias	las iglesias	le chiese	0
27	1	2	los monumentos	los monumentos	i monumenti	0
29	1	3	los nombre	los nombres	i nomi	0
29	1	4	todos santos (1)	todos santos	tutti santi	1
29	1	5	los santos (2)	los santos	i santi	1
46	1	6	las personas	las personas	le persone	1

Cuadro 4.9: Errores acumulados

Lo siguiente que se hizo fue codificar cada instancia de concordancia de la variable ESP (o sea, la instancia en español, no aquella observada) utilizando información de las variables ESP1, ESP2, LDA_C y MOD. Primero se indicó un marcador de concordancia a larga distancia si la hubiere [«L»]; luego se indicó la clase de palabra de cada término de la concordancia según MOD, en el orden en que aparecían en la instancia. Después se indicó la terminación de cada término según ESP1 y ESP2. Si la instancia pertenecía a una concordancia con más de dos términos entonces se codificó la instancia con todos los términos. En el caso de que se tratara de larga distancia, se especificó el lema del verbo y/o el pronombre relativo; también alguna estructura que implicara interferencia para computar la concordancia. En total se crearon 104 TYPES. Por ejemplo, para los casos siguientes:

SUJETO	LINEA	SESIÓN	INSTANCIA	ESP	TYPE
NATI	142	2	mi compañeros	mis compañeros	d-n-is-os
SONIA	197	2	[romanos] <muy> alegre	romanos alegres	L-n-<SER>-j-os-es
JAKO	118	1	muchísimas habitaciones (1)	muchísimas habitaciones	d-n-j-j-j-as-*es-*es-as-<y>-as
			habitaciones culturales (2)	habitaciones culturales	
			habitaciones arquitectónicas (3)	habitaciones arquitectónicas	
			<y> habitaciones artísticas (4)	habitaciones artísticas	
MIRKA	114	1	[nuestros sitios] ocupados	sitios ocupados	L-n-<ESTAR>-j-os-os

Cuadro 4.10: Ilustración de TYPES.

El primer caso se trata de una concordancia de dos términos; su TYPE es: «determinante-nombre-is-os». El segundo caso es una concordancia a larga distancia marcada por «L» [«los romanos son muy alegres»]. Consta de un nombre («n») luego se especifica el verbo «SER», seguido de un determinante «j», después vienen las terminaciones de ambos términos: «os», «es» [sin -e- epentética]. En el caso de las concordancias a larga distancia no se especificó el número de palabras intervinientes entre controlador y objetivo. El tercer ejemplo, producido por JAKO, es «muchísimas habitaciones culturales arquitectónicas y artísticas»¹. Se trata de una concordancia con un controlador y cuatro objetivos, codificada en cuatro instancias observadas. Sin embargo, en lo que respecta a la variable en objeto, todas las instancias poseen el mismo TYPE: «d-n-j-j-j-as-*es-*es-as-<y>-as». O sea: un determinante «d», seguida de un nombre «n», seguida de tres adjetivos «j»; cuyas terminaciones son: «as», «*es», «*es» [(*) marca «-e- epentética»], «as», «as»; también se marca que antes del último adjetivo aparece un coordinante «y». En lo que atañe al último ejemplo, producido por MIRKA, se trata de otra concordancia a larga distancia [«nuestros sitios estaban ocupados»], marcada por «L», luego se especifica un nombre «n», el lema del verbo «estar», seguido de un adjetivo «j»; ambos términos terminan en «os».

Ahora bien, el objetivo de codificar ésto fue el de buscar las frecuencias de cada TYPE en el Corpus de datos propio y en un Corpus del español electrónico *online*. Para esto último, se apeló al Corpus del español *EsTenTen* de *Sketch Engine* (Kilgarriff et. al., 2014). Consiste en un corpus de más de 10 billones de palabras con textos extraídos de internet; anotado para clases de palabras (POS) usando el *FreeLing Morphological Tagger v. 3.1*. O sea que hay disponible información sobre género y número. *Sketch Engine* permite búsquedas con expresiones regulares y utilizando el *Corpus Query Language (CQL)* [Jakubíček et. Al., 2010]. El siguiente ejemplo ilustra el uso de *CQL* para realizar la búsqueda del TYPE «d-n-as-as» [ej.: «muchas personas»]:

```
[tag="D..FP.*" & (word=".*as" & word!="(L|l)as" & word!="(U|u)nas")]
[word=".*as" & tag="N.FP.*"]
[(word!=".*as" & word!=".*es" & word!=".*os")&(tag!="A..FP.*" & tag!="V.P.*")]
```

Cada corchete especifica una palabra con determinados rasgos. El primer corchete pide una palabra que sea un determinante, femenino plural; que termine en «as» pero que no sean las formas «la/las/unos/unas». El segundo corchete pide una palabra que termine es «as» y que sea un nombre, femenino plural. El último corchete es general y se aplicó siempre: evita que al TYPE le siga otra palabra que concuerde también con el controlador. Específicamente, aquí prohíbe que le sigan palabras terminadas en «as» / «es» / «os» y que sean adjetivos (femeninos plurales) o verbos en participio.

Se registró la frecuencia absoluta y la frecuencia absoluta por millón de palabras del TYPE. En el caso del Corpus de la tesis, se especificó la frecuencia absoluta. El Cuadro 11 lo ilustra.

Por último, se crearon siete atributos binarios de «estrategia» para la formación del plural: cada atributo registraba «1» en aquella instancia donde la estrategia de plural podía ser aplicada en *alguno* de los dos términos de concordancia (o en ambos). Se definieron como sigue:

1. Estrategia 1 (EST1): si la palabra plural del italiano termina en *-i* poner en español plural en *-os*.

¹El sujeto usa un cognado, ya que «abitazione» en italiano significa «vivienda».

EJEMPLO	TYPE	Fabs_ES	Fmill_ES	Fabs_C
muchas personas	d-n-as-as	5842065	531,8	105
grandes ciudades	j-n-es-*es	45466	41,39	10
Muchos libros interesantes	d-n-j-os-os-es	61118	5,56	10
los idiomas	l-n-os-as	1670272	152,04	29

Cuadro 4.11: Frecuencias en el Corpus propio y en *EsTenTen*. Leyenda: Fabs_ES = Frecuencia absoluta en *EsTenTen*; Fmill_ES = Frecuencia por millón en *EsTenTen*; Fabs_C = frecuencia absoluta en el Corpus propio; TYPE = tipo de concordancia.

2. Estrategia 2 (EST2): si la palabra plural del italiano termina en *-e* poner en español plural en *-as*.
3. Estrategia 3 (EST3): si la palabra plural del italiano termina en *-o* o en *-a* no acentuada (*le foto / las fotos, le osa / los huesos*), poner el plural del italiano.
4. Estrategia 4 (EST4): si la palabra plural del italiano termina en *-e*, poner en español el plural en *-es*. Por ejemplo: *vacanze [it.] > vacaciones [esp.]*; *strade [it.] > calles [esp.]*; *volte [it.] > veces [esp.]*.
5. Estrategia 5 (EST5): si la palabra *singular* del italiano termina en *-e*, poner en español el plural en *-es*. Por ejemplo la palabra *sole* (esp. *sol*) podría ser la base para formar el plural español agregando «s»: *sole [it.] > soles [esp.]*; y el singular también, sacando «s»: *sole [it.] > sol [esp.]*; *istituzione [it.] > instituciones [esp.]*. Es decir, casos en los cuales el español coincide con la aplicación del plural con *-e-* epentética.
6. Estrategia 6 (EST6): si la palabra *singular* del italiano termina en *-e*, poner en español el plural en *-es*. Por ejemplo, la palabra *grande* (esp. *grande*) o *studente* (esp. *estudiante*) podrían formar plural (y singular) a partir de una base singular en italiano: *grandes, estudiantes*. Otros casos: *fonte [it.] > fuentes [esp.]*; *abitudine [it.] > costumbres [esp.]*; *dolce [it.] > dulces [esp.]*. Son casos que no coinciden con *-e-* epentética.
7. Estrategia 7 (EST7): si la palabra plural del italiano termina en *-a* acentuada (*università / universidades*) o es invariante terminada en consonante (*i film / las películas*) poner , en general, plural en *-es* o bien buscar la base léxica en la memoria y aplicar otra estrategia.

El Cuadro que sigue ejemplifica los casos en italiano, español y la instancia efectivamente producida por el alumno.

italiano PLURAL	italiano SINGULAR	español	INSTANCIA	EST1	EST2	EST3	EST4	EST5	EST6	EST7
molte volte	(molta) volta	muchas veces	muchas vesas	0	1	0	1	0	0	0
vacanze gradevoli	vacanza gradevole	vacaciones agradables	vacacione agreeables	0	0	0	1	0	0	0
uniche moto	unica moto	únicas motos	únicas moto	0	1	1	0	0	0	0
molti cinema	(molto) cinema	muchos cines	muchos cines	1	0	1	0	0	0	0
responsabilità sociali	responsabilità sociale	responsabilidades sociales	responsabilidades sociales	0	0	0	0	1	0	1
le abitudini	l'abitudine	las costumbres	los costumbre	0	0	0	0	0	1	0

Cuadro 4.12: Ejemplos de Estrategias.

A modo de resumen, el Cuadro 13 muestra las variables descritas hasta ahora, sus clases y sus niveles.

Variable	Descripción	Clase (R / A)	Niveles
RES_BIN	Errores: Acierto / error	Cualitativa (R)	1 = error; 0 = acierto
RES_CAT	Errores: Tipos	Cualitativa (R)	0 = sin error; 1 = error de género; 2 = error de «(-e)»; 3 = error de plural; 4 = errores acumulados
TEV	Tiempo hasta el evento «error»	Cuantitativa (R)	de 0 a ... (infinito)
S	Símbolos con el alfabeto {0, 1, 2, 3, 4}	Cualitativa (R)	{a, b} donde $a, b \in \{0, 1, 2, 3, 4\}$
NIVEL	Nivel de Competencia Lingüística	Cualitativa (A)	1 a 4
SESIÓN	Sesión transcripta	Cualitativa (A)	1 a 12 (14)
LINEA	Línea en la transcripción .CHA	Cuantitativa (A)	[6,515]
INSTANCIA	Instancia de Concordancia observada	Secuencia de caracteres	-
MOD	Tipo de modificador del controlador	Cualitativa (A)	0 = artículo definido; 1 = artículo indefinido; 2 = determinante; 3 = adjetivo
LDA	Concordancia a larga distancia	Cualitativa (A)	0 = sin larga distancia; 1 = con larga distancia
ES	Presencia de -e- epentética	Cualitativa (A)	0 = sin «e» epentética; 1 = con «e» epentética en un término; 2 = con «e» epentética en ambos términos
GRAM	Concordancia de 2 términos o más	Cualitativa (A)	0 = dos términos; 1 = más de dos términos
POS	Clase de palabra en LDA	Cualitativa (A)	<NA> = 0 = sin larga distancia (LDA), 1 = <V> = sin subordinación, 2 = <REL.><V> = con subordinación
DIS	Distancia lineal en LDA	Cualitativa (A)	0 = sin LDA; 1 = distancia corta; 2 = distancia larga; 3 = distancia a otros enunciados
LDA_C	Contexto en LDA	Secuencia de Caracteres	-
CONC	Concretud del controlador	Cuantitativa (A)	Escala de 1 a 7
FAM	Familiaridad del controlador	Cuantitativa (A)	Escala de 1 a 7
IMA	Imaginabilidad del controlador	Cuantitativa (A)	Escala de 1 a 7
LEXESP	Frecuencia del controlador	Cuantitativa (A)	[2.32,744.6]
ANIM	Animicidad del controlador	Cuantitativa (A)	0 = inanimado; 1 = animado
ESP	Instancia en español	Secuencia de caracteres	-
ITA	Instancia en italiano	Secuencia de caracteres	-
ESP1	Terminación del primer término en Esp.	Secuencia de caracteres	«E», «A», «EI», «I», «IE», «O», «OI», «UE», «V», «(V)AS», «(V)OS», «(V)ES», «IS», «US»
ESP2	Terminación del segundo término en Esp.	Secuencia de caracteres	Idem
ITA1	Terminación del primer término en Ita.	Secuencia de caracteres	Idem
ITA2	Terminación del segundo término en Ita.	Secuencia de caracteres	Idem
O1	Terminación observada	Secuencia de caracteres	-
O2	Terminación observada	Secuencia de caracteres	-
MORF	Similaridad entre terminaciones	Cuantitativa (A)	[2.2,3.6]
STEM	Similaridad entre raíces	Cuantitativa (A)	[1.8;10.2]
ACUM	Errores acumulados	Cuantitativa	[0,28]
TYPE	«Tipo de concordancia»	Secuencia de caracteres	-
Fabs_S	TYPE, Frecuencia absoluta: <i>EsTenTen</i>	Cuantitativa (A)	0 a ... (infinito)
Fmill_S	TYPE, Frecuencia por millón: <i>EsTenTen</i>	Cuantitativa (A)	0 a ... (infinito)
Fabs_C	TYPE, Frecuencia absoluta: Corpus propio	Cuantitativa (A)	[1,208]
EST1	Estrategia 1	Cualitativa (A)	0 = no aplica; 1 = aplica
EST2	Estrategia 2	Cualitativa (A)	0 = no aplica; 1 = aplica
EST3	Estrategia 3	Cualitativa (A)	0 = no aplica; 1 = aplica
EST4	Estrategia 4	Cualitativa (A)	0 = no aplica; 1 = aplica
EST5	Estrategia 5	Cualitativa (A)	0 = no aplica; 1 = aplica
EST6	Estrategia 6	Cualitativa (A)	0 = no aplica; 1 = aplica
EST7	Estrategia 7	Cualitativa (A)	0 = no aplica; 1 = aplica

Cuadro 4.13: Resumen. Variables creadas: R = respuesta; A = atributo.

4.4. Pre-procesamiento.

Se describen y aplican métodos para tratar la transformación, discretización, colinealidad entre atributos y los datos faltantes.

4.4.1. Datos faltantes.

Se recolectaron 1857 casos de concordancia en total. Sin embargo, los atributos relacionados con el controlador (excepto ANIM) a veces no tenían datos registrados en la base de datos de «BuscaPalabras». Debido a ello, hubo 161 casos en los que faltaban datos en una o más de estas variables. Los casos faltantes representaron el 8.6% de la base de datos. La proporción de datos faltantes por variable fue: (i) FAM: 7.32%; (ii) IMA: 8.61%; CONC: 7.32%; LEXESP: 2.26%. El Cuadro que sigue muestra los casos por sujeto. Se indica el porcentaje de casos del total de casos; y el porcentaje de faltantes del total de casos de cada sujeto.

Sujeto	Casos totales (%)	Faltantes (%)
SONIA	312 (0.16)	16 (0.05)
NATI	390 (0.21)	28 (0.07)
JAKO	464 (0.24)	40 (0.08)
MIRKA	691 (0.37)	77 (0.11)
TOTAL	1857 (100)	161 (100)

Cuadro 4.14: Datos totales y faltantes por sujeto.

El siguiente Cuadro muestra los patrones de datos (filas), resaltando las configuraciones de datos faltantes. Así, había 1696 casos con todas las variables completas; 24 casos donde solo faltaba el dato para CONC; 1 caso donde solo faltaba LEXESP; 95 casos donde FAM, CONC e IMA no tenían datos y 41 casos sin datos en las variables LEXESP, FAM, CONC, IMA.

#	OTRAS VARIABLES	LEXESP	FAM	CONC	IMA
1696	1	1	1	1	1
24	1	1	1	0	1
1	1	0	1	1	1
95	1	1	0	0	0
41	1	0	0	0	0

Cuadro 4.15: Patrones de datos (faltantes)

Se han teorizado tres mecanismos que explican los datos faltantes (Enders, 2010):

1. **MAR** («Missing At Random»): Cuando la probabilidad de que ocurra el dato faltante en la variable *Y* depende de los datos observados de las otra(s) variable(s) pero no de los datos de la variable *Y* misma.
2. **MCAR** («Missing Completely At Random»): Cuando la probabilidad de que ocurra el dato faltante en la variable *Y* no depende ni de los datos las otras variables ni de los datos de *Y*. Los datos faltantes son puramente azarosos.
3. **MNAR** («Missing Not At Random»): Cuando la probabilidad de que ocurra el dato faltante en la variable *Y* depende de los datos observados de las otra(s) variable(s) y además de los datos de la variable *Y* misma.

Generalmente, los métodos de imputación más recientes asumen que el mecanismo generador es MAR; que es menos restringido que MCAR, el cual resulta bastante implausible en la práctica. No existen tests estadísticos para distinguir entre MAR y MNAR, por lo que es necesario entender los propios datos. En este caso, faltan datos que simplemente no se recogieron en la base de datos de «BuscaPalabras». Es decir que parece razonable asumir que los datos faltantes de una variable, supóngase «Imaginabilidad» [IMA], no dependen de los datos de la variable misma. O sea que, por ejemplo, no se da el caso de que los datos que faltan sean los de menor (mayor) imaginabilidad en la escala subjetiva. Por lo tanto, es plausible suponer que MAR es el mecanismo generador.

Se utilizó el paquete *mice* [*Multivariate Imputation by Chained Equations*] de R (Van Buuren & Groothuis-Oudshoorn, 2011), que realiza imputación múltiple. Una imputación múltiple consiste en tres fases: la fase de imputación, la de análisis, y la de agregación [*pooling*] de estimaciones. En la primera se crean m copias de la base de datos, conteniendo cada una de ellas diferentes estimaciones de los datos faltantes. En la fase de análisis se aplica a cada base de datos la técnica de regresión elegida, obteniéndose m estimaciones de los parámetros (y sus errores típicos). Por último la fase de agregación combina las m estimaciones en un único resultado agregado. Si se utiliza la siguiente notación:

- m = número de imputaciones múltiples;
- $Y = (Y^{obs}, Y^{miss})$ representa la matriz $n \times p$ con n casos y p variables; y donde Y^{obs} son los datos observados; Y^{miss} son aquellos que faltan;
- R es la matriz indicadora $n \times p$ donde $r_{ij} = 1$ donde en y_{ij} falta un dato y $r_{ij} = 0$ donde y_{ij} tiene un dato observado.

El algoritmo *mice* funciona del siguiente modo para la fase de imputación (Van Buuren, 2013, p.110):

1. Elegir un modelo de imputación adecuado a la distribución de cada variable con datos faltantes: $P(Y_j^{miss} | Y_j^{obs}, Y_{-j}, R, \phi_j)$ $j = 1, \dots, p$. Por ejemplo, dada una variable binaria con distribución Bernoulli, se elegirá un modelo logístico. El modelo estimará la probabilidad condicional de un valor no observado la variable j –ésima dados los valores observados de la misma variable, de las otras (posiblemente también con valores faltantes), de la matriz indicadora y de los parámetros ϕ .
2. Para cada variable j –ésima rellenar con valores imputados iniciales \dot{Y}_j^0 muestreados al azar de los valores observados de dicha variable: Y_j^{obs} .
3. Bucle: repetir para $t = 1, \dots, T$ iteraciones.
4. Bucle: repetir para $j = 1, \dots, p$ variables con datos faltantes.
5. Definir: $\dot{Y}_{-j}^t = (\dot{Y}_1^t, \dots, \dot{Y}_{j-1}^t, \dot{Y}_{j+1}^{t-1}, \dots, \dot{Y}_p^{t-1})$ como los datos completos de la iteración corriente; excepto Y_j .
6. Muestrear al azar valores para los parámetros $\dot{\phi}_j^t \sim P(\dot{\phi}_j^t | Y_j^{obs}, \dot{Y}_{-j}^t, R)$ a partir de una distribución posterior bayesiana.
7. Usar estos valores para muestrear otros valores de imputación a partir de la posterior: $\dot{Y}_j^t \sim P(Y_j^{miss} | Y_j^{obs}, \dot{Y}_{-j}^t, R, \dot{\phi}_j^t)$.
8. END repetir j .
9. END repetir t .

El algoritmo comienza con valores imputados iniciales sacados de los observados de cada variable con datos faltantes y las rellena una por una. Una iteración t es un ciclo que comprende a todas las variables con datos faltantes Y_j^{miss} . Cada variable a imputar puede seguir un modelo de imputación diferente. En este caso, siendo todas cuantitativas, se elige uno por *Predictive Mean Matching* para datos multi-nivel, que tiene en cuenta la naturaleza agrupada y anidada de los datos (por SESIÓN dentro de SUJETO). El método consiste en ajustar una regresión lineal mixta a los datos observados usando como variable respuesta a la que tiene

datos faltantes a imputar y como predictoras a las demás (y con factores aleatorios), que podrían no ser todas. Después se usa dicha regresión para predecir cada dato faltante de la variable en cuestión. A continuación se elige de los valores observados, para cada entrada faltante, un conjunto de candidatos («donantes») de 3 a 10 miembros (por lo general). Dichos «donantes» tienen valores observados cercanos (definidos por alguna métrica de distancia) al valor predicho para el faltante y se selecciona un «donante» del conjunto al azar. Imputaciones fuera del rango de valores observados no ocurren, evitando, por ejemplo, valores negativos. Las variables se imputan secuencialmente en orden de menor a mayor cantidad de datos faltantes. Una vez que se imputa una variable, esta se trata dentro del ciclo como variable con todos los datos observados para imputar las que siguen. Los parámetros del modelo de imputación (por ejemplo, las componentes de varianza del modelo mixto) se muestrean de una distribución posterior bayesiana. Cada iteración usa los valores predichos por el modelo de imputación de la iteración anterior y los valores muestreados de la posterior bayesiana. Los valores imputados de la última iteración (cuando se llega a una distribución estable) constituyen los datos del $m - \text{ésimo}$ conjunto de datos. Ahora bien, es preferible poner como predictoras para la imputación a las variables con más datos completos en la fila donde se encontraba el dato faltante. El coeficiente de *Outflux* (Van Buuren, 2013, p. 99) reza:

$$O_j = \frac{\sum_j^p \sum_k^p \sum_i^n r_{ij} (1 - r_{ik})}{\sum_k^p \sum_i^n 1 - r_{ij}}$$

donde hay $i = 1, \dots, n$ datos y $j = 1, \dots, p$ es la j -ésima variable y_j ; y es $k = 1, \dots, p$ la k -ésima variable y_k , con $k \neq j$. Entonces: $R[i, l]; l \in \{j, k\}$ es la matriz indicadora $n \times p$ donde $r_{il} = 1$ es la celda del dato $n - \text{ésimo}$ de la variable $l - \text{ésima}$ donde en y_{il} hay un dato observado y $r_{il} = 0$ donde y_{il} falta un dato. El coeficiente calcula el número de pares de variables (Y_j, Y_k) con Y_j observado y Y_k faltante dividido por el número total de celdas con datos *incompletos*. Es un indicador de la utilidad de la variable y_j para imputar a otras variables. Una variable $j - \text{ésima}$ con todos sus datos observados tendrá $O_j = 1$; por el contrario si dicha variable tuviera todos sus datos faltantes $O_j = 0$, es decir, sería inútil como predictora para imputar. Se eliminaron pues aquellas potenciales predictoras con $O < 0,5$: FAM, IMA, CONC.

Asimismo resultaría deseable no elegir predictoras sin correlación con la variable a imputar. Por ello no se eligieron aquellas con coeficiente de correlación de Spearman mayor a 0,1 (para calcularla las variables categóricas se convirtieron en discretas). Con todo, los predictores elegidos para cada variable a imputar resultaron: (i) LEXESP: ADJ, MORF, EST2, EST4; (ii) FAM: ADJ, ANIM, STEM, MORF, ES, LEXESP, EST1, EST2; (iii) IMA: ADJ, ANIM, MOD, LEXESP, EST1, EST2, EST3; (iv) CONC: ADJ, MOD, ANIM, STEM, LEXESP, EST1, EST2, EST3, EST5, EST6.

Luego se aplicó el algoritmo *mice* con $m = 5$ con las predictoras seleccionadas para cada variable a imputar. Luego se agregaron las estimaciones de las imputaciones usando la mediana. Ahora bien, los valores de las variables del controlador son iguales para instancias donde se repite un mismo controlador. No obstante, las estimaciones surgidas de la imputación pueden ser diferentes. Por ende, se agregaron las imputaciones estimadas para un mismo controlador mediante la mediana.

4.4.2. Transformaciones.

Se transformaron al logaritmo los atributos que tienen que ver con las frecuencias y los errores acumulados, sumándoles una unidad; de la forma que sigue:

- Frecuencia de controlador (Corpus «BuscaPalabras»): $LOG.LEX = \log(LEXESP + 1)$.
- Frecuencia absoluta de TYPE en corpus propio: $LOG.Fabs_C = \log(Fabs_C + 1)$.
- Frecuencia absoluta de TYPE en corpus *EsTenTen*: $LOG.Fabs_S = \log(Fabs_S + 1)$.
- Frecuencia por millón de TYPE en corpus *EsTenTen*: $LOG.Fmill_S = \log(Fmill_S + 1)$.
- Errores acumulados: $CUMRES = \log(CUMRES + 1)$.

4.4.3. Colinealidad.

Colinealidad es la presencia de correlación entre atributos cuantitativos. La siguiente gráfica muestra la *correlación de Spearman* entre las variables extraídas de «BuscaPalabras» Imaginabilidad (IMA), Concretud (CONC), Familiaridad (FAM), Frecuencia en escala logarítmica (LOG.LEX); así como aquellas que expresan el logaritmo de la frecuencia de los «tipos de concordancia»: TYPE en el Corpus de datos (LOG.Fabs_C) y en el de *EsTenTen* de *Sketch Engine* (Log.Fabs_S). Las variables fueron escaladas ($\frac{x - \text{mean}(x)}{\text{sd}(x)}$). Se observa una correlación positiva moderada (0.68) entre Imaginabilidad y Concretud, lo cual es esperable: cuanto más concreta es una palabra más fácil resulta evocar imágenes de ésta. Además existe una correlación positiva entre Familiaridad y Frecuencia (0.50): cuánto más frecuente resulta una palabra, más familiar parece. También resulta alta la correlación entre las frecuencias de TYPE (0.88). Sin embargo, nótese que éstas no se hallan correlacionadas con las primeras.

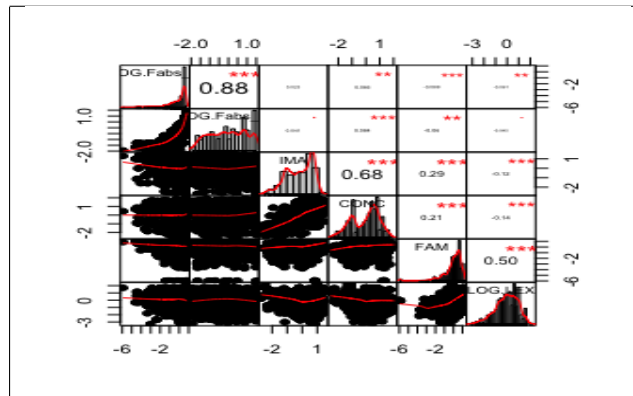


Figura 4.1: Correlación de Spearman entre: Imaginabilidad (IMA), Concretud (CONC), Familiaridad (FAM) y $\log(\text{FREQ} + 1)$, $\log(\text{Fabs}_S + 1)$, $\log(\text{Fabs}_C + 1)$.

Se aplicó un Análisis de Componentes Principales (PCA, por sus siglas en inglés). La técnica permite obtener nuevas variables ortogonales llamadas componentes principales, que se calculan como combinación lineal de las variables cuantitativas originales (Peña, 2002). Se aplicó la técnica a los grupos de variables de frecuencia de TYPE y a las relacionadas con el controlador por separado, ya que resulta indispensable que las variables estén correlacionadas entre sí. Se calcularon los componentes basados en la matriz de correlación. En los Cuadros de a continuación se observan los autovalores, el porcentaje de varianza total y la varianza acumulada para cada componente, según el grupo de variables. En cuanto a las variables del controlador, los dos primeros componentes representan el 78% de la variabilidad total. Esto también se observa en el «Scree Plot» que le sigue; donde se nota una caída abrupta de la varianza total explicada en el tercer componente. Por ende, se concluye que dos componentes resultan suficientes para explicar las cuatro variables. En lo que atañe a las otras dos variables, un solo componente basta para explicar el 88% de la varianza.

Componente	Auto-valor	% varianza explicada	% varianza acumulada
PC1	1.76	44.21	44.21
PC2	1.41	35.30	79.52
PC3	0.52	13.08	92.60
PC4	0.29	7.39	100

Cuadro 4.16: Variables del controlador: Autovalores y porcentaje de la varianza a partir de la matriz de correlación.

Componente	Auto-valor	% varianza explicada	% varianza acumulada
PC1	1.77	88.54	88.54
PC2	0.22	11.45	100

Cuadro 4.17: Variables de frecuencia TYPE: Autovalores y porcentaje de la varianza a partir de la matriz de correlación.

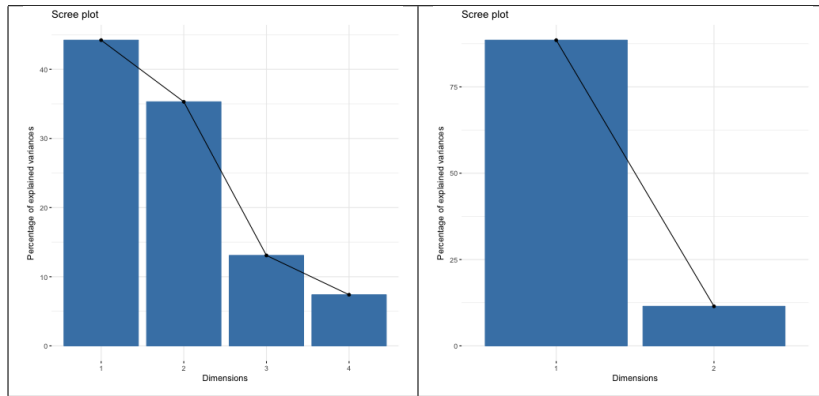


Figura 4.2: Scree Plot. Izquierda: variables del controlador; Derecha: variables de frecuencia TYPE.

El Cuadro que sigue muestra los pesos de las variables con respecto a las dos primeras componentes; y el otro que le sigue, la correlación entre las variables y dichas componentes. Las componentes se construyen como las siguientes combinaciones lineales:

$$PC1("IMA.CONC") = 0,68 \times IMA + 0,68 \times CONC + 0,10 \times FAM - 0,22 \times LOG.LEX$$

$$PC2("FAM.LEX") = 0,09 \times IMA - 0,01 \times CONC + 0,72 \times FAM + 0,68 \times LOG.LEX$$

$$PC1("Fabs.C.S") = 0,70 \times LOG.Fabs_S + 0,70 \times LOG.Fabs_C$$

Es decir que, en el caso de las variables del controlador, en la primera componente (se llamará «IMA.CONC») cargan las variables Imaginabilidad y Concretud; y en la segunda (se llamará «FAM.LEX»), Familiaridad y LOG.LEX. En lo que atañe a las correlaciones, se observa que IMA y CONC correlacionan positivamente y en alto grado con la primera componente; y lo propio sucede con FAM y LOG.LEX respecto de la segunda componente. Correlación positiva indica que a mayor / menor valor en la componente aumenta / disminuye la variable en cuestión. En lo que respecta a las variables de frecuencia de TYPE, en la primera componente cargan con fuerza LOG.Fabs_S y LOG.Fabs_C, lo cual se refleja también en la muy alta correlación positiva entre la primera componente y dichas variables. En suma, se resolvió la colinealidad pasando de cuatro variables de controlador con correlación a dos componentes sin correlación entre sí; y de dos variables de frecuencia de TYPE con correlación fuerte a una sola.

	PC1	PC2
IMA	0.68	0.09
CONC	0.68	-0.01
FAM	0.10	0.72
LOG.LEX	-0.22	0.68
LOG.Fabs_S	0.70	-
LOG.Fabs_C	0.70	-

Cuadro 4.18: Pesos de las variables con respecto a las dos primeras componentes.

	Cor(PC1, var)	Cor(PC2, var)
IMA	0.91	0.11
CONC	0.91	0
FAM	0.14	0.85
LOG.LEX	-0.29	0.81
LOG.Fabs_S	0.94	-
LOG.Fabs_C	0.94	-

Cuadro 4.19: Correlación entre las variables y las dos primeras componentes.

4.4.4. Discretización.

Se decidió discretizar los atributos cuantitativos utilizando *clustering* por mezcla de gaussianas (Pang Ning Tang, 2005; cap. 9). Se supone que las observaciones vienen de una mezcla de diferentes distribuciones de probabilidad, generalmente normales (multivariadas); que modelan *clusters* elipsoides (centrados en la media de cada distribución y cuya forma está determinada por la matriz de covarianzas de cada distribución). El procedimiento mixto es: (i) dadas k distribuciones (generalmente del mismo tipo) pero con diferentes parámetros, seleccionar una y generar un objeto x_i (u observación) a partir de ésta; (ii) repetir k veces. Formalmente se tiene:

- k distribuciones gaussianas (multivariadas) en total; m objetos (observaciones) en total, cuyo conjunto será: $\chi = \{x_1, \dots, x_m\}$, con $i = 1, 2, \dots, m$.
- la distribución j –ésima posee el vector de parámetros $\theta_j = \{\mu_j; \Sigma_j\}$, con $j = 1, 2, \dots, k$.
- Θ es el vector de todos los vectores de parámetros de las k distribuciones: $\Theta = \{\theta_1, \dots, \theta_k\}$.
- Se denota $f(x_i | \theta_j)$ a la función de densidad que arroja la probabilidad de que la observación i –ésima provenga de la j –ésima distribución.
- la probabilidad de que la distribución j –ésima sea elegida para generar el objeto x_i está dada por el peso w_j ($1 \leq j \leq k$); sujetos a la condición de que sumen uno: $\sum_{j=1}^k w_j = 1$.

Teniendo en cuenta lo enumerado, la probabilidad de que *una* observación / objeto x_i dados los parámetros de cada una de las distribuciones componentes, será:

$$p(x_i | \Theta = \{\theta_1, \dots, \theta_k\}) = \sum_{j=1}^k w_j f_j(x_i | \theta_j)$$

Si las observaciones son independientes, entonces la probabilidad del conjunto de todas las observaciones es simplemente el producto de las probabilidades de cada objeto individual x_i :

$$p(\chi | \Theta = \{\theta_1, \dots, \theta_k\}) = \prod_{i=1}^m p(x_i | \Theta) = \prod_{i=1}^m \left[\sum_{j=1}^k w_j f_j(x_i | \theta_j) \right]$$

A modo de ejemplo, la densidad de probabilidad para la distribución gaussiana unidimensional en el punto x_i es:

$$f_j(x_i | \theta_j) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x_i - \mu)^2}{2\sigma^2}}$$

Entonces la mezcla de dos distribuciones gaussianas para el punto x_i asumiendo igual desvío típico $\sigma = 2$ pero diferente media $\mu_j = \{\mu_1 = -4; \mu_2 = 4\}$ y pesos iguales $w_1 = w_2 = 0,5$; se escribe:

$$p(x_i | \Theta = \{\theta_1 = \{\mu_1 = -4, \sigma = 2\}; \theta_2 = \{\mu_1 = 4, \sigma = 2\}\}) = \sum_{j=1}^k w_j f_j(x_i | \theta_j) =$$

$$(0,5) \frac{1}{2\sqrt{2\pi}} e^{-\frac{(x+4)^2}{8}} + (0,5) \frac{1}{2\sqrt{2\pi}} e^{-\frac{(x-4)^2}{8}}$$

La Figura que sigue ilustra 20.000 observaciones generadas con el modelo ejemplificado, usando *R*.

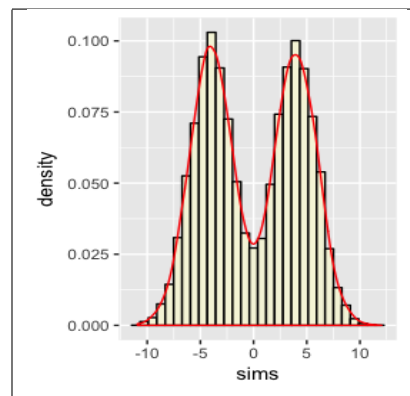


Figura 4.3: Observaciones ($n=20000$) generadas por un modelo mixto con: $\theta_1 = \{\mu_1 = -4, \sigma = 2\}; \theta_2 = \{\mu_1 = 4, \sigma = 2\}$

Cada distribución componente se considera un grupo o *cluster* diferente. En la realidad, dadas las observaciones, no se sabe cuál punto fue generado por cuál distribución; entonces, en principio, no se puede calcular la probabilidad de cada punto. Sin embargo, mediante el algoritmo *E – M (Expectation-Maximization)* es posible: (i) estimar los parámetros de las distribuciones componentes a partir de los datos; (ii) identificar cuáles observaciones pertenecen a qué distribución; o sea la probabilidad de que un determinado objeto pertenezca a un *cluster* en particular. Se utilizó el paquete *mclust* de *R* (Strucua et al., 2016) para hacer el *clustering*. En el caso unidimensional, que nos ocupa, el paquete prueba un modelo con igual varianza para las componentes (E) y uno con diferente varianza (V). La selección entre modelos se hace con la medida de información *BIC (Bayesian Information Criterion)*: $BIC = -2L(\theta) + p \log(n)$ [n = número de observaciones; p = número de parámetros; $L(\theta)$ = logaritmo de la verosimilitud del modelo]. Se discretizaron los atributos: MORF, STEM, FAM, CONC, IMA, LOG.LEX, IMA.CONC, FAM.LEX, CUMRES, Fabs.S.C, LOG.Fabs_C, LOG.Fabs_S. El atributo FAM.LEX se discretizó en dos categorías poniendo como punto de corte a la mediana; ya que el *clustering* no resultó efectivo. El Cuadro que sigue muestra la discretización de cada variable con los valores de los atributos pertenecientes a cada grupo (o categoría).

Atributo	Descripción	Discretización	Casos	Ejemplos del Corpus
MORF.f	Similaridad entre terminaciones	0 = [2,8;2,2;2,6;2,4;3)	159	«mis amigos» (2.8); «vacaciones agradables» (2.6)
		1 = [3;3,2;3,4)	1258	«las personas» (3); «mujeres jóvenes» (3.2)
		2 = [3,4;3,6]	440	«los trenes» (3.4); «relaciones industriales» (3.6)
STEM.f	Similaridad entre raíces léxicas	0 = [1,8;4)	1499	«todas reglas» (1.8); «los grupos» (2.9)
		1 = [4;10,2]	358	«alemanes fieles» (4.5); «mujeres guapas» (5.8)
FAM.f	Familiaridad	0 = [1,88;5,74)	348	«los sultanes» (3.46); «los archivos» (5.30)
		1 = [5,74;7]	1509	«otras deudas» (5.81); «muchas aventuras»(6)
CONC.f	Concretud	0 = [1,80;4,22)	714	«estos casos» (1.88); «los lenguajes» (3.93)
		1 = [4,22;6,82]	1143	«las costumbres» (5.07); «los árboles» (5.87)
IMA.f	Imaginabilidad	0 = [1,64;5,75)	1170	«los conocimientos» (1.64); «unas horas» (4.68)
		1 = [5,75;6,90]	687	«las iglesias» (5.82); «estas personas» (6.22)
LEX.f	Frecuencia: LOG.LEX	0 = [1,16;4,39)	931	«muchas librerías» (1.93); «los cantos» (3.10)
		1 = [4,39;7,22]	926	«sus juegos» (5.08); «mujeres famosas» (6.20)
IMA.CONC.f	PCA1	0 = [-3,48;0,58)	1163	«nuevos conocimientos» (-3.11); «los servicios» (-1.19)
		1 = [0,58;2,35]	694	«muchas personas» (1.31); «los hospitales» (2.17)
FAM.LEX.f (*) (*) corte: mediana	PCA2	0 = [-4,24;0,17)	934	«los sultanes» (-4.42); «las comodidades» (-0.64)
		1 = [0,17;1,98]	923	«los años» (1.22); «los hombres» (1.89)
Fabs.SC.f	PCA1	0 = [-5,56;0,45)	1007	«alemanes ingenuos» [n-j-*es-os] (-0.54)
		1 = [0,45;1,61]	850	«los latinos» [l-n-os-os] (1.61)
Fabs_C.f	LOG(Frecuencia de TYPE) en Corpus	0 = [0,69;4,09)	1263	«alemanes ingenuos» [n-j-*es-os] (1.79)
		1 = [4,09;5,54]	594	«los latinos» [l-n-os-os] (5.54)
Fabs_S.f	LOG(Frecuencia de TYPE) en EsTenTen (Sketch Engine)	0 = [0,69;15,24)	1009	«alemanes ingenuos» [n-j-*es-os] (11.53)
		1 = [15,24;16,05]	848	«los latinos» [l-n-os-os] (15.96)
CUMRES.f	Errores Acumulados	0 = [1;0,69) = [0;2)	509	«los idiomas» (0); «los profesores» (0.69)
		1 = [0,69;1,94) = [2;7)	709	«muchos lugares» (1.38); «las bromas» (1.6)
		2 = [1,94;3,33) = [7;28)	639	«los estudiantes» (2.63); «las ciudades» (2.89)

Cuadro 4.20: Discretización de atributos utilizando *clustering* por mixtura de gaussianas (*mclust*).

Se discretizó el atributo MORF en tres categorías y el resto en dos. En el primer grupo de MORF (valores de 2.2 a 2.8) se incluyen casos cuyo primer término es un posesivo o bien con un término terminado en *-es* pero *sin -e-* epentética («*agradable-s*»). En el segundo (valores: 3, 3.2) hay concordancias del tipo *i > os* o *e > as* - que son casos típicos -; o bien casos donde la terminación en italiano es *-e*, entonces: *e > es* requiere solamente agregar *-s* (*donn - e > mujer - es*). El último grupo (valores 3.4, 3.6) contiene los casos con *-e-* epentética en uno o ambos términos y los casos que en italiano son invariantes: *cittá > ciudades*. El atributo STEM se divide en dos grupos de acuerdo al grado alto / bajo la similaridad entre raíces. Así el segundo grupo indica raíces menos similares entre español e italiano: *tedesch - i > aleman - es*; *donn - e > mujer - es*; *ditt - e > empres - as*; *divers - i > diferent - es*. Luego hay Discretización de propiedades del controlador de acuerdo a: (i) Familiaridad (FAM.f): alta (*amigos*) / baja (*sultanes*); (ii) Concretud (CONC.f): alta (*árboles*) / baja (*lenguajes*); (iii) Imaginabilidad (IMA.f): alta (*personas*) / baja (*horas*); (iv) Frecuencia en escala logarítmica (LEX.f): alta (*mujeres*) / baja (*cantos*). El atributo CUMRES.f discretiza el acumulado de errores en tres categorías: (i) «0»: hasta dos errores acumulados; (ii) «1»: de 3 a 7 errores acumulados; (iii) «2»: de 8 hasta 28 errores acumulados.

En cuanto a las variables de frecuencia (en escala logarítmica) de TYPE se tiene: (i) Frecuencia de TYPE en el corpus (Fabs_C.f) y en el corpus *EsTenTen* (Fabs_S.f): alta (*los latinos: [l-n-os-os]*) / baja (*alemanes ingenuos: [n-j-*es-os]*). En lo que atañe a IMA.CONC, la categoría «0» tiende a incluir a los controladores de baja concretud e Imaginabilidad (= 0); por ejemplo: *conocimientos* (IMA = 1.64; CONC = 2.56). En cambio, la categoría «1» tiende a contener aquellos de alta Imaginabilidad y Concretud (= 1), por ejemplo: *hospitales* (IMA = 6.45, CONC = 6.61). En lo que respecta a FAM.LEX, la categoría «0» tiende a incluir a controladores de baja frecuencia y familiaridad: *sultanes* (FAM= 3.46; LOG.LEX= 1.59); y la categoría «1», a controladores de alta frecuencia y familiaridad: *años* (FAM=6.68; LOG.LEX=5.84). Por otra parte,

Fabs.SC.f tiende a incluir en la categoría «0» frecuencias de TYPE altas tanto en el corpus propio como en *EsTenTen: los latinos* [*l-n-os-os*] (Fabs_C=5.54; Fabs_S=15.96); y la categoría «0» a aquellos TYPE de frecuencia baja en ambos corpus: *alemanes ingenuos* [*n-j-*es-os*] (Fabs_C=1.79; Fabs_S=11.53).

Capítulo 5

Análisis exploratorio.

Este capítulo versa sobre el análisis exploratorio de las variables descriptas en el capítulo anterior. Específicamente se tratarán: (i) gráficos descriptivos; (ii) análisis de Correspondencias Múltiple (MCA, por sus siglas en Inglés); (iii) un método de *clustering* para la agrupación de los casos de concordancia; (iv) selección de atributos para las variables respuesta binaria y categórica. Se usará el lenguaje de programación *R* (varias librerías).

5.1. Agrupamiento de casos.

Esta sección describe el agrupamiento de casos de concordancia usando *clustering* jerárquico. Se utilizaron las siguientes variables discretizadas: ADJ, LDA, POS, DIS, GRAMS, ANIM, ES, MORF.f, STEM.f, IMA.CONC.f, FAM.LEX.f, CUMRES.f, Fabs.SC.f. Se tomaron solamente atributos que expresan características de la instancia de concordancia. Por ello, no se incluyeron las «estrategias». Como era necesario resaltar los casos de controladores inanimados, y de baja concreción, imaginabilidad y frecuencias, así como las configuraciones de concordancia (TYPE) de baja frecuencia; se invirtieron las categorías asignadas a dichas variables. Por ejemplo, ahora: ANIM = 1 = «inanimado», IMA.CONC.f = 1 = «baja concreción e imaginabilidad»; FAM.LEX.f = 1 = «baja familiaridad y frecuencia»; Fabs_C.f = 1 = «TYPE de baja frecuencia en el Corpus». La idea fue otorgar más peso a las variables con categorías no iguales a «0». Para ello, se consideró a la base de datos como si fuera una matriz $t \times d = \text{palabras} \times \text{documentos}$; donde las «palabras» son los atributos y los «documentos» son las configuraciones de atributos; por ejemplo: {ADJ = 1, LDA = 1; POS = 2; ...}. Entonces se asignó a cada atributo / palabra un peso de acuerdo al número de «ocurrencias» de dicho atributo en el documento / caso según el esquema:

$$w_{t,d} = tf \times idf = \log(tf_{t,d} + 1) \times \left(\log\left(\frac{N}{df_t}\right) + 1 \right)$$

donde: $tf_{t,d}$ es la cantidad de veces que aparece el atributo en un caso; N es la cantidad de documentos / casos; df_t es la cantidad de documentos / casos en la colección en donde ocurre el atributo. El peso aumenta con la cantidad de ocurrencias de un atributo en un documento / caso y con su rareza en la colección de casos. Por ende, con el objetivo de resaltar atributos, se sumó 1 unidad a cada atributo con categoría «1» o «2»; entonces, por ejemplo, el atributo «animado» tiene ahora las categorías: 2 = *no animado*; 0 = *animado*; interpretadas como frecuencias absolutas. A continuación se aplicó *latent semantic analysis* (Landauer et al., 1998; Jurafsky & Martin, 2017, cap. 16). La técnica aplica descomposición de valores singulares [SVD, por sus siglas en inglés] a una matriz $Z = t \times d = \text{palabras} \times \text{documentos}$ (aquí la matriz de pesos); la cual queda factorizada como $Z = W\Sigma C^T$. La matriz W resulta $t \times m$ donde las filas t son palabras y las columnas $m = \text{rango}[Z]$ ¹ son dimensiones en el espacio latente ortogonales entre sí y ordenadas según la cantidad de varianza explicada de los datos de la matriz de pesos. La matriz Σ es $m \times m$ y diagonal con autovalores, indicando la importancia de cada dimensión. Por último, C^T es $m \times d$; con los documentos en las columnas. Como las dimensiones están ordenadas por importancia (varianza) explicada se toman las

¹El rango de una matriz es la cantidad de columnas linealmente independientes.

primeras $k \leq m$ dimensiones con lo cual se obtiene: $Z_{t \times d} = W_{t \times k} \Sigma_{k \times k} C_{k \times d}^T$. Luego se aplicó la medida de similaridad «coseno» (equivalente a la correlación) entre los documentos (casos) de la matriz Z surgida de LSA, según (Manning *et al.*, 2008; p. 111):

$$\text{sim}(d_i, d_j) = \frac{\vec{d}_i \vec{d}_j}{\|\vec{d}_i\| \|\vec{d}_j\|}$$

donde el numerador es el producto interno de los vectores de los documentos / casos \vec{d}_i y \vec{d}_j ; y el denominador es el producto de sus normas (el largo de un vector según la norma Euclídea). Haciendo $1 - \text{sim}(d_i, d_j)$ se obtiene una matriz de distancias (o disimilaridad).

Dicha matriz de distancias fue el insumo para llevar a cabo un *clustering* jerárquico aglomerativo. Dado un conjunto de n objetos (instancias de concordancia) descriptos por un conjunto de p variables (atributos), el *clustering* determina las clases (grupos, *clusters*) a las que estos objetos pertenecen. Dichos objetos se agrupan de modo que el grado de similitud / asociación entre los miembros del grupo / *cluster* sea más fuerte que con los miembros de grupos diferentes. Los métodos *jerárquicos* encuentran particiones consecutivas en una jerarquía de clases anidadas. Si el método es aglomerativo, se parte de tantas clases como objetos haya y en pasos sucesivos éstas se juntan para obtener clases de objetos similares. El resultado se aprecia en un *dendograma* (diagrama en forma de árbol en dos dimensiones) donde se pueden ver las uniones o divisiones que se hacen en el proceso de construcción. En el gráfico, las ramas representan los *clusters*, los *nodos* muestran dónde se unen los *clusters* y la posición del nodo indica el nivel (similitud) en la cual ocurre la fusión entre *clusters*. El nodo raíz indica el nivel en que todos los objetos forman un único *cluster* (Pang Ning Tang, 2005; cap. 8). Se han propuesto diferentes medidas para establecer la distancia entre los *clusters*. El coeficiente de correlación cofenética indica la correlación entre las distancias originales (coseno) de los pares de objetos y aquellas del dendograma, o sea la distancia entre *clusters* a recorrer en el árbol para unir dos grupos. A continuación se tabula dicho índice contra las diferentes medidas de distancia entre *clusters*. Se eligió, de acuerdo al valor más alto, la distancia «average», definida como el promedio de las distancias entre pares de objetos, uno de cada *cluster*.

	coef_cofenetico
average	0.7615
complete	0.5457
single	0.5578
ward.D	0.5098
ward.D2	0.6394
median	0.7211
centroid	0.7518
mcquitty	0.7057

Cuadro 5.1: Coeficiente cofenético según medidas de distancia entre grupos.

La gráfica que sigue muestra el dendograma. Se podría proponer una agrupación de como mínimo dos *clusters* ($k = 2$) y como máximo en 6 ($k = 6$), dependiendo del punto de corte de distancia entre *clusters* (medido por la escala de la izquierda) en el árbol, por ejemplo en 0.5 ($k = 2$) ó 0.15 ($k = 6$).

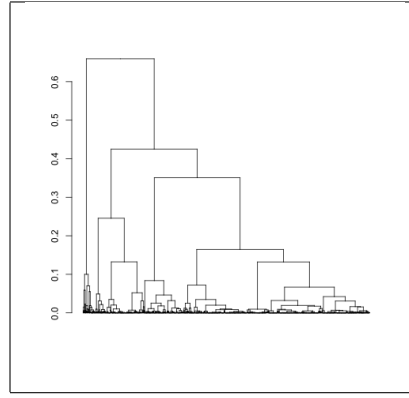


Figura 5.1: *Clustering* jerárquico aglomerativo: dendrograma.

Determinar el número k óptimo de *clusters* significa establecer la altura en la que se corta el dendrograma. Para ello se usarán las medidas que siguen (Halkidi et al., 2001; Hennig, 2013, 2019):

(1) suma de errores al cuadrado (SSW): la suma de las distancias al cuadrado entre sí, de los objetos dentro de cada *cluster* (cohesión), se generaliza como:

$$SSW = \frac{1}{2 \sum_{i=1}^k n_j(n_j - 1)} \sum_{j=1}^k \sum_{x \neq y \in C_j} dist(x, y)^2$$

donde C_j son los *clusters*, $n_j = |C_j|$ es la cardinalidad (número de objetos) del *cluster* y $dist$ una distancia (disimilaridad) cualquiera. Se trata de minimizarla (ya que así se maximiza la separación entre *clusters*) o bien de encontrar el k para el que la diferencia en SSW con $k - 1$ es mayor (primera diferencia).

(2) el coeficiente *Silhouette*: si es negativo la distancia promedio de un objeto a los demás en un *cluster* es mayor que la distancia promedio de dicho objeto a los objetos de otros *clusters* (que no lo contengan). Se debe maximizar. Se obtiene como sigue: (i) calcular a_i : la distancia promedio entre el objeto i -ésimo a los demás objetos dentro de un *cluster*; (ii) calcular b_i : la distancia promedio entre el objeto i -ésimo a los demás objetos y los objetos de otro *cluster* que no contenga al objeto i -ésimo; encontrando el mínimo de este valor para todos los *clusters*; (iii) calcular el coeficiente para el objeto i -ésimo: $S_i = (b_i - a_i) / \max(a_i, b_i)$; $S \in [-1, 1]$. Si $a_i > b_i$ entonces S_i será negativo; si $a_i < b_i$, será positivo y será igual a 1 cuando $a_i = 0$. Se maximiza: $\frac{1}{n} \sum_{i=1}^n S_{i,k}$; entre todos los k *clusters*.

(3) el coeficiente de Dunn: la razón entre la distancia mínima entre los *clusters* y la distancia máxima (dispersión) dentro de los *clusters*; si hay una buena clasificación el diámetro de los *clusters* es pequeño y la distancia entre ellos, grande. Entonces, se lo debe maximizar.

(4) el índice de separación [SINDEX]: está basado en las distancias $d(x_i, y)$ de cada punto x_i dentro de un *cluster* j -ésimo al punto más próximo y que *no* esté en el mismo *cluster*, y tomando aquella mínima: $d_{j:i} = \min_{y \notin C_j} d(x_i, y)$; $x_i \in C_j$ ($i = 1, \dots, n$; $j \in \{1, \dots, K\}$). A continuación se ordenan de menor a mayor los valores de $d_{j:i}$ para cada objeto $x_i \in C_j$; o sea: $d_{j:(i)} \leq \dots \leq d_{j:(n_j)}$. Luego se toma el 10% de dichas distancias por *cluster*, es decir $\lfloor pn_j \rfloor$; donde $\lfloor \rfloor$ es el número entero más grande $\leq pn_j$. Esto produce que se seleccionen los objetos de cada *cluster* más cercanos a los de otro *cluster*. El índice de separación debe minimizarse y es un promedio para todos los *clusters* del porcentaje de estas distancias:

$$I_{p-sep}(\mathcal{C}) = \frac{1}{\sum_{j=1}^k \lfloor pn_j \rfloor} \sum_{j=1}^K \sum_{i=1}^{\lfloor pn_j \rfloor} d_{j:(i)}$$

(5) Calinsky & Harabasz index [CH]: el índice maximiza: $\frac{B(k)(n-k)}{W(k)(k-1)}$; donde $W(k)$ es la suma de distancias al cuadrado *within* [SSW] y $B(k)$ la suma de distancias el cuadrado *between* [SSB]; ya que la suma de distancias al cuadrado total es $SST = SSB + SSW \implies SSW = SST - SSB$, se tiene:

$$W(k) = \sum_{h=1}^k \frac{1}{|C_h|} \sum_{x_i, x_j \in C_h} \text{dist}(x_i, x_j)^2$$

$$B(k) = \frac{1}{n} \sum_{i,j=1}^n \text{dist}(x_i, x_j)^2 - W(k)$$

(6) El coeficiente de Correlación Γ (Pearson-Gamma): maximiza el coeficiente de correlación de Pearson entre el vector de todas las distancias y un vector binario donde «0» indica que el par de objetos se encuentra en el mismo *cluster* y «1» indica que están en *clusters* diferentes.

La gráfica que sigue muestra las medidas para $k = 2, \dots, 10$. Se observa que para WSS (y su primera diferencia), $k = 4$; para SINDEXX, $k = 6$; según Pearson-Gamma, $k = 4$; DUNN arroja $k = 2$, CH establece $k = 8$ y Silhouette retorna $k = 8$. Proponer ocho *clusters* parece mucho y proponer dos, muy poco. Se exploran entonces las alternativas con cuatro y seis *clusters*.

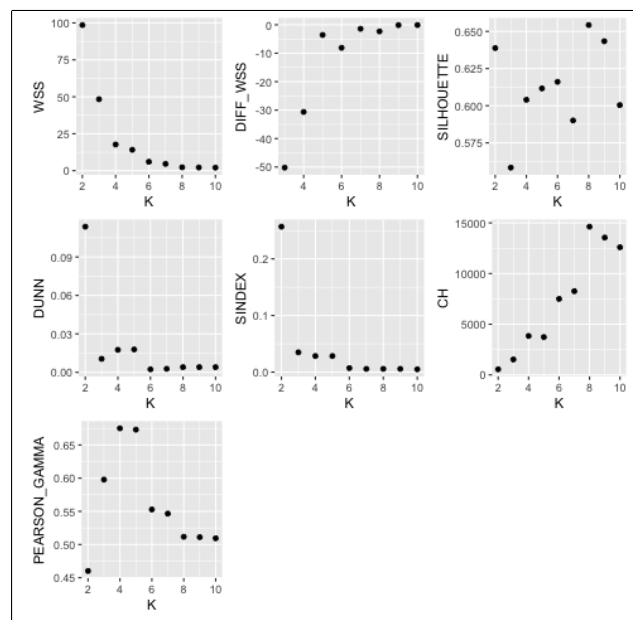


Figura 5.2: Medidas de validación interna.

Se compara pues una partición con cuatro *clusters* con otra de seis. Las gráficas de las Figuras 3 a 5 que siguen ilustran la participación de cada variable en la partición de cuatro y de cinco *clusters*. Comenzando con la partición de cuatro, se nota la siguiente composición de cada *cluster* como sigue:

(i) El cuarto *cluster* contiene concordancias sin artículos definidos y de larga distancia solamente, de dos términos en su mayoría y con *-e-* epentética en uno o ambos. El tipo de configuración [TYPE] es de baja frecuencia. Además contiene distancias medias y altas entre las desinencias [MORF.f].

(ii) El tercer *cluster* se distingue por contener solamente distancias de STEM altas. Por otra parte las distancias MORF.f son en su mayoría distancias medias. Además hay mayoritariamente artículos definidos.

(iii) El segundo *cluster* contiene, en su mayoría, distancias MORF altas y solamente casos de *-e-* epentética en un solo término. Por otro lado, consiste exclusivamente de distancias altas de raíz STEM.f.

(iv) Por último, el primer *cluster* posee mayoría de términos sin *-e-* epentética, y minoría en un solo término; y de distancia MORF media.

Nótese que las variables IMA.CONC.f, FAM.LEX.f y CUMRES.f no ayudan a caracterizar los *clusters*, como tampoco la mayoría de las estrategias. Asimismo, los *clusters* 3 y 4 se encuentran en orden ascendente de animicidad del controlador.

En lo que atañe a la partición con seis *clusters*, los grupos 2 y 4 no cambian, pero ahora el «4» pasa a ser el grupo «5». El *cluster* 3 se divide en dos grupos, el «4» y el «6» solo caracterizados por: (i) Grupo «4»: distancias medias MORF.f en su mayoría; instancias sin *-e-* epentética; (ii) Grupo «6»: distancias altas MORF.f en su mayoría; instancias con *-e-* epentética en un término. Como en el caso del grupo «3» con $k = 4$, ambos grupos tienen STEM.f alta. El grupo «1» se dividió en dos: los grupos «1» y «3» en esta nueva partición. El grupo «3» se caracteriza por tener en su mayoría más de dos términos de concordancia y por tener distancias MORF.f medias, STEM.f baja y sin *-e-* epentética. Por otro lado, el grupo «1» posee distancias MORF.f medias y altas; y un 25% de instancias con *-e-* epentética. En lo que atañe a la partición en 5 grupos, solamente se parte el grupo tres en dos grupos con idénticas características que en la partición de seis grupos. El resto de los grupos son iguales a la partición de $k=4$.

Con el objetivo de cuantificar el análisis gráfico que se acaba de hacer, se calcularon las medidas de *pureza* y *entropía* (Rendón et al., 2011) entre cada variable y la clasificación de 4 o 6 *clusters*. Para cada *cluster* j la «pureza» es el número de objetos en j que tienen la categoría i : $P_j = \frac{1}{n_j} \max_i (n_j^i)$. El índice se obtiene haciendo una suma de P_j , pesada por la proporción de objetos de cada *cluster* del total:

$$Pureza = \sum_{j=1}^K \frac{n_j}{n} P_j$$

donde: K es el número de *clusters*; n_j es el tamaño del *cluster* j y n es el número total de objetos. $Purity \in [0, 1]$, por ende, un *cluster* que consiste solamente de una categoría de una variable tendrá pureza máxima «1».

La *entropía*, por otra parte, mide la pureza de las categorías de cada *cluster*; es decir que si un *cluster* consiste solo de una categoría su entropía es cero y aumenta a medida que el *cluster* se vuelve más variado en categorías. En suma, al aumentar la pureza disminuye la entropía y viceversa. Se define, para cada *cluster*, como: $E_j = -\sum_i p_{ij} \log(p_{ij})$. La medida total es la suma pesada de las entropías de todos los *clusters*:

$$Entropía = \sum_{j=1}^K \frac{n_j}{n} E_j$$

El Cuadro siguiente muestra ambas medidas para cada variable, según la partición en tres o cuatro *clusters* (valores redondeados al segundo decimal). Las diferencias se encuentran en pocas variables: ADJ, GRAMS, ES, Fabs_C.f. En dichas variables, las entropías son más altas y las purezas más bajas en $k = 4$ respecto de $k = 6$. La partición de 4 grupos casi no difiere de aquella de 5 grupos. Por ende, las categorías se reparten más homogéneamente en la partición de cuatro / cinco respecto a la de seis para estas variables, haciendo que los grupos se distingan menos entre sí debido a la preponderancia de una categoría (menos puros).

	Pureza(k = 4)	Entropía(k = 4)	Pureza(k = 5)	Entropía(k = 5)	Pureza(k = 6)	Entropía(k = 6)
ADJ	0.577	0.942	0.577	0.942	0.619	0.919
LDA	1.000	0.000	1.000	0.000	1.000	-0.000
POS	0.990	0.021	0.990	0.021	0.990	0.021
DIS	0.987	0.025	0.987	0.025	0.987	0.025
GRAMS	0.683	0.786	0.683	0.768	0.831	0.589
ANIM	0.718	0.850	0.718	0.843	0.718	0.843
ES	0.849	0.397	0.887	0.314	0.887	0.277
MORF.f	0.785	0.596	0.811	0.542	0.811	0.505
STEM.f	0.978	0.123	0.978	0.123	0.978	0.115
IMA.CONC.f	0.589	0.974	0.601	0.969	0.601	0.967
FAM.LEX.f	0.535	0.994	0.535	0.994	0.539	0.991
CUMRES.f	0.382	0.989	0.382	0.989	0.390	0.988
Fabs.SC.f	0.593	0.942	0.593	0.940	0.683	0.807

Cuadro 5.2: Pureza y Entropía por variable, según partición de 4 a 6 clusters.

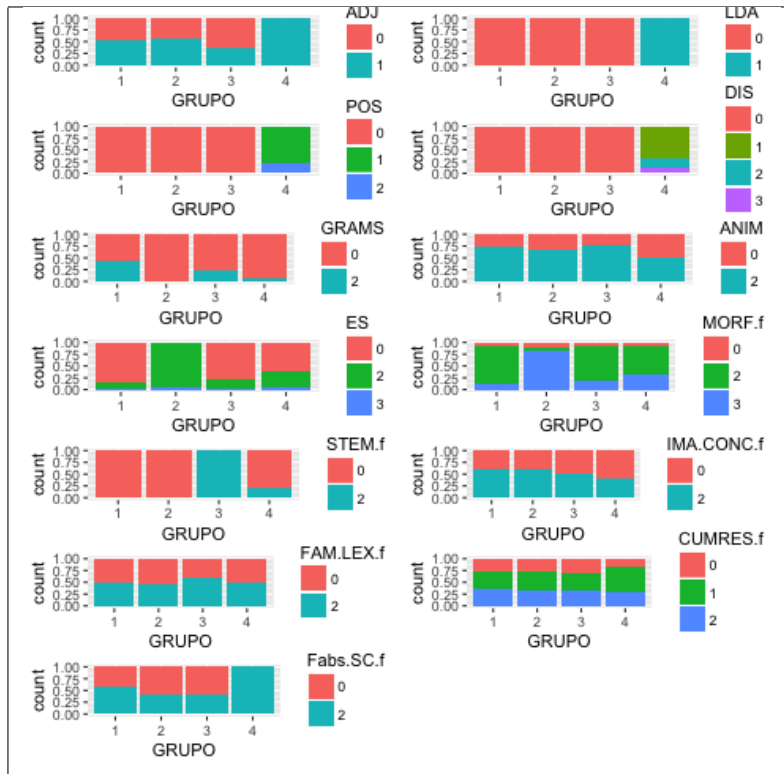


Figura 5.3: Análisis de variables para cuatro *clusters*.

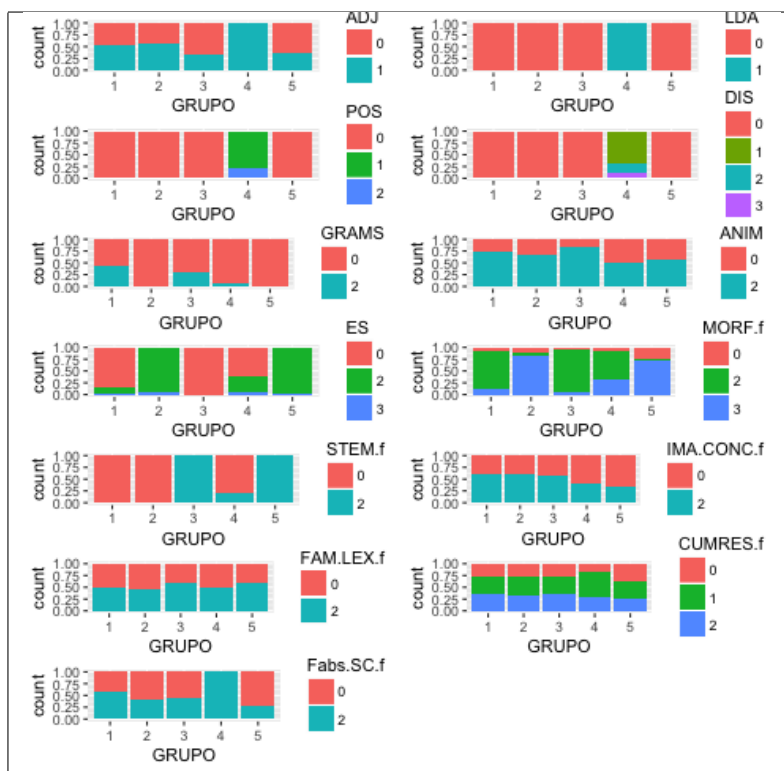


Figura 5.4: Análisis de variables para cinco *clusters*.

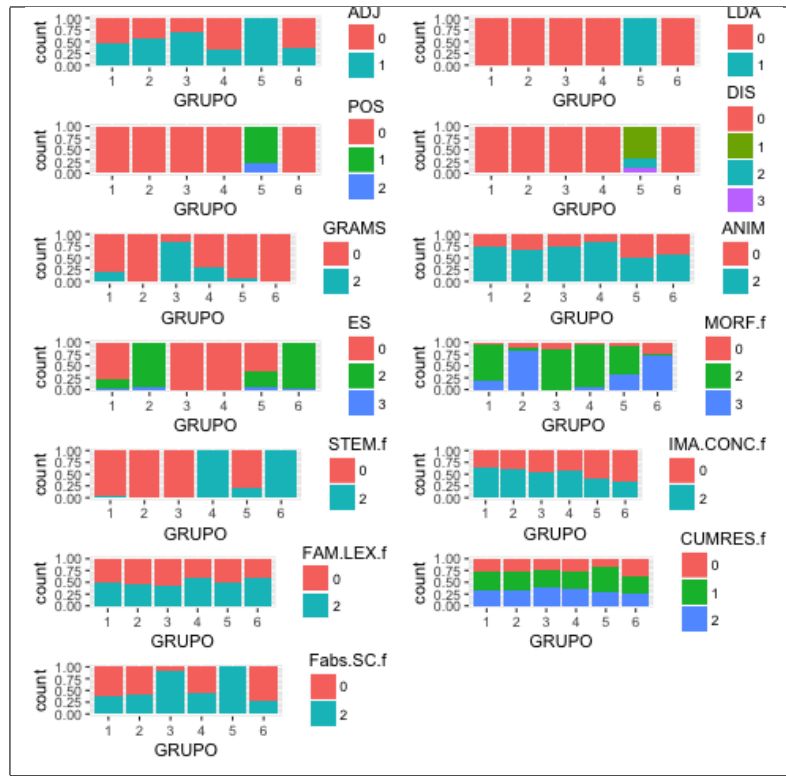


Figura 5.5: Análisis de variables para seis *clusters*.

Por último, es preciso evaluar la estabilidad de las particiones. Ello significa que un *cluster* «válido» no debería desaparecer ante perturbaciones de los datos; por ejemplo, datos que vengan de igual distribución, por adición de error o datos atípicos, o muestreo de subconjuntos de los datos. Sin embargo; estabilidad no implica validez, ya que un *cluster* sin significado para el investigador puede ser estable. El método aplicado se describe en Hennig (2007). En principio, por supuesto, se aplica el *clustering* a los datos originales según $E_n(\mathbf{x}_n)$; donde E_n es el método (aquí jerárquico aglomerativo) que mapea el conjunto de objetos $\mathbf{x}_n = \{x_1, \dots, x_n\}$ a una colección de grupos $\mathcal{C} = \{C_1, \dots, C_s\}$ de \mathbf{x}_n . A continuación, se repiten los pasos siguientes para $i = 1, \dots, B$:

1. Obtener una muestra \mathbf{x}_n^i de n puntos (vectores de atributos) de los datos originales \mathbf{x}_n , o bien usando muestreo con reposición o un subconjunto de $m = \lfloor n/2 \rfloor$ datos (« $\lfloor \cdot \rfloor$ » denota la parte entera), con muestreo sin reposición.
2. Computar el *clustering* $E_n(\mathbf{x}_n^i)$; donde E_n mapea el conjunto de objetos $\mathbf{x}_n^i = \{x_1, \dots, x_n\}$ a una colección de grupos $\mathcal{C} = \{D_1, \dots, D_r\}$ de \mathbf{x}_n^i . Notar que podría ser que $r < s$.
3. Denótese a $\mathbf{x}_*^i = \mathbf{x}_n \cap \mathbf{x}_n^i$ al conjunto de datos de la base original que están en la i -ésima muestra obtenida. Entonces, se denota a cada *cluster* en la clasificación original como $C_*^i = C \cap \mathbf{x}_*^i$; es decir, el cluster original asignado a los datos muestreados. Además, escríbase como $\Delta = E_n(\mathbf{x}_n^i) \cap \mathbf{x}_*^i$ a la nueva clasificación compuesta por $D = \{D_1, \dots, D_r\}$, $j = 1, \dots, r$.
4. Si el *cluster* $C_*^i \neq \emptyset$ (si no está vacío), entonces computar la medida de similitud de *Jaccard* entre C_*^i y D_r : $\gamma(C, D) = \frac{|C \cap D|}{|C \cup D|}$; o sea, la cantidad de puntos que están en ambos *clusters* respecto del total de puntos que pertenecen a, al menos, uno de los *clusters*. Luego, maximizar para $D_i \in \Delta$: $\gamma_{C,i} = \max_{D \in \Delta} \gamma(C_*^i, D)$. Si no hay máximo, $\gamma_{C,i} = 0$.

Esto genera, para cada *cluster* original, una secuencia $\gamma_{C,i}$, para $i = 1, \dots, B$. Por último se toma la media, que es el valor que se retorna como medida de estabilidad (para cada cluster original):

$$\bar{\gamma}_C = \frac{1}{B} \sum_{i=1}^B \gamma_{C,i}$$

Si $\gamma_{C,i} < 0,5$ el cluster se declara «disuelto» bajo la nueva partición. Se registra el número de disoluciones ocurridas en las B muestras. El cluster se declara «estable» si $\bar{\gamma}_C \geq 0,75$. El Cuadro que sigue muestra los resultados para las particiones de tres y cuatro *clusters*; usando ambos tipos de muestreo (con reposición y por subconjunto). Se establece $B = 10$. Entre paréntesis se pone el número de objetos de cada *cluster*. Las particiones resultan estables, sin ningún grupo disuelto. Sin embargo, se anotó más arriba que la partición en cuatro hace menos distinguibles los clusters entre sí. Tampoco resulta de interés proponer un solo *cluster* de 1205 objetos.

6 Clusters		Grupo	1	2	3	4	5	6
		n (%)	794 (42.75)	257 (13.83)	411 (22.13)	245 (13.19)	78 (4.20)	72 (3.87)
Con reposición	Jaccard: $\bar{\gamma}_C$		0.94	0.92	0.97	0.94	1	1
	Disueltos: $\gamma_{C,i} < 0,5$		0	0	0	0	0	0
	Recuperados		8	8	10	8	10	10
Sin reposición	Jaccard: $\bar{\gamma}_C$		0.87	0.89	0.91	0.91	1	1
	Disueltos: $\gamma_{C,i} < 0,5$		0	0	0	0	0	0
	Recuperados		8	7	8	7	10	10

5 Clusters		Grupo	1	2	3	4	5
		n (%)	1205 (42.75)	257 (13.83)	245 (22.13)	78 (4.20)	72 (3.87)
Con reposición	Jaccard: $\bar{\gamma}_C$		0.96	0.92	0.94	1	1
	Disueltos: $\gamma_{C,i} < 0,5$		0	0	0	0	0
	Recuperados		10	8	8	10	10
Sin reposición	Jaccard: $\bar{\gamma}_C$		0.95	0.89	0.91	1	1
	Disueltos: $\gamma_{C,i} < 0,5$		0	0	0	0	0
	Recuperados		10	7	7	10	10

4 Clusters		Grupo	1	2	3	4
		n (%)	1205 (64.88)	257 (13.83)	317 (17.07)	78 (4.20)
Con reposición	Jaccard: $\bar{\gamma}_C$		0.96	0.92	0.95	1
	Disueltos: $\gamma_{C,i} < 0,5$		0	0	0	0
	Recuperados		10	8	10	10
Sin reposición	Jaccard: $\bar{\gamma}_C$		0.95	0.89	0.93	1
	Disueltos: $\gamma_{C,i} < 0,5$		0	0	0	0
	Recuperados		10	7	10	10

Cuadro 5.3: Estabilidad de las particiones de tres a seis *clusters*.

En conclusión, se elige la partición en 6 grupos y se la denota como variable cualitativa «GRUPO6».

5.2. Gráficas descriptivas.

En esta sección se muestran gráficas de proporciones e intervalos de confianza de cada uno de los atributos *cualitativos* contra las variables respuesta binaria y categórica (para el tipo de respuesta simbólica ver cap. 10). El Cuadro que sigue muestra los casos (y su proporción del total) por niveles para ambas respuestas. En general los errores de concordancia son un cuarto del total (25.47%); y, de estos, más de la mitad son errores de plural (15.56%).

	RES_BIN	%	RES_CAT	%
0	1384	74.53	1384	74.53
1	473	25.47	62	3.34
2	-	-	32	1.72
3	-	-	289	15.56
4	-	-	90	4.85

Cuadro 5.4: Estadísticos descriptivos para RES_BIN y RES_CAT, en total.

Por otra parte, los Cuadros que siguen muestran el panorama desglosando por sujeto. En el caso de RES_BIN, la proporción de la categoría «0» constituye una medida de precisión (porcentaje correcto del total). Se nota el orden: JAKO > SONIA > MIRKA > NATI. Vale la pena observar que con solo 120 horas de español el alumno SONIA logra casi un 80% tomando en cuenta todas las sesiones. En cambio, MIRKA, con muchas más horas de práctica, alcanza casi el 70%. Esto muestra una vez más que el aprendizaje nunca el lineal: más horas no es igual a más precisión. En lo que atañe a los tipos de errores, el de plural siempre lleva la delantera. Los errores de género y por uso de terminación «(-e)-s» disminuyen en los alumnos de los niveles más altos (B2 y C1).

	SONIA (A1/A2)	NATI (B1)	JAKO (B2)	MIRKA (C1)
0 (%)	247 (79.17)	258 (66.15)	387 (83.41)	492 (71.20)
1 (%)	65 (20.83)	132 (33.85)	77 (16.59)	199 (28.80)

Cuadro 5.5: Estadísticos descriptivos para RES_BIN por sujeto.

	SONIA (A1/A2)	NATI (B1)	JAKO (B2)	MIRKA (C1)
0 (%)	247 (79.17)	258 (66.15)	387 (83.41)	492 (71.20)
1 (%)	13 (4.17)	34 (8.72)	8 (1.72)	7 (1.01)
2 (%)	15 (4.81)	11 (2.82)	1 (0.22)	5 (0.72)
3 (%)	28 (8.97)	60 (15.38)	58 (12.50)	143 (20.69)
4 (%)	9 (2.88)	27 (6.92)	10 (2.16)	44 (6.37)

Cuadro 5.6: Estadísticos descriptivos para RES_CAT por sujeto.

Los Cuadros de a continuación muestran la cantidad de casos y el porcentaje del total, de cada atributo cualitativo.

	ADJ	MOD	POS	DIS	LDA	ES	GRAMS	MORF.f	CUMRES.f
0 (%)	883 (47.55)	870 (46.85)	1779 (95.80)	1779 (95.80)	1779 (95.80)	1334 (71.84)	1268 (68.28)	159 (8.56)	509 (27.41)
1 (%)	974 (52.45)	37 (1.99)	60 (3.23)	54 (2.91)	78 (4.20)	495 (26.66)	589 (31.72)	1258 (67.74)	709 (38.18)
2 (%)	-	488 (26.28)	18 (0.97)	14 (0.75)	-	28 (1.51)	-	440 (23.69)	639 (34.41)
3 (%)	-	462 (24.88)	-	10 (0.54)	-	-	-	-	-

Cuadro 5.7: Estadísticos descriptivos para atributos cualitativos.

	ANIM	FAM.f	IMA.f	CONC.f	LEX.f	IMA.CONC.f	FAM.LEX.f	Fabs.SC.f	Fabs_C.f
0 (%)	1333 (71.78)	343 (18.47)	1177 (63.38)	665 (35.81)	883 (47.55)	1078 (58.05)	929 (50.03)	1007 (54.23)	1263 (68.01)
1 (%)	524 (28.22)	1514 (81.53)	680 (36.62)	1192 (64.19)	974 (52.45)	779 (41.95)	928 (49.97)	850 (45.77)	594 (31.99)

Cuadro 5.8: Estadísticos descriptivos para atributos cualitativos.

	Fabs_S.f	STEM.f	EST1	EST2	EST3	EST4	EST5	EST6	EST7
0 (%)	1009 (54.33)	1499 (80.72)	1302 (70.11)	1390 (74.85)	1850 (99.62)	1790 (96.39)	1606 (83.68)	1775 (95.58)	1777 (95.69)
1 (%)	848 (45.67)	358 (19.28)	555 (29.89)	467 (25.15)	7 (0.38)	67 (3.61)	251 (13.52)	82 (4.42)	80 (4.31)

Cuadro 5.9: Estadísticos descriptivos para atributos cualitativos.

5.2.1. Respuesta binaria.

Las gráficas muestran los niveles de los atributos cualitativos contra la variable respuesta binaria («0» = sin error; «1» = con error). Se grafica con un punto la proporción estimada de error por nivel y con una línea, su intervalo de confianza obtenido por *bootstrap*.

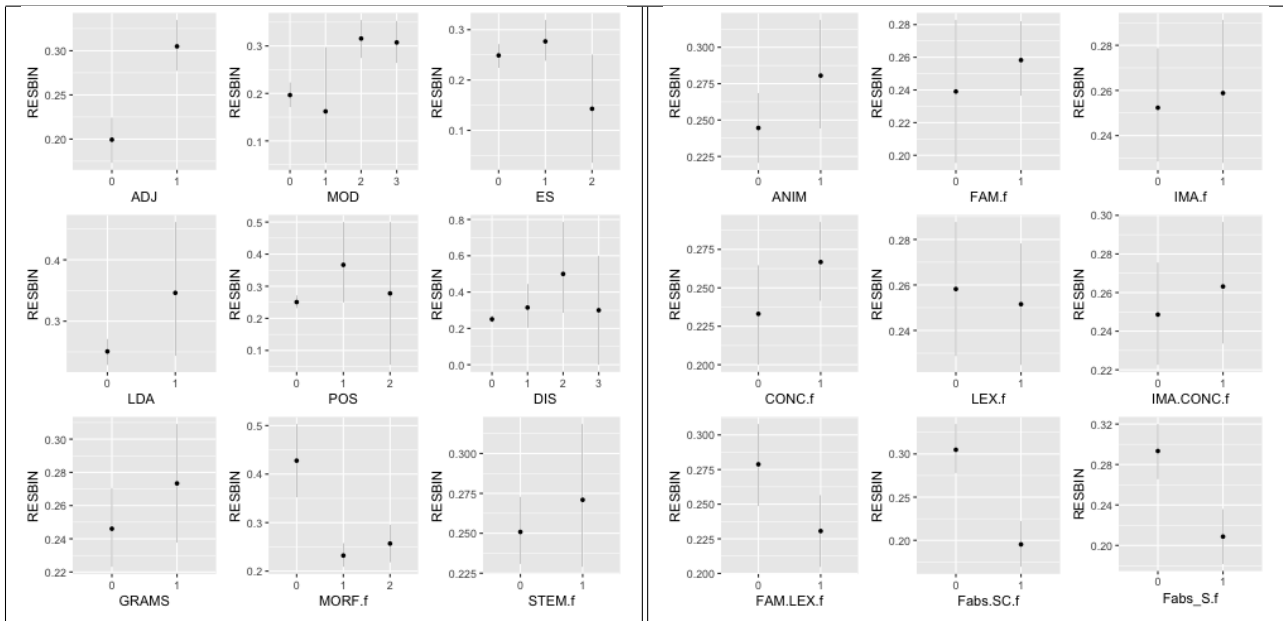


Figura 5.6: Proporciones por niveles de atributos, juntos a sus intervalos de confianza (bootstrap)

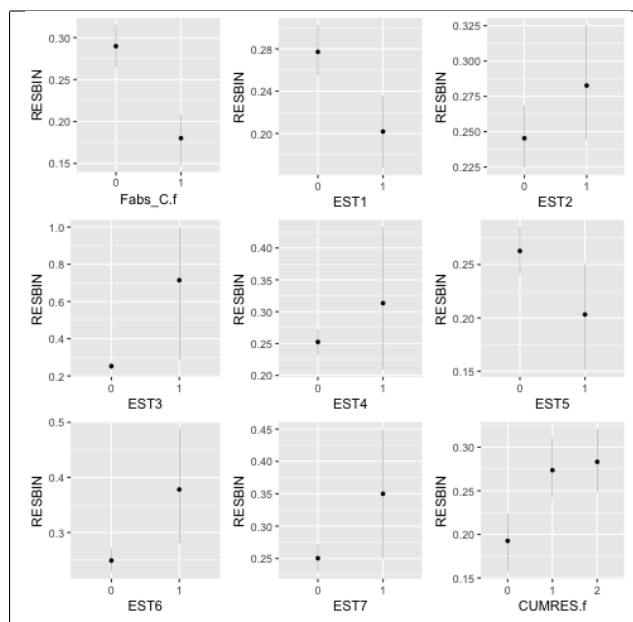


Figura 5.7: Proporciones por niveles de atributos, juntos a sus intervalos de confianza (bootstrap)

Lo importante es verificar que los intervalos de confianza no se superpongan, lo cual constituye un *indicio* de posible diferencia significativa estadísticamente. A ojo de buen cubero, se verifica que hay dife-

rencia entre proporciones de error entre las concordancias: (i) que tienen artículos definidos frente al resto [ADJ]; (ii) que tienen determinantes y adjetivos frente a los artículos definidos e indefinidos [MOD]; (iii) que poseen baja similaridad en las terminaciones respecto a los de similaridad media y alta [MORF.f]; (iv) que siguen patrones TYPE más frecuentes frente a los menos frecuentes, tanto en el corpus propio como en *EsTenTen*; (v) que siguen las estrategias: EST1, EST3, EST6; (vi) que tienen hasta 2 errores acumulados frente al resto. Ningún atributo del controlador presenta *indicios* de diferencias significativas entre proporciones de errores. Téngase también en cuenta que el largo del intervalo de confianza depende de la cantidad de casos: intervalos muy anchos implican pocos casos en el nivel del atributo.

5.2.2. Respuesta categórica.

Las gráficas de barras de a continuación ilustran el panorama cruzando cada atributo con los tipos de errores (de 1 a 4).

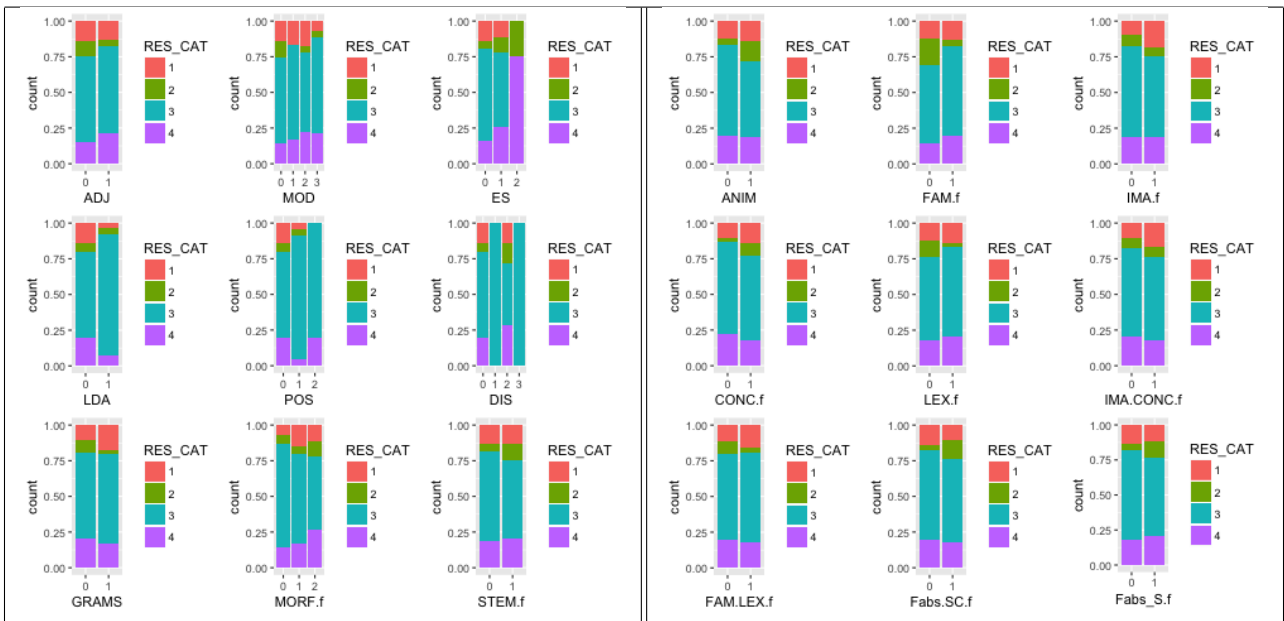


Figura 5.8: Cruzas entre atributos y tipo de error.

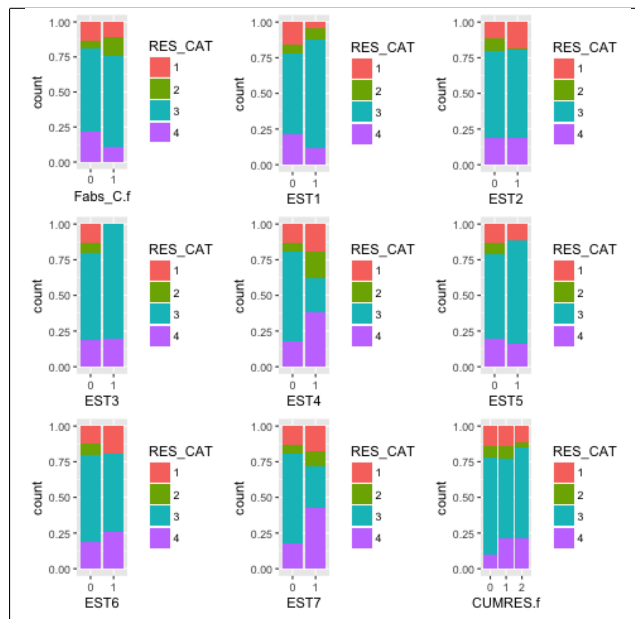


Figura 5.9: Cruzas entre atributos y tipo de error.

Se apunta lo siguiente: (i) los errores de plural predominan; (ii) en concordancias con ambos términos con *-e-* epentética [$ES = 2$] predominan los errores mixtos; (iii) en la distancia mínima y la que se da entre enunciados de LDA solamente existieron errores de plural [$DIS = 1,3$]; (iv) por otro lado, en las LDA con subordinada relativa solo se registraron errores de plural y mixtos [$POS = 2$]; (v) las estrategias EST3, EST5 y EST6 no se aplican a casos con errores de *-e-* epentética [$EST3 = 1$, $EST5 = 1$, $EST6=1$]; en EST3 tampoco se registraron errores de género.

5.3. Análisis de Correspondencias Múltiple [MCA].

El análisis de correspondencias (CA) es una técnica para analizar dependencias en las variables categóricas de tablas de contingencia. A partir de una tabla de contingencia de frecuencias relativas, resume la información contenida usando un espacio de menor dimensión, a través de la distancia χ^2 . Constituye el equivalente de PCA [*Principal Component Analysis*] para variables cualitativas. Una medida importante es la *inercia total*, equivalente a la suma de todas las distancias de las filas a su valor medio ponderadas por su importancia. Geométricamente, mide cuán lejos están los perfiles fila (columna) de su perfil medio, el cual simboliza la hipótesis de homogeneidad (igualdad) de los perfiles. Si la inercia es alta, entonces hay mucha asociación entre filas y columnas de la tabla (caso extremo: $I_T = J - 1$); por el contrario, si es baja, hay poca asociación porque los perfiles están cerca de la media (caso extremo: $I_T = 0$) [Greenacre, 2008]. También se entiende como una medida similar a la varianza total en PCA. Se define, para el caso de dos variables, como:

$$I_T = \phi^2 = \frac{\chi^2}{N} = \sum_{i=1}^I f_i \cdot (r_i - c)^T D_c^{-1} (r_i - c)$$

donde: (a) r_i es el vector de la distribuciones condicionales de las filas; (b) c es el valor medio de las filas, que viene dado por el vector cuyos componentes son las frecuencias relativas de las columnas; (c) D_c^{-1} es la matriz diagonal, cuyos componentes son las raíces cuadradas de las inversas de las frecuencias relativas de las columnas: $f_j^{-\frac{1}{2}}$; (d) f_i es la frecuencia relativa de las filas.

Se llevó a cabo un MCA con los siguientes atributos cualitativos: ID, MOD, POS, DIS, ANIM, ES, GRAMS, STEM.f, MORF.f, IMA.CONC.f, FAM.LEX.f, Fabs.SC.f, EST1, EST2, EST3, EST4, EST5, EST6, EST7. El Cuadro que sigue muestra los auto-valores para cada dimensión (inercias); el porcentaje de varianza explicada por cada dimensión (ratio entre inercia de cada dimensión y la total) y la inercia total (suma de las inercias de cada dimensión). Se reportan las primeras 10 sobre 25 componentes totales. Se observa que las primeras tres componentes representan el 47.41% de la variabilidad. Se seleccionan éstas. La inercia total (variabilidad total explicada) es $I_T \approx 0,11$; nótese que es baja, pero no cercana a cero, con lo cual se verifica una asociación *débil* entre los atributos.

	eigenvalue	variance.percent	cumulative.variance.percent
Dim.1	0.0256	22.7153	22.7153
Dim.2	0.0163	14.4542	37.1695
Dim.3	0.0115	10.2417	47.4112
Dim.4	0.0096	8.4916	55.9028
Dim.5	0.0068	6.0256	61.9284
Dim.6	0.0055	4.8578	66.7862
Dim.7	0.0047	4.1761	70.9623
Dim.8	0.0041	3.6654	74.6277
Dim.9	0.0039	3.5023	78.1300
Dim.10	0.0036	3.2334	81.3634

Cuadro 5.10: MCA: Inercia por dimensión, Varianza explicada y Varianza acumulada.

Para la correcta interpretación de las gráficas que siguen es pertinente prestar atención a lo siguiente respecto de los pares «variable:categoría» (llamadas «modalidades»): (i) cuánto más lejos del centro de

masa (correspondiente a al punto (0,0) de cruce entre las dimensiones), más inercia; o sea, mayor asociación entre modalidades; (ii) La gradación de colores indica la contribución absoluta de inercia (más claros, menos contribución) a cada dimensión; (iii) cuánto más cerca se encuentren dos modalidades, más co-ocurren en los datos.

- Dimensiones 1 y 2: (a) las modalidades que más contribuyen son: EST5:1; MORF.f:2, ES:1, DIS:1, POS:1, MOD:3; (b) las distancias [DIS:1] cortas entre concordancias co-ocurren con falta de subordinación relativa [POS:1] (44 casos); (c) MORF.f:2 (distancias medias de las terminaciones) co-ocurren con: los dos términos de las concordancias con *-e-* epentética [ES:2] (404 casos); EST5 (210 casos); EST7 (78 casos).
- Dimensiones 1 y 3: (a) las modalidades que más contribuyen son: EST5:1; MORF.f:2, ES:1, Fabs.SC.f:1; GRAMS:1; (b) las concordancias con adjetivos [MOD:3] tienden a pertenecer a TYPE de baja frecuencia [Fabs.SC:0] (441 casos); (c) en el TYPE de alta frecuencia [Fabs.SC.f:1] tiende a aplicarse la estrategia 6 [EST6:1] (30 casos);
- Dimensiones 2 y 3: (a) las modalidades que más contribuyen son: POS:1, DIS:1, MORF.f:2, ES:1; (b) Las concordancias con *-e-* epentética en ambos términos [ES:2] tienden a tener adjetivos [MOD:3] (28 casos); (c) en las concordancias con controlador animado [ANIM:1] se tiende a aplicar la estrategia 6 [EST6:1] (42 casos).

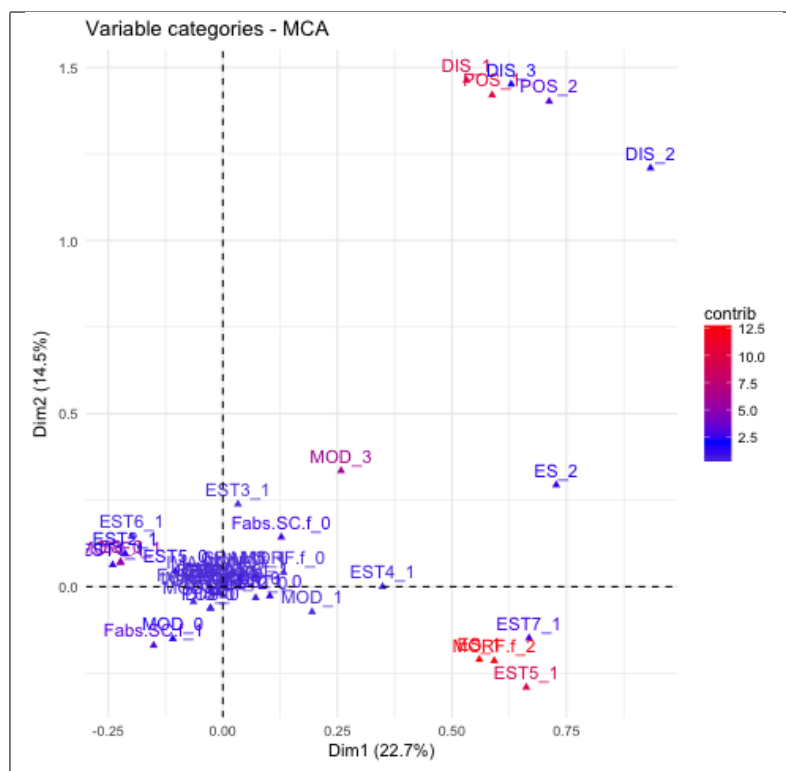


Figura 5.10: Gráfica de Correspondencias: dimensión 1 y 2.

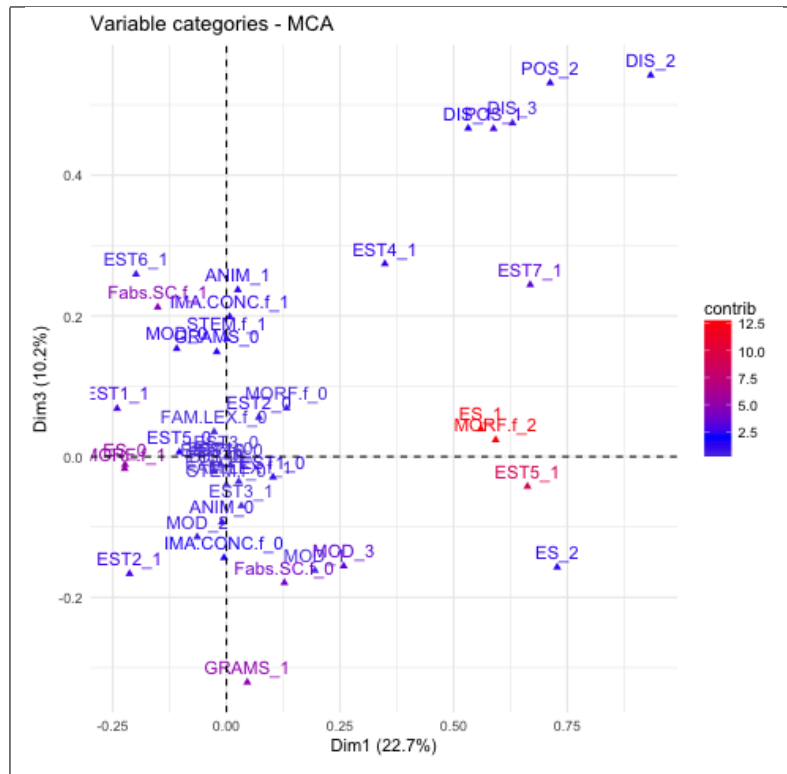


Figura 5.11: Gráfica de Correspondencias: dimensión 1 y 3.

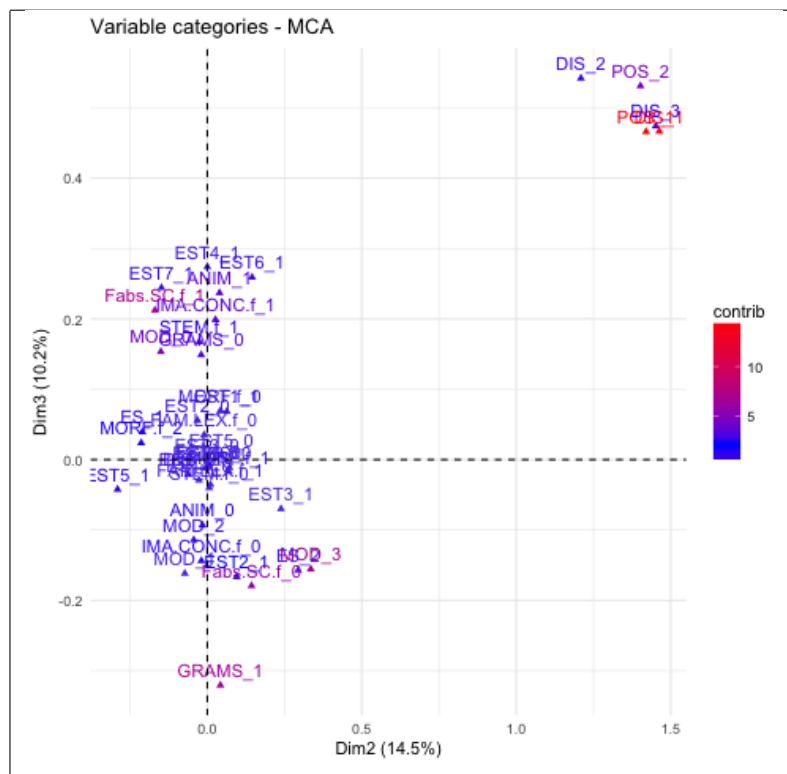


Figura 5.12: Gráfica de Correspondencias: dimensión 2 y 3.

5.4. Selección de variables.

En un modelo de regresión con gran cantidad de predictoras, se hace necesario la elección de los coeficientes que sean más informativos (que produzcan los efectos mayores en la respuesta media) con el objetivo de reducir la dimensionalidad del problema. Para ello, es preciso encontrar el mejor subconjunto

de tamaño k , $k \in \{1, 2, \dots, p\}$, de regresoras que *minimice* la función de pérdida (por ejemplo, la suma de cuadrados de los residuos). Sucede que la curva de la función de pérdida contra k decrece necesariamente con el aumento de las regresoras; por lo cual no puede usarse para elegir el mejor subconjunto de tamaño k . Además, explorar el espacio entero de modelos se hace imposible². Otra cuestión es que no solamente interesan los efectos de las variables independientemente de las demás sino también la interacción entre ellas. Entonces, aun teniendo una cantidad módica de predictoras, las interacciones dobles, triples, etc. entre todas ellas también hacen intratable el problema. El objetivo de este apartado es detectar los atributos más influyentes y establecer si existe diferencia entre las versiones continuas y discretas de los atributos. Se usan tres estrategias: (i) selección univariada (cada atributo por separado) usando medidas de asociación; (2) selección recursiva dentro de un modelo usando el resto de los atributos; (3) LASSO generalizado.

5.4.1. Selección univariada.

Se aplicaron medidas de asociación univariada entre cada atributo aisladamente y la variable respuesta binaria y categórica. Se describen a continuación las medidas utilizadas.

(1) Basadas en entropías: miden la cantidad de información compartida entre dos variables (en el rango $[0,1]$) y deben ser maximizadas. Dadas las definiciones de:

(i) Entropía (incertidumbre presente en una variable): $H(X) = \sum_{x \in X} p(x) \log [p(x)]$;

(ii) Entropía condicional (incertidumbre en una variable sabiendo el valor de otra) :

$$H(X | Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log [p(x | y)]$$

(iii) Información mutua (reducción en la incertidumbre de una variable si se sabe el valor de otra):

$$I(X, Y) = H(X) - H(X | Y)$$

se calcularon las siguientes medidas de información:

- *Normalized Mutual information* (NMI): $\frac{I(\text{Respuesta}, \text{Atributo})}{\sqrt{H(\text{Atributo})H(\text{Respuesta})}}$.
- *InformationGain* (IG): $H(\text{respuesta}) + H(\text{Atributo}) - I(\text{Respuesta}, \text{Atributo})$.
- *Gain Ratio* (GR): $\frac{H(\text{Respuesta}) + H(\text{Atributo}) - I(\text{Respuesta}, \text{Atributo})}{H(\text{Atributo})}$.
- *Symmetrical Uncertainty* (SU): $2 \frac{H(\text{respuesta}) + H(\text{Atributo}) - I(\text{Respuesta}, \text{Atributo})}{H(\text{Atributo}) + H(\text{Respuesta})}$

Area bajo la curva ROC (AUC): Gráfica en la que el eje «y» contiene la proporción de verdaderos positivos y el eje «x», la proporción de falsos positivos. Si el predictor separara perfectamente las clases habría un corte de probabilidad de asignación de la clase cuyo valor de verdaderos positivos sería 1 y de falsos positivos sería 0; por lo tanto, el área total bajo la curva sería $AUC = 1$ también. Por otro lado, si un predictor no separa bien las clases, el área se acercaría a $AUC = 0,50$. Se debe maximizar. En el caso de respuesta categórica, se calcula la AUC de la categoría en cuestión respecto de *todas* las demás.

Relief Score (Kira & Rendell, 1992; ver también: Kuhn & Johnson, 2013, cap. 18) : el algoritmo selecciona un punto al azar y luego busca los puntos más cercanos de ambas clases (denominados «error» [*miss*] y «acierto» [*hit*]). Luego se calcula una medida de *score* de la diferencia entre el punto elegido y los errores y aciertos. Esto se repite con m puntos y el *score* general es una suma de estas diferencias. La intuición radica en que un predictor que separa

²Sin contar interacciones, hay 2^p modelos; si hubiera 20 regresoras: $2^{20} = 1048576$.

bien las clases tendrá los «aciertos» cerca y los «errores» lejanos; por eso, *scores generales* altos indicarán predictores influyentes.

DistAngle: Coseno de la distancia angular entre particiones por la clase.

Gini index: Considera una partición $D = \{D_1, \dots, D_d\}$ binaria ($d = 2$) de cada atributo y mide la «impureza» de dicha partición como: $Gini(D) = 1 - \sum_{i=1}^m P_i^2$; donde $i = 1, \dots, m$, con m clases en la respuesta y $P_i = \frac{|C_i, D|}{|D|}$; la cantidad de veces que la clase C aparece en la partición D respecto del total de casos en la partición D . Suponiendo la clase A de la respuesta, el índice reza: $Gini(D) = \frac{|D_1|}{|D|} Gini(D_1) + \frac{|D_2|}{|D|} Gini(D_2)$. Se debe minimizar (Rokach & Maimon, 2008).

Maximal Information Coefficient [MIC]: Medida de asociación lineal y no lineal entre dos variables continuas propuesta en Reshef et. Al. (2011). El rango varía entre $[0, 1]$, cuyos extremos indican asociación mínima o máxima. Dado un par de variables (x, y) se buscan múltiples configuraciones (con particiones de los datos) que contienen grillas con resoluciones crecientes. Para cada configuración de grilla (Ej.: $2 \times 2, 2 \times 3, \dots, x \times y$) se almacena el valor de mayor Información Mutua ($I(x, y)$) y se lo normaliza como: $m = \frac{\max\{I_G(x, y)\}}{\log(\min\{x, y\})}$; donde el máximo se toma sobre todos los valores de las grillas de la configuración G . Luego se crea una matriz $x \times y$ con los valores m : $M = (m_{x, y})$. *MIC* es el valor máximo sobre los pares ordenados x, y de la matriz M tal que $xy < B$; donde $B = n^{0,6}$ (n es el tamaño muestral). Dicho valor también se corresponde al punto más alto en la matriz M representada como superficie de respuesta.

Odds Ratio [OR]: Se calcula para dos variables categóricas de dos niveles cada una. Si se tienen las variables $A \in \{0, 1\}$ y $B \in \{0, 1\}$; y demás: $a = |A = 0 \cap B = 0|$; $b = |A = 0 \cap B = 1|$; $c = |A = 1 \cap B = 0|$; $d = |A = 1 \cap B = 1|$; el *Odds Ratio* reza: $\frac{ab}{cd} = \frac{ad}{bc}$. El intervalo de confianza de nivel $1 - \alpha$ (aquí $\alpha = 0,05$) asintótico para $\ln(OR)$ es: $IC_{0,95}(\ln(OR)) = \exp\left\{\ln(OR) \pm z_{\frac{\alpha}{2}} \times ES(\ln(OR))\right\}$; donde: $ES(\ln(OR)) = \sqrt{\frac{1}{a} + \frac{1}{b} + \frac{1}{c} + \frac{1}{d}}$ y $z_{\frac{\alpha}{2}}$ es el percentil de la distribución normal estándar $N(0, 1)$ que deja a su derecha un área de $\frac{\alpha}{2}$. Bajo $H_0 : OR = 1$; se rechaza si $IC(\ln(OR)) \notin 1$, o sea si el intervalo de confianza no contiene al «1».

Pesos de Random Forest [RF]: Se calculan los pesos para cada atributo utilizando un modelo de *Random Forest* (ver apartado 5.4.2).

Los Cuadros 11 y 12 que siguen muestran los resultados para la respuesta binaria y la categórica (los Cuadros con los valores se presentan en el apéndice del capítulo). Se declaró a una variable como importante de acuerdo al corte $AUC > 0,6$; $GINI < 0,001$; $RF > 20$ y para el resto $> 0,15$ (las medidas vacías no cumplen con dichas condiciones).

	Variables
CO	ADJ MOD Fabs.C.S Fabs.SC.f
AUC	Fabs.C.S
GINI	POS LDA ES GRAMS ANIM STEM.f IMA.CONC.f EST2 EST4 EST5 EST7
RF	MOD Fabs.C.S STEM MORF MORF.f CUMRES IMA.CONC FAM.LEX GRUPO6 EST1
OR	LDA GRAMS ANIM STEM.f IMA.CONC.f EST2 EST4 EST7
SU	
RS	
GR	
IG	
NMI	
MIC	

Cuadro 5.11: Atributos seleccionados. Respuesta binaria. NMI: Normalized Mutual Information; IG: Information Gain; GR: Gain Ratio; SU: Symmetrical Uncertainty; AUC: Area Under the Curve; RS: Relief Score; GI: Gini; CO: Cosine; MIC: Maximal Information Coefficient; OR: Odds Ratio; RF: Pesos de los atributos según Random Forest.

	VARIABLES
CO	ADJ MOD ANIM Fabs.C.S Fabs.SC.f MORF MORF.f CUMRES.f CUMRES FAM.LEX FAM.LEX.f GRUPO6 EST2
AUC.1	Fabs.C.S EST1
AUC.2	ADJ MOD Fabs.C.S MORF CUMRES EST1
AUC.3	ES GRAMS ANIM Fabs.C.S Fabs.SC.f MORF CUMRES FAM.LEX FAM.LEX.f EST1 EST2
AUC.4	ANIM FAM.LEX FAM.LEX.f EST2
GINI	GRAMS ANIM STEM.f FAM.LEX IMA.CONC.f FAM.LEX.f EST2 EST4 EST5
RF	ADJ MOD Fabs.C.S Fabs.SC.f STEM MORF MORF.f CUMRES IMA.CONC FAM.LEX FAM.LEX.f GRUPO6 EST1
SU	
RS	Fabs.C.S STEM MORF CUMRES FAM.LEX GRUPO6
GR	
IG	
NMI	
MIC	

Cuadro 5.12: Atributos seleccionados. Respuesta categórica. Leyenda: NMI: Normalized Mutual Information; IG: Infor- mation Gain; GR: Gain Ratio; SU: Symmetrical Uncertainty; AUC.1: Area Under the Curve (1 vs 0); AUC.2: Area Under the Curve (2 vs 0); AUC.3: Area Under the Curve (3 vs 0); AUC.4: Area Under the Curve (4 vs 0); RS: Relief Score; GI: Gini; CO: Cosine; MIC: Maximal Information Coefficient; RF: Pesos de los atributos según Random Forest.

5.4.2. Selección basada en eliminación recursiva de atributos con modelos de minería de datos [RFE].

Se trata de un método *backward* [«hacia atrás»] de selección de variables. El modelo inicial contiene los P predictores totales. Estos se remueven de forma interactiva según aquellos que no contribuyen de manera significativa al modelo. Guyon et al. (2002) propusieron un algoritmo de este tipo, que fue mejorado encapsulando el mecanismo en el remuestreo con validación cruzada (Kuhn y Johnson, 2013, cap. 19). El algoritmo se esquematiza a continuación. Primero se dividen los datos en un conjunto de entrenamiento (80%, por ejemplo) y otro de validación (20% restante). Luego el algoritmo selecciona el modelo con la cantidad óptima de variables usando *solamente* el conjunto de entrenamiento. A continuación se aplica validación cruzada:

1. Se parte el conjunto de datos T en k conjuntos disjuntos de igual tamaño. Supóngase $T = T_1, T_2, \dots, T_k$; eligiendo aquí $k = 10$.
2. Para cada $k = 1, 2, \dots, K$; se ajusta un modelos con p regresoras excluyendo el conjunto T_k -ésimo: $\hat{f}_{-k}(p)$, se calcula la importancia de las variables.
3. Se calculan los predichos para el T_k -ésimo (validación), con el modelo ajustado con los datos en 2, o sea con un conjunto de *training* que excluía al T_k -ésimo.
4. Se calcula una medida de desempeño de validación cruzada (CVM) para cada T_k -ésimo de *testing*: $(CVM) = AUC$.
5. Se repiten los pasos 2-4 para $p = 1, 2, \dots, P$.
6. Se calcula una medida *global* de validación cruzada $(CVM)_p = K^{-1} \sum_{k=1}^k (CVM)_k$, o sea una media sobre $k - folds$; para cada p .
7. Se elije el modelo con CVM máximo (porque AUC debe ser lo más alta posible). De la cantidad p de variables del modelo seleccionado, se eligen las primeras cinco en importancia.

Se ajustaron modelos con árboles de decisión, árboles aleatorios (*Random Forest*) y Naïve Bayes.

Naïve Bayes es un método de aprendizaje automático probabilístico, basado en el teorema de *Bayes*. Supone independencia entre las predictoras.

Un árbol de decisión clasifica un ejemplar (caso de concordancia) filtrándolo de manera descendente, hasta encontrar una hoja que coincide con la clasificación buscada. Genera una regla para clasificar nuevas

instancias. El algoritmo explora el espacio de todos los árboles de decisión posibles; empezando por el árbol vacío y yendo hacia árboles cada vez más complejos. Se construye el árbol de «arriba hacia abajo» (de la raíz a las hojas). Selecciona el mejor atributo y lo pone en la raíz, usando una medida de ganancia de información (el algoritmo usado aquí utiliza el «índice de información»³). Para cada valor de dicho atributo se crea una rama y un nodo de acuerdo a sus valores. Los ejemplos de entrenamiento se reparten en los nodos. Luego se aplica otra vez dicha medida para colocar el mejor atributo en cada nuevo nodo, usando los ejemplares de entrenamiento ya distribuidos. Los atributos que se evalúan no se vuelven a usar para sucesivas evaluaciones de atributos restantes (no existe *back – tracking*). Los atributos que se colocan cerca de la raíz son los más influyentes. El algoritmo encuentra el primer árbol consistente con el conjunto de entrenamiento, en una búsqueda de tipo *hill – climbing* guiada por esta medida de ganancia de información. Idealmente cada rama del árbol crecería en profundidad hasta que se logren clasificar *perfectamente* los casos del conjunto de entrenamiento. Sin embargo, ésto llevaría al sobre ajuste; por lo cual se debe establecer un nivel del árbol L de «poda»; aquí se estableció $L = 12$ (Mitchel, 1997; Therneau & Atkinson, 2015).

Random Forest es un tipo de ensamble de árboles de decisión. Combina las clasificaciones hechas por múltiples (miles de) árboles de decisión. El siguiente esquema ilustra el algoritmo. Para cada clasificador M_t ; $t = 1, \dots, T$ se muestrean μ instancias del conjunto de entrenamiento S sin reemplazo. Luego se hace crecer un árbol de decisión sin podarlo. En *cada* nodo, se seleccionan aleatoriamente N atributos del total y se elige la mejor partición teniendo en cuenta estos N atributos. La clasificación de una instancia desconocida se hace por voto mayoritario de las decisiones de cada árbol (Rokach & Maimon, 2008).

Sean: DT : método de árbol de decisión; T : número de interacciones; S : conjunto de entrenamiento; μ : subconjunto muestreado; N : número de atributos usados en cada nodo.

Para cada clasificador M_t ; $t = 1, \dots, T$.

1. $t \leftarrow 1$

2. **repetir:**

3. $S_t \leftarrow$ Muestrear μ instancias de de S sin reemplazo.

4. Construir un clasificador M_t mediante $DT(N)$ usando S_t .

5. $t = t + 1$

6. **hasta** $t > T$.

Ahora bien, recuérdese que los casos de error son aproximadamente un 25%. Es decir que ambas clases se encuentran desbalanceadas. El desbalance es más severo todavía en el caso de la respuesta categórica. Estos métodos son sensibles al desbalance de clases, lo cual afecta las medidas (globales) de predicción. Se decidió recurrir a un método de remuestreo, la llamada *synthetic minority over-sampling technique* (SMOTE, por sus siglas en inglés), propuesta por Chawla et al. (2002) [ver también: Kuhn y Johnson, 2013, cap. 16]. Para balancear, el método reduce la(s) clase(s) que contienen más cantidad de instancias y aumenta aquella(s) con menos instancias. Para la clase minoritaria se crean «casos sintéticos». Para ello, una instancia es seleccionada de la clase minoritaria y se buscan los K vecinos más cercanos (aquí $K = 5$). El nuevo «caso sintético» se forma mediante una combinación aleatoria de los atributos del caso seleccionado originalmente y de sus K vecinos más cercanos. Por otra parte, también es posible reducir la(s) clase(s) mayoritaria(s) mediante remuestreo para ayudar al balanceo.

El Cuadro siguiente muestra las primeras cinco variables más influyentes seleccionadas por cada método, según sea la respuesta binaria o categórica. Se detalla también el valor global de área bajo la curva (AUC) obtenido y la cantidad de atributos p del modelo seleccionado. Por otro lado, la Figura detalla, para cada tipo de clasificador, la medida global AUC para cada conjunto de variables considerado en la validación cruzada.

³Supóngase que f es una función de la impureza del nodo A , tal que: $I(A) = \sum_{i=1}^C f(p_{i,A})$; se define el «índice de información»: $f(p) = -p \log(p)$. Se parte el nodo A según la máxima reducción entre nodo padre A e hijos $A_{derecha}$ y $A_{izquierda}$; tal que: $\Delta I = p(A)I(A) - p(A_{izq})I(A_{izq}) - p(A_{dcha})I(A_{dcha})$.

Respuesta		Random Forest	Árbol de decisión	Naïve Bayes
Binaria	Variables	Fabs.C.S, FAM.LEX, IMA.CONC STEM, CUMRES	Fabs.C.S, GRUPO6, LDA IMA.CONC, POS	Fabs.C.S, GRUPO6, POS DIS, LDA
	AUC global	0.9473	0.8521	0.7982
	p atributos	27	21	25
Categorica	Variables	Fabs.C.S, IMA.CONC, FAM.LEX STEM, CUMRES	Fabs.C.S, IMA.CONC, MORF STEM, FAM.LEX	EST1, EST2, Fabs.C.S FAM.LEX, Fabs.SC.f
	AUC global	0.9294	0.7956	0.7280
	p atributos	9	7	28

Cuadro 5.13: *Recursive Feature Elimination*. Resultados de clasificadores basados en árboles de decisión, *Random Forest* y Bayes ingenuo.

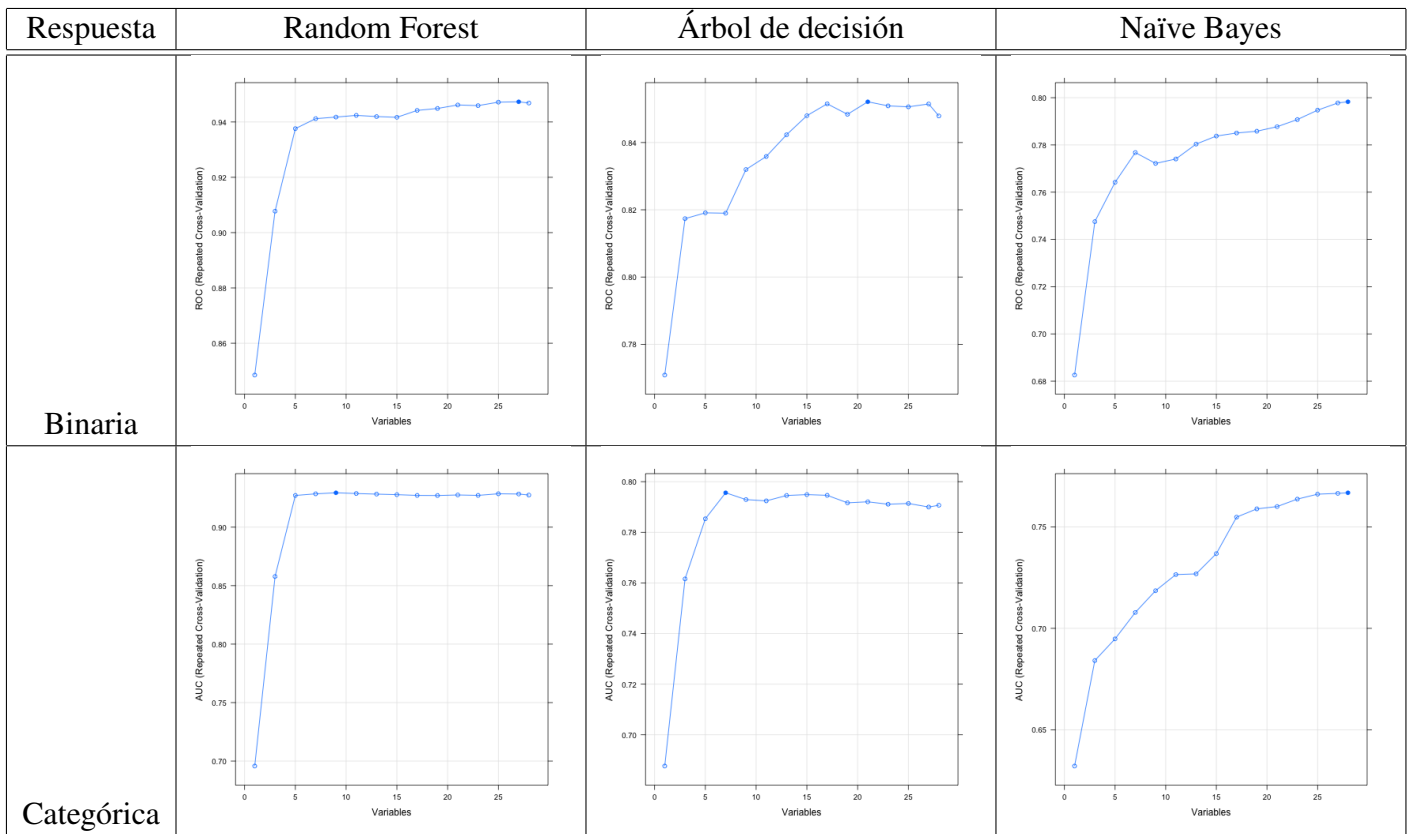


Figura 5.13: *AUC* global para cada grupo de atributos, según el método y el tipo de respuesta.

5.4.3. LASSO.

La regresión LASSO («Least Absolute Selection and Shrinkage Operator») Fue propuesta por R. Tibshirani (1996; ver también: Hastie, Tibshirani & Friedman, 2017) y, al igual que la regresión *Ridge*, impone una penalización en el tamaño de los coeficientes para seleccionar variables. Dada una regresión lineal con x_{ij} predictoras e y_i valores de la variable respuesta ($i = 1, \dots, N$ y $j = 1, \dots, p$), *Lasso* resuelve el siguiente problema de optimización minimizando:

$$\hat{\beta}^{Lasso} = \left\{ \min_{\beta} \frac{1}{N} \sum_{i=1}^N w_i \mathcal{L}(y_i, \eta) + \lambda \sum_{j=1}^p |\beta_j| \right\}$$

donde $\beta = (\beta_0, \beta_1, \dots, \beta_p)$ es el vector de parámetros y $\mathcal{L}(y_i, \eta)$ el logaritmo negativo de la verosimilitud para la observación i -ésima; $\eta = \beta_0 - \sum_{j=1}^p x_{ij} \beta_j$ es el predictor lineal; y w_i son posibles pesos para las

observaciones. En el caso de la familia gaussiana, $\mathcal{L}(y_i, \eta) = \frac{1}{2} (y_i - \eta)^2$. Por lo tanto, la fórmula anterior reza:

$$\hat{\beta}^{Lasso} = \left\{ \min_{\beta} \frac{1}{2N} \sum_{i=1}^N (y_i - \beta_0 - \sum_{j=1}^p x_{ij} \beta_j)^2 + \lambda \sum_{j=1}^p |\beta_j| \right\}$$

El parámetro positivo $\lambda \geq 0$ controla el grado de penalización impuesto: cuanto más alto, más penalización y, entonces, los coeficientes caen hacia cero; cuanto más bajo, los coeficientes se acercan a una estimación por OLS. Es decir que $\hat{\beta}^{Lasso} = \hat{\beta}^{OLS}$ si $\lambda = 0$; y $\hat{\beta}^{Lasso} = 0$ si $\lambda = \infty$. Ya que la respuesta binaria se distribuye Bernoulli y la categórica sigue una distribución multinomial, se usarán las familias binomial y multinomial. Se usó el paquete *glmnet* de R (Friedman, Hastie & Tibshirani, 2010). Las clases se balancearon como se explicó en el apartado anterior (SMOTE). Lambda es elegido por el investigador o por medio de validación cruzada de la siguiente manera:

1. Se parte el conjunto de datos T en k conjuntos disjuntos de igual tamaño. Supóngase $T = T_1, T_2, \dots, T_k$; eligiendo $k = 5$ o $k = 10$.
2. Para cada $k = 1, 2, \dots, K$; se ajusta un modelo con z regresoras con un determinado λ (o t) excluyendo el conjunto T_k -iésimo: $\hat{f}_{-k}^{\lambda}(z)$.
3. Se calculan los predichos para el T_k iésimo (testing), con el modelo ajustado con los datos en 2, o sea con un conjunto de *training* que excluía al T_k iésimo.
4. Se calcula el error de validación cruzada (CV error) para cada T_k -iésimo de *testing*: $(CVerro)r_k^{\lambda}$; aquí se usa *AUC* y *deviance* = $\mathcal{L}(\hat{f}_k^{\lambda}(z))$, se sea: $-2 \times \log(\text{verosimilitud})$ usando T_k .
5. Dicho modelo con determinado λ , tendrá un error de validación cruzada total, que es el promedio de los errores de cada T_k iésimo:

$$(CVerro)r^{(\lambda)} = K^{-1} \sum_{k=1}^k (CVerro)r_k^{(\lambda)}$$

6. De todos los λ probados para el modelo, se selecciona aquel λ^* con menor error de validación cruzada total: $(CVerro)r^{(\lambda^*)}$.

En la Figura que sigue se grafica el resultado de la validación cruzada con $k = 20$ folds para elegir el valor óptimo de λ . El eje «y» es la medida (error) de validación usada y el «x», los diferentes valores de $\log(\lambda)$. Como medidas de desempeño, se *maximizó* área bajo la curva (*AUC*) para la respuesta binaria; y se *minimizó* la *deviance* para el caso multinomial.

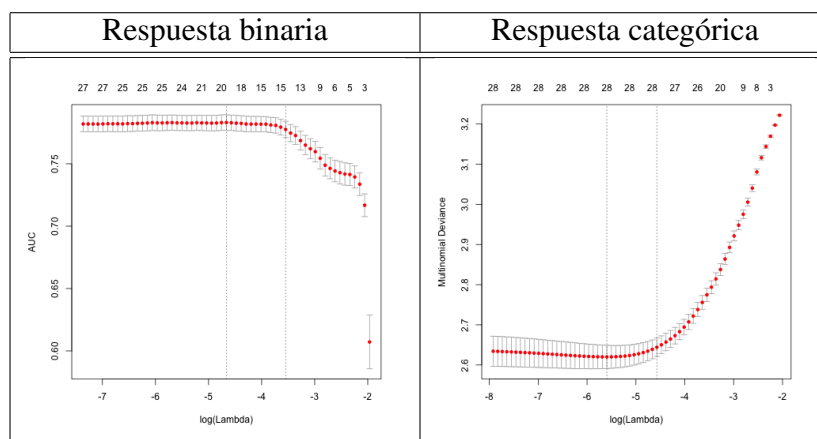


Figura 5.14: Lasso generalizado: elección de λ por validación cruzada ($k = 20$), para la respuesta binaria y la categórica.

El Cuadro que sigue muestra las cinco variables más importantes⁴ para ambos tipos de respuesta; el valor de $\log(\lambda)$ obtenido por la validación cruzada y $\log(\lambda_{1.sd})$, el valor de λ que se encuentra a un desvío típico de λ . Se utilizó este último para extraer los coeficientes del modelo. En el caso de la respuesta categórica, el modelo multinomial arroja estimaciones de los coeficientes por categoría de respuesta (es decir que ajusta en este caso cinco modelos). Se tomaron las variables influyentes que se repetían al menos dos veces.

	$\log(\lambda)$	$\log(\lambda_{1.sd})$	atributos influyentes
Respuesta binaria	-4.663461	-3.547056	DIS, Fabs.C.S, LDA, GRUPO6, EST7
Respuesta categórica	-5.595087	-4.571716	ES, EST2, EST5, FAM.LEX, FAM.LEX.f, Fabs.SC.f, IMA.CONC, IMA.CONC.f

Cuadro 5.14: Lasso generalizado: las variables más influyentes, según respuesta binaria o categórica

A modo de resumen, los Cuadros 5.10 y 5.11 muestran cuántas veces fue elegido un atributo según todos los métodos de selección de variables considerados. Para RES_BIN no se secciona nunca CUMRES.f, EST3, EST6. Por otro lado, para RES_CAT no aparecen POS, DIS, LDA, EST3, EST6.

Se observa que las versiones discretas de las variables aparecen menos veces que las versiones continuas cuando ambos tipos de variables resultan seleccionadas⁵. Para la respuesta binaria los dos atributos más seleccionados resultaron LDA y Fabs.C.S; en tanto que para la categórica fueron FAM.LEX y Fabs.C.S. Entonces, la influencia de la frecuencia de TYPE de concordancia es muy importante.

5.5. Resumen.

El análisis hecho en el presente capítulo ha evidenciado lo siguiente:

- Se identificaron seis grupos para aglomerar los casos de concordancia.
- Los errores de concordancia son un cuarto del total, y de éstos, más de la mitad son de plural. En lo que respecta a la precisión, se observó el orden: JAKO > SONIA > MIRKA > NATI. Los errores de género y por uso de terminación «(-e)s» disminuyen en los alumnos de los niveles más altos (B2 y C1).
- Se vieron diferencias en la proporción de errores para las variables: ADJ, MOD, Fabs.SC.f, MORF.f, CUMRES.f y las estrategias: EST1, EST3, EST6.
- En el MCA, se observó que las variables que más co-ocurren son: DIS:1 con POS:1; MORF.f:2 con ES:2 / EST5 / EST7; Fabs.SC.f:0 con MOD:3; Fabs.SC.f:1 con EST6; ES:2 con MOD:3 y ANIM:1 con EST6.
- En la selección de atributos, EST3 y EST6 nunca se eligieron para ambas respuestas. Se favoreció a las versiones continuas de los atributos. La frecuencia de TYPE resultó muy influyente.

⁴Siguiendo a Menard (2011), se multiplicó la estimación de cada coeficiente $\hat{\beta}$ por el desvío típico del atributo x correspondiente: $\hat{\beta}_{standar}^{Lasso} = \hat{\beta}^{Lasso} \times \hat{\sigma}_x$. Luego se tomó el valor absoluto. Dichos valores absolutos se ordenaron de modo decreciente y se seleccionaron los primeros cinco como los más importantes.

⁵Lo cual parece a priori esperable ya que discretizar implica perder información.

	Conteo	Atributos
1	1	ADJ ES EST1 EST5 Fabs.SC.f MORF.f
2	2	ANIM CUMRES DIS EST2 EST4 GRAMS IMA.CONC.f MOD MORF POS STEM.f
3	3	EST7 FAM.LEX GRUPO6 IMA.CONC STEM
4	4	LDA
5	7	Fabs.C.S

Cuadro 5.15: Conteo de atributos para respuesta binaria, seleccionados por todos los métodos.

	Conteo	Atributos
1	1	CUMRES.f EST7 POS STEM.f
2	2	EST4 GRAMS MORF.f
3	3	ADJ ES GRUPO6 IMA.CONC.f MOD
4	4	STEM
5	5	ANIM EST5 IMA.CONC
6	6	CUMRES EST1 MORF
7	7	FAM.LEX.f Fabs.SC.f
8	8	EST2
9	9	Fabs.C.S
10	11	FAM.LEX

Cuadro 5.16: Conteo de atributos para respuesta categórica, seleccionados por todos los métodos.

5.6. Apéndice de capítulo.

5.6.1. Selección univariada de variables.

	NMI	IG	GR	SU	AUC	RS	GI	CO	RF	MIC	OR
ADJ	0.0118	0.0074	0.0143	0.0118	0.5694	-0.026	0.0056	0.1918	18.2817		1.76 [1.42,2.2]
MOD	0.0107	0.0091	0.0127	0.0107	0.5744	-0.014	0.0068	0.2128	24.1174		
POS	0.0027	0.001	0.019	0.0027	0.51	0.014	9e-04	0.0147	6.6575		
DIS	0.0035	0.0014	0.034	0.0035	0.5102	0	0.0012	0.0104	4.3797		
LDA	0.0024	9e-04	0.0178	0.0024	0.5101	0.016	7e-04	0.0157	4.9324		1.58 [0.94,2.61]
ES	0.0016	0.001	0.0107	0.0016	0.5096	-0.028	7e-04	0.035	17.7997		
GRAMS	7e-04	4e-04	0.0048	7e-04	0.5156	0.026	3e-04	0.0392	15.6313		1.15 [0.92,1.45]
ANIM	0.0012	7e-04	0.0082	0.0012	0.5192	0.004	5e-04	0.0462	15.694		1.20 [0.95,1.52]
Fabs.C.S	0.0481	0.0097	0.0186	0.0159	0.6001	0.0355	0.0075	0.2096	42.9239	0.1072	
Fabs.SC.f	0.0127	0.008	0.0169	0.0127	0.5716	0.04	0.006	0.196	18.3766		0.55 [0.44,0.69]
STEM	0.0144	0	0	0	0.5321	0.1081	0.003	0.1155	28.194	0.0598	
MORF	0.0084	0	0	0	0.5575	0.0261	0.0029	0.1022	25.7781	0.0204	
MORF.f	0.0101	0.007	0.013	0.0101	0.5286	0.046	0.0058	0.0932	26.3736		
STEM.f	3e-04	2e-04	0.0027	3e-04	0.5082	0.016	1e-04	0.0171	12.37		1.11 [0.85,1.45]
CUMRES.f	0.0049	0.004	0.0072	0.0049	0.5482	0.03	0.003	0.1307	15.1743		
CUMRES	0.0084	0	0	0	0.5575	0.0261	0.0029	0.1022	27.2094	0.0204	
IMA.CONC	0.0649	0	0	0	0.5301	0.0758	0.0018	0.0897	29.1985	0.1112	
FAM.LEX	0.0649	0	0	0	0.5051	0.0399	0.0012	0.0905	29.6611	0.1103	
IMA.CONC.f	2e-04	1e-04	0.0022	2e-04	0.5093	0.018	1e-04	0.0258	14.3657		1.08 [0.87,1.34]
FAM.LEX.f	0.0024	0.0015	0.0044	0.0024	0.5317	0.056	0.0012	0.0895	19.1252		0.78 [0.63,0.96]
GRUPO6	0.0033	0.0034	0.0099	0.0033	0.5396	0.03	0.0026	0.115	22.922		
EST1	0.0055	0.0032	0.0144	0.0055	0.5417	0.09	0.0024	0.0983	21.3624		0.66 [0.51,0.84]
EST2	0.0012	7e-04	0.0095	0.0012	0.5185	0.07	5e-04	0.0426	14.2762		1.21 [0.95,1.54]
EST3	0.0059	0.0018	0.0953	0.0059	0.5046	-0.002	0.0016	0.0065	5.3708		7.37 [1.2,77.58]
EST4	9e-04	3e-04	0.0257	9e-04	0.5056	0.018	3e-04	0.0085	9.5414		1.35 [0.76,2.34]
EST5	0.0024	0.0011	0.0089	0.0024	0.5183	0.018	8e-04	0.0333	14.0909		0.72 [0.51,1]
EST6	0.0046	0.0017	0.0186	0.0046	0.5143	0.042	0.0014	0.0225	14.1713		1.83 [1.12,2.96]
EST7	0.0027	0.001	0.027	0.0027	0.5108	0.014	8e-04	0.0168	7.4939		1.61 [0.97,2.64]

Cuadro 5.17: Selección de Atributos: respuesta binaria. Leyenda: NMI: Normalized Mutual Information; IG: Infor- mation Gain; GR: Gain Ratio; SU: Symmetrical Uncertainty; AUC: Area Under the Curve; RS: Relief Score; GI: Gini; CO: Cosine; MIC: Maximal Information Coefficient; OR: Odds Ratio [Intervalo de confianza 95 por ciento]; RF: Pesos de los atributos según Random Forest.

	NMI	IG	GR	SU	AUC.1	AUC.2	AUC.3	AUC.4	RS	GI	CO	RF	MIC
ADJ	0.013	0.01	0.014	0.013	0.57	0.61	0.6	0.57	0.06	0.0042	0.2	21.56	
MOD	0.015	0.014	0.013	0.015	0.58	0.62	0.58	0.58	0.11	0.0055	0.27	25.18	
POS	0.007	0.004	0.019	0.007	0.52	0.51	0.51	0.52	0	0.0014	0.02	5.18	
DIS	0.014	0.007	0.034	0.014	0.52	0.51	0.51	0.52	0	0.0017	0.03	6.8	
LDA	0.006	0.003	0.018	0.006	0.52	0.51	0.51	0.52	0	0.0013	0.02	6.62	
ES	0.009	0.007	0.011	0.009	0.52	0.57	0.61	0.6	0.04	0.0011	0.14	19.22	
GRAMS	0.004	0.003	0.005	0.004	0.57	0.57	0.65	0.58	0.05	4e-04	0.15	18.9	
ANIM	0.007	0.005	0.008	0.007	0.53	0.53	0.64	0.68	0.05	5e-04	0.2	18.88	
Fabs.C.S	0.094	0.012	0.019	0.016	0.63	0.63	0.63	0.6	0.16	0.006	0.26	47.47	0.11
Fabs.SC.f	0.015	0.012	0.017	0.015	0.6	0.6	0.68	0.58	0.1	0.0048	0.24	21.62	
STEM	0.042	0	0	0	0.55	0.55	0.56	0.55	0.18	0.0026	0.11	32.2	0.06
MORF	0.018	0	0	0	0.56	0.61	0.61	0.56	0.17	0.0022	0.19	32.53	0.02
MORF.f	0.013	0.011	0.013	0.013	0.56	0.53	0.56	0.56	0.09	0.0053	0.16	33.45	
STEM.f	0.002	0.001	0.003	0.002	0.51	0.52	0.57	0.58	0.09	1e-04	0.08	12.95	
CUMRES.f	0.008	0.008	0.007	0.008	0.55	0.6	0.57	0.55	0.13	0.0023	0.21	16.17	
CUMRES	0.018	0	0	0	0.56	0.61	0.61	0.56	0.17	0.0022	0.19	33.09	0.02
IMA.CONC	0.115	0	0	0	0.59	0.59	0.59	0.6	0.14	0.0013	0.15	31.36	0.11
FAM.LEX	0.115	0	0	0	0.54	0.54	0.68	0.66	0.18	9e-04	0.23	33.34	0.11
IMA.CONC.f	0.002	0.001	0.002	0.002	0.57	0.57	0.58	0.5	0.09	2e-04	0.1	17.9	
FAM.LEX.f	0.004	0.003	0.004	0.004	0.53	0.54	0.63	0.62	0.1	8e-04	0.16	22.02	
GRUPO6	0.013	0.015	0.01	0.013	0.55	0.52	0.58	0.58	0.16	0.0037	0.27	28.85	
EST1	0.012	0.009	0.014	0.012	0.62	0.62	0.62	0.52	0.08	0.0017	0.13	21.8	
EST2	0.008	0.005	0.009	0.008	0.57	0.57	0.68	0.61	0.12	6e-04	0.16	16.34	
EST3	0.005	0.002	0.095	0.005	0.51	0.5	0.5	0.51	0.01	0.0015	0.01	6.49	
EST4	0.008	0.004	0.026	0.008	0.52	0.53	0.53	0.55	0.04	7e-04	0.04	12.03	
EST5	0.006	0.004	0.009	0.006	0.52	0.53	0.55	0.57	0.05	5e-04	0.05	15.15	
EST6	0.007	0.003	0.019	0.007	0.53	0.53	0.55	0.52	0.04	0.001	0.04	9.49	
EST7	0.009	0.005	0.027	0.009	0.52	0.55	0.52	0.53	0.03	0.0011	0.04	6.36	

Cuadro 5.18: Selección de Atributos: respuesta binaria. Leyenda: NMI: Normalized Mutual Information; IG: Infor- mation Gain; GR: Gain Ratio; SU: Symmetrical Uncertainty; AUC.1: Area Under the Curve (1 vs 0); AUC.2: Area Under the Curve (2 vs 0); AUC.3: Area Under the Curve (3 vs 0); AUC.4: Area Under the Curve (4 vs 0); RS: Relief Score; GI: Gini; CO: Cosine; MIC: Maximal Information Coefficient; RF: Pesos de los atributos según Random Forest.

Capítulo 6

Análisis de respuesta binaria.

En el presente capítulo se explora el modelo logístico mixto para la repuesta binaria. En primera instancia se presenta el modelo (generalizado) logístico, luego se introduce el modelo (generalizado) mixto y se presentan medidas de selección de modelos. Seguidamente, se presenta el análisis de los datos. Luego de llegar al modelo final, se analizan los errores de concordancia a la luz de los resultados.

6.1. El modelo logístico.

Las respuestas binarias del tipo $1 = \text{«ocurre el evento»}$, $0 = \text{«no ocurre el evento»}$ resultan muy comunes en investigación. La regresión logística (Hilbe, 2009; Hosmer & Lemeshow, 2013) describe la asociación entre dicha respuesta binaria y un conjunto de variables predictoras. Sea π_i la probabilidad de respuesta positiva: $Y_i = 1$ para el caso i -ésimo, con $i = 1, \dots, n$ casos. Por tanto, $P(Y_i = 0) = 1 - \pi_i$ resulta la probabilidad de respuesta negativa para el caso i -ésimo. Sea el vector de dimensión $(n \times (p + 1))$ de covariables para el caso i -ésimo: $\mathbf{x}_i = (1, x_{i1}, \dots, x_{ip})^T$; y el vector de coeficientes del modelo $\beta = (\beta_0, \beta_1, \dots, \beta_p)^T$, de dimensión $(p + 1) \times 1$. Entonces el modelo logístico teórico puede escribirse como:

$$g(E[Y_i | X_i = x_i] = \pi_i) = \text{logit}(\pi_i) = \log \left[\frac{\pi_i}{1 - \pi_i} \right] = x_i^T \beta = \eta_i$$

donde η es el predictor lineal y, despejando π_i , se obtiene:

$$\pi_i = \frac{\exp \{x_i^T \beta\}}{1 + \exp \{x_i^T \beta\}} = \frac{e^{\eta_i}}{1 + e^{\eta_i}}$$

O bien, en notación para cada individuo, sin la esperanza:

$$\text{logit}(\pi_i) = \log \left[\frac{\pi_i}{1 - \pi_i} \right] = x_i^T \beta + \varepsilon_i = \eta_i$$

El *logit* se denomina función de enlace porque mapea el rango $(0, 1)$ de las probabilidades al rango $(-\infty, +\infty)$ del predictor lineal. Sin embargo, también están como funciones de enlace: (i) complemento *cloglog*: $g(\pi_i) = \log(-\log(1 - \pi_i))$; (ii) *probit*: $g(\pi_i) = \Phi^{-1}(\pi_i)$ [Φ^{-1} es la función inversa de distribución acumulada de la normal estándar $N(0, 1)$]. El modelo resulta pues lineal en su función de enlace pero no en las proporciones. El Cuadro que sigue ilustra la relación lineal entre el *logit* de la proporción de ocurrencias del evento y los coeficientes β_j del modelo, según x_i sea continua o discreta (indicadora con $I = 1, \dots, k$ categorías), manteniendo fijos los valores de las restantes variables x_{-i} :

Tipo	β_j	Relación
Continua	$\beta_j = 0$	Sin relación entre logit y variable
	$\beta_j > 0$	Logit crece cuando x_i aumenta en una unidad
	$\beta_j < 0$	Logit decrece cuando x_i aumenta en una unidad
Discreta	$\beta_j = 0$	Sin relación entre logit y variable
	$\beta_j > 0$	Logit crece cuando x_i pasa de la categoría de referencia $I = k_{ref}$ a la categoría $I = k$
	$\beta_j < 0$	Logit decrece cuando x_i pasa de la categoría de referencia $I = k_{ref}$ a la categoría $I = k$

Cuadro 6.1: Relación lineal entre $\text{logit}(\pi_i)$ y β_j para variables continuas y discretas.

Por otra parte, la chance (*Odds*) de que ocurra el evento (respecto de que no ocurra) cuando $x_j = a$ (x_j representa ahora una covariable cualquiera), es:

$$\text{Odds}(x_i) = \frac{\pi(x_j = a)}{1 - \pi(x_j = a)} = \exp\{\beta_0 + \beta_j(x_j = a)\}$$

De modo análogo, la chance (*Odds*) de que ocurra dicho evento (respecto de que no lo haga) cuando $x_j = a + 1$ será:

$$\text{Odds}(x_j = a + 1) = \frac{\pi(x_j = a + 1)}{1 - \pi(x_j = a + 1)} = \exp\{\beta_0 + \beta_j(x_j = a + 1)\}$$

El cociente de chances (*Odds Ratio* = OR) compara $\text{Odds}(x + 1)$ con $\text{Odds}(x)$; o sea, cuando la covariable aumenta en un nivel. Al aumentar x en una unidad, dejando las otras predictoras fijas, se ve cómo aumenta o disminuye la chance de que $y = 1$ respecto de que $y = 0$. Si $a = 0$, se tiene:

$$\text{OR} = \frac{\text{Odds}(x_j = a + 1)}{\text{Odds}(x_j = a)} = \frac{\frac{\pi(x_j = a + 1)}{1 - \pi(x_j = a + 1)}}{\frac{\pi(x_j = a)}{1 - \pi(x_j = a)}} = \frac{\frac{\pi(1)}{1 - \pi(1)}}{\frac{\pi(0)}{1 - \pi(0)}} = \frac{\exp\{\beta_0 + \beta_j\}}{\exp\{\beta_0\}} = \exp\{\beta_0 + \beta_j - \beta_0\} = \exp\{\beta_j\}$$

Entonces $\exp\{\beta_j\}$ representa el cambio en el *Odds* (en la chance de que ocurra el evento respecto de que no ocurra) al incrementarse x en una unidad, dejando las demás predictoras fijas. El Cuadro que sigue muestra la interpretación para $\exp\{\beta_j\}$:

Tipo	β_j	$\exp\{\beta_j\}$	Relación
Continua	$\beta_j = 0$	$e^0 = 1$	La chance de $y = 1$ vs $y = 0$ no cambia si x_i aumenta en una unidad
	$\beta_j > 0$	$e^{\beta_j} > 1$	La chance de $y = 1$ vs $y = 0$ crece si x_i si x_i aumenta en una unidad
	$\beta_j < 0$	$e^{\beta_j} < 1$	La chance de $y = 1$ vs $y = 0$ decrece si x_i aumenta en una unidad
Discreta	$\beta_j = 0$	$e^0 = 1$	La chance de $y = 1$ vs $y = 0$ no cambia si x_i entre $I = k_{ref}$ y la categoría $I = k$
	$\beta_j > 0$	$e^{\beta_j} > 1$	La chance de $y = 1$ vs $y = 0$ crece entre $I = k_{ref}$ y la categoría $I = k$
	$\beta_j < 0$	$e^{\beta_j} < 1$	La chance de $y = 1$ vs $y = 0$ decrece entre $I = k_{ref}$ y la categoría $I = k$

Cuadro 6.2: Relación lineal entre $\text{Odds}(x_i)$ y β_j para variables continuas y discretas.

Ahora bien, para estimar $\pi_i(x_i)$, es decir, la probabilidad π_i de $y = 1$ dado cada nivel de x_i , es necesario observar su distribución: $y | X = x_i \sim \text{Bernoulli}(\pi(x_i))$ o lo que es igual: $y | X = x_i \sim \text{Binomial}(n = 1, \pi(x_i))$. Por lo tanto, la probabilidad de observar cada par (x_i, y_i) , para $y = \{0, 1\}$ será, según la función de masa de probabilidad:

$$\pi_i(y_i) = \text{Bi}(1, \pi(x_i)) = \pi_i(x_i)^{y_i} [1 - \pi_i(x_i)]^{1 - y_i}$$

Como se supone que las observaciones son independientes, la función de verosimilitud es el producto de la función anterior sobre N individuos:

$$\pi_i(y_1, \dots, y_n) = Bi(N, \pi(x)) = \mathcal{L}(\beta) = \prod_{i=1}^N \pi_i(x_i)^{y_i} [1 - \pi_i(x_i)]^{1-y_i} = \pi_i(x_i)^{\sum y_i} [1 - \pi_i(x_i)]^{N - \sum y_i}$$

La log-verosimilitud será:

$$\mathcal{L}(\beta) = \ln[L(\beta)] = \sum_{i=1}^N [y_i \ln(\pi_i(x_i)) + (1 - y_i) \ln(1 - \pi_i(x_i))]$$

El vector de parámetros estimado $\hat{\beta}$ serán los valores del vector que maximizan la función anterior de obtener los datos observados. Para encontrar dicho máximo se requiere diferenciar la función respecto del vector de parámetros:

$$\frac{\partial \log \mathcal{L}(\beta)}{\partial \beta} = \sum_{i=1}^N (y_i - \pi_i(x_i)) x_i$$

Como la ecuación es no lineal, se debe resolver mediante métodos numéricos (por ejemplo, *Newton-Raphson*). Si la solución converge, la varianza de los estimadores (de los coeficientes) por máxima verosimilitud se obtienen de la diagonal de la inversa negativa de la Hessiana (matriz de derivadas segundas). Los correspondientes desvíos típicos ($\widehat{SE}(\hat{\beta}_i)$) serán las raíces cuadradas de dichos valores diagonales. Estos se usan para calcular el test asintótico de *Wald*, lo cual permite calcular p-valores e intervalos de confianza, para cada coeficiente estimado. El estadístico de *Wald* para testear la hipótesis nula de que cada coeficiente es $H_0 : \beta_i = 0$ versus su alternativa $H_1 : \beta_i \neq 0$, reza:

$$W = \frac{\hat{\beta}_i}{\widehat{SE}(\hat{\beta}_i)} \approx N(0, 1)$$

Rechaza H_0 si $P(|z| > z_{obs}) < \alpha$. El Intervalo de Confianza (I.C.) de *Wald* de nivel de 95% para cada coeficiente β_i es: $\beta_j \pm \hat{z}_{1-\frac{\alpha}{2}} \widehat{SE}(\hat{\beta}_i)$; y su correspondiente I.C. para el OR [Odds Ratio] se obtiene exponenciando el anterior: $\exp\left\{\beta_j \pm \hat{z}_{1-\frac{\alpha}{2}} \widehat{SE}(\hat{\beta}_i)\right\}$. Para este caso, si $1 \notin IC(OR(\beta_i))$ entonces se rechaza la hipótesis nula.

La Devianza compara la verosimilitud del modelo ajustado con aquella del modelo saturado, que contiene tantos parámetros como datos. Por supuesto, el modelo saturado es muy complicado pero puede usarse como referencia. Si la diferencia entre los modelos ajustado y saturado es pequeña, entonces el modelo ajustado explica bien, con la ventaja de tener menos parámetros. En el modelo saturado, $\hat{\pi}_i = y_i$ y su verosimilitud es 1. Por lo tanto:

$$\mathcal{L}(\text{Modelo Saturado}) = \prod_{i=1}^N \pi_i(x_i)^{y_i} [1 - \pi_i(x_i)]^{1-y_i} = \prod_{i=1}^N y_i^{y_i} [1 - y_i]^{1-y_i} = 1$$

La devianza se define como:

$$\begin{aligned} D &= -2 \ln \left[\frac{\mathcal{L}(\text{Modelo Ajustado})}{\mathcal{L}(\text{Modelo Saturado})} \right] = -2 (\ln \mathcal{L}(\text{Modelo Ajustado}) - \ln \mathcal{L}(\text{Modelo Saturado})) \\ &= -2 (\ln \mathcal{L}(\text{Modelo Ajustado}) - \ln(1)) = -2 (\ln \mathcal{L}(\text{Modelo Ajustado}) - 0) \\ &= -2 (\ln \mathcal{L}(\text{Modelo Ajustado})) \end{aligned}$$

El test de cociente de verosimilitud (LRT) compara las devianzas del modelo más pequeño *anidado* en otro más grande, o sea con más parámetros. Se distribuye como χ^2 con $k - p$ grados de libertad, con k parámetros del modelo pequeño y p parámetros del grande. La hipótesis nula es que el modelo pequeño es correcto. Se rechaza cuando el estadístico toma valores grandes (p-valor chico).

$$G = -2[\ln \mathcal{L}(\text{Modelo Chico}) - \ln \mathcal{L}(\text{Modelo Grande})] = -2[D(\text{Modelo Chico}) - D(\text{Modelo Grande})] \sim \chi_{k-p}^2$$

Para la respuesta binaria, la bondad de ajuste del modelo se puede evaluar mediante el test de *Hosmer-Lemeshow*. La idea es que si el modelo ajusta bien no debería haber gran diferencia entre los valores observados y los esperados. Dicho test procede como sigue:

- Se definen g grupos aproximadamente del mismo tamaño según el orden de las probabilidades de ocurrencia del evento estimadas; o sea que, por ejemplo, las n/g observaciones con \hat{p} más grande definen el primer grupo.
- Se cuenta el número de ocurrencias del evento ($y_i = 1$) en cada uno de los grupos.
- Se estima el número de ocurrencias del evento en cada grupo por medio de la suma de las probabilidades estimadas de ocurrencia del evento para las observaciones de dicho grupo.
- Se compara el número de éxitos observados con los estimados en cada grupo con el estadístico que sigue:

$$\hat{C} = \sum_{i=1}^g \frac{(O_i - E[O_i])^2}{\text{Var}[O_i]} = \sum_{i=1}^g \frac{(O_i - n_i \hat{p}_i)^2}{n_i \hat{p}_i (1 - \hat{p}_i)} = \sum_{i=1}^g \frac{\left(\sum_{j=1}^{n_i} y_{ij} - \sum_{j=1}^{n_i} \hat{p}_{ij} \right)^2}{\left(\sum_{j=1}^{n_i} \hat{p}_{ij} \right) \left[1 - \left(\frac{1}{n_i} \sum_{j=1}^{n_i} \hat{p}_{ij} \right) \right]}$$

donde: n_i es el número total de casos en el i -ésimo grupo, $O_i = \sum_{j=1}^{n_i} y_{ij}$ ($j = 1, \dots, n_i$ observaciones en el i -ésimo grupo); y $\hat{p}_i = \frac{1}{n_i} \sum_{j=1}^{n_i} \hat{p}_{ij}$. El valor observado del estadístico \hat{C} se compara con el valor crítico de una distribución χ_{g-2}^2 , rechazando si $\hat{C}_{obs} > \chi_{g-2, 1-\alpha}^2$. Valores grandes de \hat{C} indican falta de ajuste, o sea que *no* se debe rechazar. Como cualquier otro test, su potencia (la probabilidad de rechazar correctamente un modelo que ajusta mal) aumenta con el tamaño muestral. Con el objetivo de independizar al test del tamaño muestral, Prabasaj et. al. (2013) sugieren calcular el número de grupos g mediante:

$$g = \max \left(10, \min \left\{ \frac{m}{2}, \frac{n-m}{2}, 2 + 8 \left(\frac{n}{1000} \right)^2 \right\} \right)$$

donde m es el número de ocurrencias del evento y n es el tamaño muestral. Para muestras con $n \leq 1000$, se obtiene $g = 10$; para $1000 \leq n \leq 25000$ se recomienda usar g usando la fórmula y para $n > 25000$ no se recomienda usar el test. Por otra parte, $g < 6$ tampoco debe usarse; y, al igual que para el test estándar de χ^2 , ningún valor esperado debe ser menor a 1 y el 80% deben ser mayores a 5.

También es posible agrupar los datos en patrones de covariables. Cada patrón es una determinada configuración de valores para las regresoras: $X_j^T = (x_1, x_2, \dots, x_p)$, con p variables independientes. Sean:

- j , el número de patrones (valores distintos de X observadas). Si hay tantos patrones como individuos, $n = j$. En cambio si algunos comparten los mismos valores de X_j , entonces $j < n$.
- m_j , los casos con los mismos valores de X_j . La suma de m_j constituye el número total de casos: $\sum m_j = n$.
- y_j , el número de respuestas $y = 1$ en los m_j casos con X_j . El número total de casos con $y = 1$ será $n_1 = \sum y_j$.

Por ejemplo, para la pequeña base de datos siguiente, con predictoras binarias x_1 y x_2 : $j = 2$; $m_1 = 3$; $m_2 = 1$; $y_1 = 2$; $y_2 = 1$; $n_1 = 3$. Los patrones observados son: $X_1 = (1, 0)$ y $X_2 = (0, 0)$. Los otros patrones posibles $(0, 1)$ y $(1, 1)$ no se observan.

ID	x_1	x_2	y
1	1	0	1
2	0	0	1
3	1	0	1
4	1	0	0

Dejando las p regresoras constantes y aumentando n total de los datos, cada valor de m_j tenderá a hacerse grande, se llamará a esto «m-asintótico». Esto sucede con $j < n$. En cambio, si $j \approx n$ (por ejemplo si alguna predictora es continua) o bien si $\hat{\pi}_j \approx 0$ o $\hat{\pi}_j \approx 1$ dicha condición m-asintótica *no* se puede usar. Con los datos agrupados de este modo es posible calcular residuales, con el objetivo de evaluar bondad de ajuste. Obsérvese que y_j se distribuye como binomial: $y_j \sim Bi(m_j, \pi_j = P(y = 1 | X_j))$. Por lo tanto, el valor predicho de y_j es:

$$\hat{y}_j = E[y_j] = m_j \hat{\pi}_j = m_j \frac{\exp\{\hat{g}(X_j)\}}{1 + \exp\{\hat{g}(X_j)\}}$$

donde: $\hat{g}(X_j) = \hat{\beta}_0 + \hat{\beta}_1 x_{j1} + \hat{\beta}_2 x_{j2} + \dots + \hat{\beta}_p x_{jp}$. Entonces, el residual para cada patrón resulta $y_j - \hat{y}_j$, o sea la distancia entre observados y predichos. El residual de Pearson será el residual estandarizado por el desvío típico estimado para cada patrón:

$$r(y_j, \hat{\pi}_j) = e_j = \frac{y_j - \hat{y}_j}{ES(\hat{y}_j)} = \frac{y_j - (m_j \hat{\pi}_j)}{\sqrt{m_j \hat{\pi}_j (1 - \hat{\pi}_j)}}$$

El estadístico de resumen se obtiene sumando sobre todos los patrones, distribuido como χ^2 con $J - (p + 1) = J - p - 1$ grados de libertad ($p + 1$ parámetros)¹:

$$\chi^2 = \sum_{j=1}^J [r(y_j, \hat{\pi}_j)]^2 \sim \chi_{J-p-1}^2$$

Por otra parte, el residual de devianza se define según el valor de y_j , con tres casos: (i) $y_j = 0$, o sea no hay en el patrón m_j casos de $y = 1$; (ii) $y_j = m_j$, es decir que todos los casos en m_j son $y = 1$; (iii) en cualquier otro caso (el signo \pm se corresponde con el del residual $y_j - \hat{y}_j$).

$$d(y_j, \hat{\pi}_j) = \begin{cases} \pm \left\{ 2 \left[y_j \ln \left(\frac{y_j}{m_j \hat{\pi}_j} \right) + (m_j - y_j) \ln \left(\frac{(m_j - y_j)}{m_j (1 - \hat{\pi}_j)} \right) \right] \right\}^{\frac{1}{2}} & ; y_j \neq \{0, m_j\} \\ -\sqrt{2m_j |\ln(1 - \hat{\pi}_j)|} & ; y_j = 0 \\ \sqrt{2m_j |\ln(\hat{\pi}_j)|} & ; y_j = m_j \end{cases}$$

Su estadístico de resumen también se distribuye como χ^2 con $J - p - 1$ grados de libertad:

$$D = \sum_{j=1}^J [d(y_j, \hat{\pi}_j)]^2 \sim \chi_{J-p-1}^2$$

Para hacer el test de bondad de ajuste se supone que la hipótesis nula indica «buen ajuste»; por ello, se rechaza cuando χ^2 y D toman valores grandes hacia la cola derecha de la distribución bajo la hipótesis nula.

Usando «m-asintótico», el test de Hosmer-Lemeshow reza:

$$\hat{C} = \sum_{k=1}^g \frac{(O_k - E[O_k])^2}{Var[O_k]} = \sum_{k=1}^g \frac{(O_{1k} - n_k \bar{\pi}_k)^2}{n_k \bar{\pi}_k (1 - \bar{\pi}_k)} = \sum_{k=1}^g \left[\frac{(O_{1k} - \hat{e}_{1k})^2}{\hat{e}_{1k}} + \frac{(O_{0k} - \hat{e}_{0k})^2}{\hat{e}_{0k}} \right]$$

¹La distribución χ^2 es una suma de normales estándar al cuadrado. Nótese que si m_j es grande, como sucedería si $J < n$, la variable aleatoria binomial y_j se aproximará a una normal de parámetros $\mu = m_j \hat{\pi}_j$ y $\sigma = \sqrt{m_j \hat{\pi}_j (1 - \hat{\pi}_j)}$. Entonces $r(y_j, \hat{\pi}_j)$ se distribuirá aproximadamente como normal estándar, y su cuadrado, como χ^2 .

donde: n_k es el número total de casos en el k -ésimo grupo; c_k es el número de patrones m_j en el k -ésimo grupo; $O_{1k} = \sum_{j=1}^{c_k} y_{kj}$ ($j = 1, \dots, c_k$ observaciones $y = 1$ en los c_k patrones del k -ésimo grupo); $O_{0k} = \sum_{j=1}^{c_k} (m_{kj} - y_{kj})$ ($j = 1, \dots, c_k$ observaciones $y = 0$ en los c_k patrones del k -ésimo grupo); $\hat{e}_{1k} = \sum_{j=1}^{c_k} m_{kj} \hat{\pi}_{kj}$; $\hat{e}_{0k} = \sum_{j=1}^{c_k} m_{kj} (1 - \hat{\pi}_{kj})$ y $\bar{\pi}_k = \frac{1}{n_k} \sum_{j=1}^{c_k} m_j \hat{\pi}_{kj}$.

La varianza para cada patrón supuesta por el modelo con datos agrupados es: $\sigma_e^2(X_j) = m_j \pi_j (1 - \pi_j)$. Sin embargo, la varianza estimada puede estar por arriba o por debajo de aquella supuesta, en cuyo caso el modelo sufrirá de sobre-dispersión o de sub-dispersión, respectivamente. Una forma de diagnosticar problemas de dispersión es mediante el ratio $Disp_1 = \frac{\chi^2}{n-p}$, donde hay p parámetros a estimar y $\chi^2 = \sum_{j=1}^J$

$[r(y_i, \hat{\pi}_j)]^2$ es el estadístico de resumen de Pearson presentado más arriba. Otra forma es mediante la devianza: $Disp_2 = \frac{D}{n-p}$; donde D es el estadístico de resumen de Devianza. Si $Disp \approx 1$ no habrá problemas de dispersión. En cambio, si $Disp > 1,25$ (tamaño muestral medio-moderado) o bien si $Disp > 1,05$ (tamaño muestral grande) hay sobre-dispersión (Hilbe, 2009, p. 321).

6.2. El modelo logístico mixto.

Si los datos conllevan alguna estructura de grupos (alumnos en cursos, ítems de respuesta en individuos, ítems en sesiones) entonces se puede dar que los casos dentro de cada grupo estén más correlacionados entre sí que con casos de otros grupos. En una situación como ésta el supuesto de independencia de las observaciones ya no rige. En un modelo mixto se plantean efectos fijos y aleatorios. Los segundos sirven para dar cuenta de la posible correlación intra-grupo e inter-grupo. Supóngase un experimento donde cada individuo (= grupo) responde correctamente ($Y = 0$) o incorrectamente ($Y = 1$) a una batería de ítems. Las respuestas de cada sujeto i tendrán una variabilidad «dentro» del sujeto pero también una «entre» sujetos. Por tanto, tendremos efectos aleatorios para dar cuenta de la varianza *dentro* y otros para la varianza *entre*.

Siguiendo a Hedeker & Gibbons (2006) y a Stroup (2013), sean $i = 1, \dots, N$ las unidades de segundo nivel (individuos / ítems); y $j = 1, \dots, n_i$, las unidades de primer nivel (grupos). El total de unidades de individuos es $n = \sum_{i=1}^N n_i$. Ahora $Y_{ij} \in \{0, 1\}$ representa el valor de la variable binaria del individuo i anidado en el grupo j . Por tanto, π_{ij} será la probabilidad de respuesta positiva: $Y_{ij} = 1$ para el caso i -ésimo del j -ésimo grupo. Por otro lado, $P(Y_{ij} = 0) = 1 - \pi_{ij}$ resulta la probabilidad de respuesta negativa para el caso i -ésimo en el j -ésimo grupo. Sea el vector de dimensión $(1 \times (p+1))$ de covariables para el caso i -ésimo en el grupo j -ésimo (incluido el intercepto): $\mathbf{x}_{ij} = (1, x_{ij1}, \dots, x_{ijp})$; el vector de coeficientes del modelo es $\boldsymbol{\beta} = (\beta_0, \beta_1, \dots, \beta_p)$, de dimensión $(p+1) \times 1$. Además, sea el vector de dimensión $(1 \times q_v)$ de factores aleatorios «entre grupos» para el caso i -ésimo en el grupo j -ésimo (uno para cada individuo): $\mathbf{z}_v = (z_{ij1}, \dots, z_{ijq_v})$; y el vector de coeficientes aleatorios del modelo es $\mathbf{v} = (v_{1ij}, v_{2ij}, \dots, v_{q_v ij})$, de dimensión $q_v \times 1$. Asimismo, sea el vector de factores aleatorios «dentro» del modelo $\boldsymbol{\omega} = (\omega_0, \omega_1, \dots, \omega_{q_\omega})$, de dimensión $q_\omega \times 1$ y el vector de dimensión $(1 \times q_\omega)$ de factores aleatorios «dentro de los grupos» para el caso i -ésimo en el grupo j -ésimo (uno para cada individuo): $\mathbf{z}_\omega = (z_{ij1}, \dots, z_{ijq_\omega})$. Entonces se puede expresar el modelo mixto para cada individuo en el grupo j -ésimo en términos del logit como:

$$g(E[Y_{ij} | z_{ij}; w_{ij}] = \pi_{ij}) = \text{logit}(\pi_{ij}) = \log \left[\frac{\pi_{ij}}{1 - \pi_{ij}} \right] = x_{ij} \boldsymbol{\beta} + z_v \mathbf{v}_{ij} + z_\omega w_{ij} = \eta_{ij}$$

Como antes, la inversa del *logit* arroja la probabilidad:

$$g(E[Y_{ij} | z_{ij}; w_{ij}] = \pi_{ij}) = g^{-1}[\eta_{ij} = x_{ij} \boldsymbol{\beta} + z_v \mathbf{v}_{ij} + z_\omega w_{ij}]$$

Obsérvese que el modelo es *condicional* a los efectos aleatorios de grupo «dentro» y «entre». Es decir que el modelo podrá estimar una esperanza $\hat{\pi}_j$ para cada grupo y no solamente una general para todos los datos, como en la regresión logística ordinaria. Por ejemplo, el i -ésimo ítem en un grupo j -ésimo, con dos variables binarias (y el intercepto) $x_1 = 1$ y $x_2 = 0$; un efecto aleatorio «entre» v_{0ij} (por individuo) de ordenada al origen (grupo) y efecto aleatorio «dentro» w_{0i} , será:

$$\text{logit}(\pi_{ij}) = \eta_{ij} = \beta_0 + \beta_1(x_{ij1} = 1) + \beta_2(x_{ij2} = 0) + v_{0ij}(1) + w_{ij}(1) = (\beta_0 + [v_{0ij} + w_{ij}]) + \beta_1(x_{ij1} = 1)$$

O sea que el efecto aleatorio de ordenada al origen es la *diferencia o desviación* entre la ordenada al origen de la regresión general para todos los datos y la ordenada al origen de *cada uno* de los grupos. Obsérvese que, al contrario de un modelo gaussiano, la variabilidad «dentro» forma parte del predictor lineal η_{ij} . Por esta razón, se dice que los efectos «dentro» se modelan en el «G-side». Ahora, las probabilidades son condicionales al efecto aleatorio, o sea que la probabilidad de éxito («error») será: $P(Y_{ij} = 1 | \theta = z_i) = g^{-1}(\eta_{ij})$ (para simplificar solo se considera el efecto aleatorio de ordenada al origen z_i). En tanto que aquella de fracaso será: $P(Y_{ij} = 0 | \theta = z_i) = 1 - g^{-1}(\eta_{ij})$. El supuesto fundamental es que las observaciones *dentro* de un mismo sujeto (grupo), *condicionales* al efecto aleatorio son independientes; es decir que los efectos aleatorios explican la correlación intra-grupo. Por lo tanto en la función de verosimilitud basta multiplicar las probabilidades condicionales (a los efectos aleatorios) sobre las n_i observaciones para *cada* grupo j -ésimo [Hedeker & Gibbons, 2006, p. 162]:

$$\mathcal{L}(Y_j | \theta = z_i) = \prod_{j=1}^{n_i} g^{-1}(\eta_{ij})^{Y_{ij}} [1 - g^{-1}(\eta_{ij})]^{1 - Y_{ij}}$$

En cambio, para obtener la verosimilitud de los j grupos, es decir, en la población; resulta necesario promediar (integrar) sobre los efectos aleatorios para cada grupo y luego multiplicar los grupos, para así obtener la expresión de la verosimilitud marginal (poblacional) para todos los grupos.

$$\mathcal{L} = \prod_{j=1}^N \left[\int_{\theta} \mathcal{L}(Y_j | \theta) g(\theta) d\theta \right]$$

El modelo mixto logístico posee los siguientes supuestos:

- (1) los errores del sujeto i (efectos aleatorios de π_{ij} *dentro* de los grupos) se distribuyen como normal multivariada con vector de media θ y matriz de varianza R_i (una para cada grupo): $w_{ij} \sim N_{n_i}(\theta, R_i)$, n_i es la cantidad de observaciones repetidas por grupo;
- (2) los efectos aleatorios *entre grupos* se distribuyen como normal multivariada con vector de media θ y matriz de varianza G : $v \sim N_q(\theta, G)$, q es la cantidad de grupos;
- (3) G y R_i son matrices positivas definidas: $G > 0$; $R_i > 0$;
- (4) Los factores aleatorios *dentro* son independientes de los factores aleatorios *entre*: $\text{cov}(v, w') = 0$;
- (5) Las observaciones del grupo i (π_{ij}) *condicional* a los efectos aleatorios de los vectores v y ω se distribuyen como binomial con media $\pi_{ij} = g^{-1}(X\beta + Z_v v + Z_\omega w)$ y varianza $\text{var}(y_{ij} | v, \omega) = V_{n_i x n_i} = \text{diag}(\pi_{ij}(1 - \pi_{ij}))$; $\text{var}(\omega_{ij}) = R_i = W_{n_i} \sigma_\omega^2$; $\text{var}(v) = G = A \sigma_v^2 = \sigma_\omega^2$. Además de (1) y (2): $\begin{bmatrix} v \\ \omega \end{bmatrix} \sim N\left(\theta, \begin{bmatrix} G & \mathbf{0} \\ \mathbf{0} & R \end{bmatrix}\right)$.

Generalmente: $A = I$, o sea con efectos aleatorios *entre sujetos* independientes y $W_{n_i} = I_{n_i}$; $\text{cov}(w_{ij}, w_{ij'}) = 0$ [$j = j'$] (los efectos aleatorios *dentro de un mismo sujeto* son independientes).

6.3. Selección de modelos.

Se usaron las siguientes medidas para seleccionar modelos (Stroup, 2013; p. 193; Burnham & Anderson, 2010, caps. 2 y 4) [$\theta = \hat{\sigma}, \hat{\beta}$ es el vector de los coeficientes fijos y aleatorios estimados]:

- (a) *Schwarz*: $BIC = -2L(\theta) + (p \times \log(s))$ [$s =$ número de grupos; y $p = p_\sigma + (p_\beta = \text{rank}[X])$; o sea el número de parámetros fijos más los aleatorios]; Menos es mejor.
- (b) *Akaike*: $AIC = -2L(\theta) + 2p$; Menos es mejor.

- (c) Akaike corregido: $AIC_c = -2L(\theta) + 2p(n^*/(n^* - p - 1))$ [$n^* = N$, tamaño muestral]; Corrige por muestra pequeña. Menos es mejor. Como heurística, se debería usar cuando²: $\frac{n}{p} < 40$.
- (d) Prueba de cociente de máxima verosimilitud [LRT], detallado más arriba. Si se testean los efectos aleatorios en el límite del espacio del parámetro (o sea: $H_0 : \sigma_v^2 = 0$), debe procederse a una corrección del p-valor según las distribuciones mixtas: $0,5(\chi_{(0)}^2 + \chi_{(1)}^2) = 0,5\chi_{(1)}^2$ [modelo sin efecto aleatorio de intercepto vs modelo con efecto aleatorio de intercepto]; y $0,5(\chi_{(1)}^2 + \chi_{(2)}^2)$ [modelo con efecto aleatorio de intercepto vs modelo aleatorio con efecto aleatorio adicional].
- (e) Delta de Akaike: $\Delta AIC = \Delta = AIC_i - AIC_{min}$. Indican la distancia del modelo al mejor de todos (el de menor AIC). $\Delta_i \leq 2$ indica evidencia substancial para el modelo i .
- (f) Pesos de Akaike (ω_i): indican el peso de la evidencia en favor de que modelo sea el mejor de entre todos los modelos candidatos. Es decir, responde a la pregunta: ¿Cómo soportan los datos al modelo i con respecto al resto de los modelos? Se define como:

$$\omega_{modelo_i} = \frac{\exp(-\Delta_i/2)}{\sum_{i=1}^R \exp(-\Delta_i/2)}$$

donde $i = 1, \dots, R$ son los modelos considerados; y $\sum_{i=1}^R \omega_i = 1$.

- (g) Ratio de evidencia («Evidence Ratio», [ER]): Ratio entre el peso de Akaike del modelo i -ésimo y el peso de Akaike del j -ésimo modelo: $\frac{W(i)}{W(j)}$. Muchas veces resulta de interés establecer i como el índice del mejor modelo: $\frac{W(1)}{W(j)}$. Los ER son invariantes a los demás modelos, a parte de i y j . Responden a la pregunta: ¿Cuántas más veces apoyan los datos al (mejor) modelo i respecto del modelo j ?

Una vez ordenados los modelos según alguno de los criterios, se puede reducir dicho conjunto por medio de un «conjunto de confianza» [*confidence set*] para el mejor modelo hallado. Burnham & Anderson (2010, p. 169) plantean tres alternativas: (i) sumar los pesos de Akaike de los modelos hasta alcanzar $\geq 0,95$ (recuérdese que los pesos de Akaike suman 1); (ii) tomar los modelos tal que $\Delta_i \leq 2$, ya que indican evidencia substancial para el modelo i ; (iii) establecer un corte usando ratios de evidencia (poniendo ahora el mejor modelo en el denominador), tal que³: $\frac{W(i)}{W(1)} > \frac{1}{8}$ ($\Delta_i = 2$). Los autores prefieren el tercer criterio debido a su invariancia por adición o borrado de modelos del conjunto de confianza.

Resulta imperativo tener en cuenta la incerteza debida al proceso de selección de modelos. De R modelos considerados se selecciona el mejor modelo i . Sin embargo, ¿Si hubieran cambiado los datos, se elegiría igualmente el modelo i como el mejor o habría variabilidad de entre las muestras de datos en cuanto al modelo elegido?. Una forma de tener en cuenta dicha incerteza es estimar la probabilidad de que un determinado predictor x_j esté en el mejor modelo si se pudiera recoger una nueva muestra de datos. Se trata de una medida de importancia relativa de los predictores. Se lleva a cabo sumando los pesos de Akaike de los modelos en los cuales el predictor x_j está presente: $W_+ = w_i I_j(g_i)$; donde $I_j(g_i)$ es la función indicadora que es «1» si x_j está en el modelo g_i o cero, si no. Entonces, la importancia relativa es la proporción de modelos en los cuales la predictora está presente.

Si se diera el caso de que, por ejemplo, $w(i) > 0,9$, entonces el modelo i es un claro ganador. En dicho caso es válido hacer inferencia mediante la estimación de los coeficientes β_i y sus errores típicos serán condicionales al modelo seleccionado. Sin embargo, muchas veces, especialmente si el conjunto de modelos a considerar es grande, los modelos con $\Delta_i \leq 2$ poseen pesos de Akaike similares o bien deltas de Akaike cercanos al cero. En este caso, β_i puede diferir en los modelos del conjunto considerado. Una solución es usar la información de todos los modelos involucrados mediante un promedio pesado de los coeficientes.

²Las medidas AIC y AIC_c convergen para n grande (manteniendo p constante). Es decir que cuando dicho ratio es suficientemente grande, tienden a seleccionar el mismo modelo. Entonces, en la práctica conviene usar siempre AIC_c .

³También podrían usarse: 0.135 ($\Delta_i = 4$); 0.082 ($\Delta_i = 5$); 0.05 ($\Delta_i = 6$)

En este caso, los errores típicos de los coeficientes estimados no son condicionales al modelo (ganador) en cuestión sino a todo el conjunto de modelos. Por lo tanto, dichos errores típicos «incondicionales» tienen en cuenta la varianza que proviene del proceso de selección de modelos. Para promediar los coeficientes se utilizó:

$$\bar{\beta}_j = \sum_{i=1}^R w_i I_j(g_i) \hat{\beta}_{j,i} = W_+ \hat{\beta}_{j,i}$$

donde:

$$I_j(g_i) = \begin{cases} 1 & x_j \in g_i \\ 0 & x_j \notin g_i \end{cases}$$

y la suma es sobre todos los modelos del conjunto: $i = 1, \dots, R$. En este estimador se usan todos los modelos («full average»), y cuando la predictora x_j no estuviera presente en un determinado modelo entonces $\beta_j = 0$. Tiene la ventaja de «correr hacia cero» [*Shrinkage*] las estimaciones de parámetros presentes en «modelos malos». La varianza del estimador resulta:

$$\widehat{\text{var}}(\bar{\beta}_j) = \left[\sum_{i=1}^R w_i \sqrt{\widehat{\text{var}}(\hat{\beta}_j | g_i) + (\hat{\beta}_j - \bar{\beta}_j)^2} \right]^2$$

y su error típico: $\sqrt{\widehat{\text{var}}(\bar{\beta}_j)}$.

6.4. Expectativa de análisis.

Se desea encontrar las variables que influyen en que un caso de concordancia tenga un error ($Y = 1$). Se esperan los siguientes resultados de los predictores, en cuanto a la *facilitar* con menor (\uparrow) / mayor ($\uparrow\uparrow$) tamaño de efecto, o *inhibir* (\downarrow) errores de concordancia⁴. Solo se pusieron en el Cuadro los predictores discretos, pero la dirección esperada del efecto es igual para la contraparte continua.

⁴Debido a la naturaleza exploratoria de la tesis, no se trata de hipótesis a ser confirmadas porque se requeriría un estudio experimental aleatorizado que controlara los sesgos.

Predictor	Efecto	dirección / tamaño
MOD	MOD(1) - MOD(0)	↑
	MOD(2) - MOD(0)	↑
	MOD(3) - MOD(0)	↑
Fabs.SC.f	Fabs.SC.f(1) - Fabs.SC.f(0)	↓
MORF.f	MORF.f(1) - MORF.f(0)	↑
	MORF.f(2) - MORF.f(0)	↑↑
CUMRES.f	CUMRES.f(1) - CUMRES.f(0)	↑
	CUMRES.f(2) - CUMRES.f(0)	↑↑
FAM.LEX.f	FAM.LEX.f(1) - FAM.LEX.f(0)	↓
EST1	EST1(1) - EST1(0)	↓
EST2	EST2(1) - EST2(0)	↓
EST3	EST3(1) - EST3(0)	↑
EST4	EST4(1) - EST4(0)	↓
EST5	EST5(1) - EST5(0)	↓
EST4	EST6(1) - EST6(0)	↓
EST7	EST7(1) - EST7(0)	↑
ANIM	ANIM(1) - ANIM(0)	↑
ES	ES(1) - ES(0)	↑
	ES(2) - ES(0)	↑↑
POS	POS(1)-POS(0)	↑
	POS(2)-POS(0)	↑↑
DIS	POS(1)-POS(0)	↑
	POS(2)-POS(0)	↑↑
LDA	LDA(1)-LDA(0)	↑
GRAMS	GRAMS(1)-GRAMS(0)	↑
STEM.f	STEM.f(1)-STEM.f(0)	↑
IMA.CONC.f	IMA.CONC.f(1) - IMA.CONC.f(0)	↓
GRUPO6	GRUPO6(2)-GRUPO6(1)	↑
GRUPO6	GRUPO6(3)-GRUPO6(1)	↑
GRUPO6	GRUPO6(4)-GRUPO6(1)	?
GRUPO6	GRUPO6(5)-GRUPO6(1)	↑
GRUPO6	GRUPO6(6)-GRUPO6(1)	?

Cuadro 6.3: Expectativas del análisis. Leyenda: (↑) = efecto aumenta; (↓) = efecto disminuye; (↑↑/↓↓) = mayor tamaño de efecto; ? = efecto indeterminado.

Según el Cuadro 3, se espera hallar más errores con concordancias con: (i) artículos indefinidos, determinantes y adjetivos; (ii) concordancias a larga distancia (la chance de error crece con: distancia larga / sin subordinada > distancia entre enunciados / con subordinada); (iii) *-e-* epentética, con chance de error creciente si está en un término o en ambos; (iv) concordancias de más de dos términos; (v) controlador animado; (vi) TYPE infrecuente; (vii) similaridad baja (alta distancia) entre las raíces léxicas del español y el italiano; (viii) similaridad media (media distancia) entre las terminaciones del español y el italiano, según el tamaño de efecto *baja > media*; (ix) cantidad media > alta de errores acumulados; (x) controlador infrecuente y/o no familiar; (xi) controlador de baja imaginabilidad y/o concretud; (xii) las «estrategias» EST3 y EST7 porque identifican casos de difícil producción para los italianos; (xiii) en GRUPO6, los grupos (5), (2), (3); el primero porque concentra casos de larga distancia, infrecuentes y sin artículos definidos; el segundo, por aglomerar casos de MORF.f alta y con *-e-* epentética en un término; el tercero, por concentrar casos de más de dos términos y con distancias MORF.f media. En los demás niveles de GRUPO6 el efecto se considera indeterminado, por eso se marca «?».

6.5. Análisis de datos.

Los grupos son las sesiones, *anidados* en los sujetos; lo cual arrojó un total de 52 grupos. Cada grupo tenía entre 4 y 81 ítems de concordancia. Por lo tanto, había desbalance, que un modelo mixto logra manejar sin problemas. Se ajustaron dos modelos mixtos condicionales en el «*G side*» (Stroup, 2013; p. 188). Debido a la cantidad de variables, no fue posible aplicar un estrategia «de arriba hacia abajo» con el modelo casi saturado, para construir el modelo. Se seleccionaron modelos utilizando el paradigma de información (Burnham & Anderson, 2010). Se estimaron los parámetros con el método de *aproximación de Laplace* para evitar el uso de *pseudo-verosimilitud* y poder hacer comparación de modelos con diferentes estructuras de covarianza. Las sesiones (grupos) estaban espaciadas en el tiempo. Por ende, no se esperó que hubiera correlación en los efectos aleatorios «entre» (matriz $G = \sigma_0^2 I$). En cambio, en el transcurso de cada sesión se esperaba que errores en concordancias cercanas a la que el hablante procesa influyeran; por lo cual ya no cabría suponer independencia en los efectos aleatorios «dentro». Se usó el software *R* (diferentes librerías).

6.5.1. Modelo logístico.

A modo exploratorio, se comenzó con un modelo nulo (solo ordenada al origen) y se agregó cada predictora a éste, creando tantos modelos como predictoras. Se evaluó la significancia estadística de cada predictora mediante la prueba de cociente de máxima verosimilitud [LRT]. El Cuadro 4 muestra los resultados. Resultaron estadísticamente significativos los predictores: ADJ, MOD, Fabs.C.S, Fabs.SC.f, MORF, MORF.f, CUMRES.f, CUMRES, IMA.CONC, FAM.LEX.f, EST1, EST3, EST5, EST6, GRUPO6. Ahora bien, MOD es una especificación de ADJ, por tanto se eligió el primero. En cuanto a MORF y MORF.f, el estadístico χ^2 de MORF.f resultó dos veces el de MORF, por ello se eligió la versión discreta. En lo que atañe a CUMRES / CUMRES.f y Fabs.C.S / Fabs.SC.f no hubo grandes diferencias entre ambas versiones. Se eligió la versión discreta para Fabs (Fabs.SC.f) y la continua para el otro (CUMRES).

La prueba de LRT así planteada sufre de inconvenientes: (i) no se tuvo en cuenta la variabilidad por el proceso de selección de modelos; (ii) no se consideró la cuestión de las comparaciones múltiples contra el modelo nulo, ajustando el p-valor. Por estas razones, no se puede decidir solamente sobre esta base. Por ende, se practicó selección de modelos basado en medidas de información. Primero, se ajustaron modelos logísticos (no mixtos) con las predictoras significativas seleccionadas. Fueron $2^{11} - 1 = 2047$ modelos, jerarquizados mediante la medida de información AIC (como: como: $\frac{n}{p} = \frac{1857}{11} \approx 168 > 40$, no se usó la versión *AICc* corregida por tamaño muestral). Luego se examinó la frecuencia de las predictoras en el conjunto completo de modelos, que da un panorama de la incerteza por la selección. A continuación se redujo la cantidad de modelos al subconjunto «de confianza» con la regla $\frac{W(i)}{W(1)} > \frac{1}{8}$. Sobre dicho subconjunto se llevó a cabo un promedio de coeficientes con la varianza calculada con «full average». Se repitió el proceso con las variables significativas elegidas, agregando interacciones hasta el segundo nivel.

El Cuadro 5 ilustra los resultados de importancia de los predictores. Los primeros ocho se encuentran en la gran mayoría de los modelos. Nótese la poca proporción de GRUPO6.

Los promedios de coeficientes resultaron los que siguen en el Cuadro 6. Las predictoras significativas fueron: FAM.LEX.f, Fabs.SC.f, MOD, MORF.f, CUMRES, EST1 y EST5. Nótese que IMA.CONC y EST3 ya no se seleccionan.

A continuación se examinaron modelos logísticos que incluían las interacciones entre las predictoras seleccionadas, menos CUMRES, EST1 EST5, que no entraron como términos de interacción. El Cuadro 7 muestra la importancia de cada predictora. La interacción que más aparece es Fabs.SC.f:MORF.f, que casi dobla la frecuencia relativa de la la siguiente más importante (FAM.LEX.f:Fabs.SC.f). Por otro lado, en el Cuadro 8 se hallan los coeficientes promediados (con *full average*), los modelos elegidos en el «conjunto de confianza». Se observa que *ninguna* interacción resulta significativa. Por tanto, se elige un modelo sin interacciones, con las variables: MOD, Fabs.SC.f, MORF.f, FAM.LEX.f, CUMRES, EST1, EST5. Para dicho modelo el test de Hosmer-Lemeshow (usando 29 grupos) arroja: $\chi_{(29)} = 27,065$, $p = 0,46$, es decir que no se rechaza la hipótesis nula de falta de ajuste. En el apéndice se detalla el Cuadro con los resultados por grupo⁵. Si bien el modelo con respuesta binaria no puede estar sobre-disperso, se evaluó

⁵No hay ningún valor esperado menor a 1 y todos son mayores que 5. Se verifica $N > 30$ en todos los grupos.

	term	df	Deviance	AIC	LRT	p.value
1	<none>		2107.5195	2109.5195		
2	ADJ	1.0000	2080.0300	2084.0300	27.4895	0.0000
3	MOD	3.0000	2073.7230	2081.7230	33.7965	0.0000
4	POS	2.0000	2103.6595	2109.6595	3.8599	0.1452
5	DIS	3.0000	2102.4288	2110.4288	5.0907	0.1653
6	LDA	1.0000	2104.1555	2108.1555	3.3640	0.0666
7	ES	2.0000	2103.9476	2109.9476	3.5718	0.1676
8	GRAMS	1.0000	2105.9547	2109.9547	1.5648	0.2110
9	ANIM	1.0000	2104.9875	2108.9875	2.5319	0.1116
10	Fabs.C.S	1.0000	2067.7046	2071.7046	39.8148	0.0000
11	Fabs.SC.f	1.0000	2077.9464	2081.9464	29.5730	0.0000
12	STEM.s	1.0000	2106.0298	2110.0298	1.4897	0.2223
13	MORF	1.0000	2096.8484	2100.8484	10.6711	0.0011
14	MORF.f	2.0000	2081.6360	2087.6360	25.8834	0.0000
15	STEM.f	1.0000	2106.9098	2110.9098	0.6097	0.4349
16	CUMRES.f	2.0000	2092.5109	2098.5109	15.0086	0.0006
17	CUMRES	1.0000	2091.6342	2095.6342	15.8853	0.0001
18	IMA.CONC	1.0000	2103.2496	2107.2496	4.2699	0.0388
19	FAM.LEX	1.0000	2107.5108	2111.5108	0.0087	0.9256
20	IMA.CONC.f	1.0000	2107.0162	2111.0162	0.5033	0.4781
21	FAM.LEX.f	1.0000	2101.8333	2105.8333	5.6862	0.0171
22	GRUPO6	5.0000	2095.0395	2107.0395	12.4799	0.0288
23	EST1	1.0000	2095.4855	2099.4855	12.0339	0.0005
24	EST2	1.0000	2104.9913	2108.9913	2.5282	0.1118
25	EST3	1.0000	2101.0136	2105.0136	6.5059	0.0108
26	EST4	1.0000	2106.3110	2110.3110	1.2085	0.2716
27	EST5	1.0000	2103.2876	2107.2876	4.2319	0.0397
28	EST6	1.0000	2101.1839	2105.1839	6.3355	0.0118
29	EST7	1.0000	2103.7768	2107.7768	3.7427	0.0530

Cuadro 6.4: Test de cociente de verosimilitud [LRT]: modelo nulo contra cada predictor.

si había una posible sobre-dispersión «oculta» (Hilbe, 2009, p. 341), pasando los datos al formato de masintótico (poniendo CUMRES como discreta), con: $j = 203 < n = 1857$. En el caso del modelo elegido, $Disp_1 = \frac{\chi^2}{n-p} = \frac{237,3471}{191} = 1,12$; $Disp_2 = \frac{D}{n-p} = \frac{214,884}{191} = 1,12$; es decir que no se detectó sobre-dispersión ($Disp > 1,25$). Por último, se examinará la asociación entre las predictoras del modelo y su posible colinealidad⁶. El Cuadro 9 muestra el estadístico de χ^2 y la medida de asociación V de Cramer, que ajusta el primero por tamaño muestral: $V = \sqrt{\chi^2/N(df)}$; donde N es el tamaño muestral y $df = \min(F - 1; C - 1)$ es el mínimo entre las filas menos uno y las columnas menos uno de la tabla de contingencia múltiple entre dos variables categóricas. Según Cohen (1988), una magnitud de efecto media es: (i) $V \geq 0,3$ si $df = 1$; (ii) $V \geq 0,21$ si $df = 2$; (iii) $V \geq 0,17$ si $df = 3$. Se observa que esto se cumple para los pares de variables MOD:Fabs.SC.f, MORF.f:EST1 y MORF.f:EST5. Se podría esperar que el efecto de MOD esté ajustado por el efecto de Fabs.SC.f; y el de MORF.f por los efectos de EST1 y EST5. Por otra parte, el Cuadro diez investiga los valores de $GVIF$, un índice de inflación de varianzas (para modelos generalizados) para cada predictora del modelo, ya que todas están debajo de cinco, no se observan problemas de colinealidad.

⁶Colinealidad es la alta correlación entre predictoras del modelo. Un posible efecto de esta es que se inflan las varianzas de las predictoras. Si la correlación es alta entre dos predictoras, ambas conllevan casi la misma información y se hace difícil discernir entre cuál de las dos es que se halla asociada con la respuesta.

	names	x
1	CUMRES	1.00
2	MOD	1.00
3	MORF.f	1.00
4	FAM.LEX.f	1.00
5	EST1	0.99
6	Fabs.SC.f	0.97
7	EST5	0.97
8	IMA.CONC	0.86
9	EST3	0.52
10	EST6	0.39
11	GRUPO6	0.02

Cuadro 6.5: Importancia relativa de las predictoras.

	Estimate	Std. Error	Adjusted SE	z value	Pr(> z)
(Intercept)	-0.563	0.240	0.240	2.344	0.019
CUMRES	0.267	0.063	0.063	4.234	0.000
EST11	-0.475	0.134	0.134	3.530	0.000
EST31	0.660	0.879	0.879	0.751	0.453
EST51	-0.585	0.199	0.199	2.938	0.003
FAM.LEX.f1	-0.407	0.114	0.114	3.578	0.000
Fabs.SC.f1	-0.408	0.135	0.135	3.018	0.003
IMA.CONC	0.091	0.049	0.049	1.861	0.063
MOD1	-0.507	0.474	0.474	1.070	0.285
MOD2	0.533	0.142	0.142	3.760	0.000
MOD3	0.324	0.161	0.161	2.006	0.045
MORF.f1	-0.718	0.186	0.186	3.869	0.000
MORF.f2	-0.512	0.218	0.218	2.344	0.019
EST61	0.102	0.204	0.204	0.498	0.618

Cuadro 6.6: Promedio de los coeficientes con FULL AVERAGE.

6.5.2. Modelo logístico mixto.

Hasta ahora no se ha tenido en cuenta la naturaleza agrupada de los datos, que puede conllevar correlación entre los elementos de un grupo. La Figura 1 ofrece un gráfico de la función de auto-correlación para el modelo logístico elegido en el apartado anterior. Se observa que existe una correlación significativa (fuera de las bandas de confianza) en el $lag=1$. El test de *Durbin-Watson* (para series temporales) testea la hipótesis nula de *no correlación* en el $lag=1$ entre los residuos del modelo, $W = 2$ corresponde al ruido blanco, o sea independencia. En este caso el test confirma el gráfico: $W = 1,7158$; $p < 0,0001$.

	names	x
1	MORF.f	1.00
2	CUMRES	1.00
3	FAM.LEX.f	1.00
4	Fabs.SC.f	1.00
5	MOD	1.00
6	EST1	0.99
7	EST5	0.99
8	Fabs.SC.f:MORF.f	0.76
9	FAM.LEX.f:Fabs.SC.f	0.39
10	FAM.LEX.f:MORF.f	0.26
11	FAM.LEX.f:MOD	0.13
12	Fabs.SC.f:MOD	0.13
13	MOD:MORF.f	0.06

Cuadro 6.7: Importancia Relativa de las predictoras.

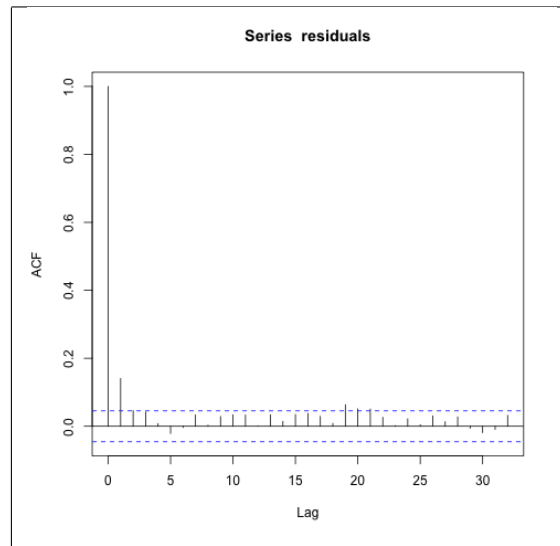


Figura 6.1: Función de Auto-correlación para el modelo logístico.

Modelar la estructura de covarianza significa representar (para cada sujeto / grupo) $var(y_i | v_i, \omega_i) = V_i$ en función de un número relativamente pequeño de parámetros (Littel *et al.*, 2000). Su especificación funcional se realiza a través de la forma que asumen G y R_i . El supuesto de errores «dentro» independientes muchas veces es irreal. Entonces es usual modelar la matriz W_i mediante diferentes estructuras. La siguiente constituye la estructura de correlación que modela un proceso «AR(1)» [Hedeker & Gibbons, 2006]:

$$R_i = \sigma_{\omega}^2 \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{t-1} \\ \rho & 1 & \rho & \dots & \rho^{t-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{t-3} \\ \dots & \dots & \dots & \dots & \dots \\ \rho^{t-1} & \rho^{t-2} & \rho^{t-3} & \dots & 1 \end{bmatrix}$$

O sea que, para un mismo sujeto / grupo, los efectos aleatorios «dentro» tienen varianza homogénea: $var(w_{ij}, w_{ij'}) = \sigma_{\omega}^2$ ($j = j'$) y covarianza heterogénea que decrece hacia cero a medida que se incrementa el lag entre observaciones: $cov(w_{ij}, w_{ij'}) = \sigma_{\omega}^2 \rho^{|j-j'|}$ ($j > j'$). La correlación es: $\phi = \frac{\sigma_{\omega}^2 \rho^{|j-j'|}}{(\sqrt{\sigma_{\omega}^2})^2} = \rho^{|j-j'|}$.

Nótese que, como se supone estacionariedad, la correlación no depende del valor del lag, es decir que, dado

	Estimate	Std. Error	Adjusted SE	z value	Pr(> z)
(Intercept)	-0.442	0.286	0.287	1.541	0.123
CUMRES	0.269	0.063	0.063	4.266	0.000
EST11	-0.447	0.134	0.135	3.318	0.001
EST51	-0.635	0.198	0.198	3.213	0.001
FAM.LEX.f1	-0.535	0.263	0.263	2.038	0.042
Fabs.SC.f1	-0.436	0.400	0.400	1.089	0.276
MOD1	-0.432	0.541	0.541	0.798	0.425
MOD2	0.505	0.148	0.148	3.418	0.001
MOD3	0.328	0.164	0.164	1.992	0.046
MORF.f1	-0.760	0.258	0.258	2.949	0.003
MORF.f2	-0.754	0.300	0.300	2.512	0.012
Fabs.SC.f1:MORF.f1	-0.151	0.402	0.402	0.375	0.708
Fabs.SC.f1:MORF.f2	0.430	0.470	0.470	0.915	0.360
FAM.LEX.f1:Fabs.SC.f1	0.085	0.179	0.179	0.473	0.636
FAM.LEX.f1:MORF.f1	0.097	0.260	0.260	0.374	0.708
FAM.LEX.f1:MORF.f2	0.089	0.259	0.259	0.343	0.731
FAM.LEX.f1:MOD1	-0.073	0.391	0.391	0.187	0.852
FAM.LEX.f1:MOD2	0.001	0.060	0.060	0.015	0.988
FAM.LEX.f1:MOD3	0.001	0.061	0.061	0.015	0.988
Fabs.SC.f1:MOD1	0.000	0.000	0.000		
Fabs.SC.f1:MOD2	-0.004	0.065	0.065	0.067	0.947
Fabs.SC.f1:MOD3	-0.009	0.139	0.139	0.066	0.948

Cuadro 6.8: Promedio de los coeficientes con FULL AVERAGE.

un determinado *lag*, la correlación entre dos observaciones son iguales para cualquier par de observaciones del mismo sujeto.

Este es un modelo plausible para los datos corrientes, ya que se esperaría que la correlación entre observaciones disminuyera a medida que nos alejáramos en el tiempo hacia atrás. Es decir que si el sujeto comete un error al hacer una concordancia, esta influencia será mayor en la concordancia sucesiva que haga, respecto de las que vengan después de ésta.

El Cuadro 11 compara los modelos logísticos mixtos: (1) M.2: con efecto aleatorio de ordenada al origen; (2) M2.1: con efecto aleatorio de ordenada al origen y de FAM.LEX.f; (3) M2.2: agregando al anterior estructura de covarianza AR(1); (4) M2.3: el modelo (3) pero agregando $CUMRES^2$, para descartar un posible efecto de curvatura. Se detalla la verosimilitud \mathcal{L} de cada modelo («logLik»), su devianza ($-2\mathcal{L}(\text{modelo})$), junto con el p-valor arrojado («p.value») por la prueba de cociente de máxima verosimilitud [LRT] entre los modelos (1)-(2); (2)-(3); (3)-(4) (basado en una mezcla de χ^2 para evitar el problema de la prueba de hipótesis para el límite del espacio paramétrico, en el caso de las primeras dos comparaciones). Se observa que el modelo (3) resulta significativo estadísticamente; y que no existe efecto de curvatura. En lo que atañe a dicho modelo, el test de *Durbin-Watson* arroja: $W = 2,0118$; $p < 0,816$. El gráfico de autocorrelación (Figura 2) confirma que la correlación del lag 1 ha desaparecido, debido al uso de la estructura AR(1).

	X2	df	P	V	V.df
MOD:Fabs.SC.f	572.54	3	<0.0001	0.56	1
MOD:MORF.f	97.72	6	<0.0001	0.16	2
MOD:FAM.LEX.f	39.88	3	<0.0001	0.15	1
MOD:CUMRES.f	20.77	6	0.002	0.07	2
MOD:EST1	11.05	3	0.0115	0.08	1
MOD:EST5	7.32	3	0.0624	0.06	1
Fabs.SC.f:MORF.f	55.26	2	<0.0001	0.17	1
Fabs.SC.f:FAM.LEX.f	11.42	1	<0.001	0.08	1
Fabs.SC.f:CUMRES.f	4.48	2	0.1066	0.05	1
Fabs.SC.f:EST1	21.40	1	<0.0001	0.11	1
Fabs.SC.f:EST5	11.96	1	<0.001	0.08	1
MORF.f:FAM.LEX.f	18.69	2	<0.001	0.10	1
MORF.f:CUMRES.f	6.59	4	0.1592	0.04	2
MORF.f:EST1	186.51	2	<0.0001	0.32	1
MORF.f:EST5	580.65	2	<0.0001	0.56	1
FAM.LEX.f:CUMRES.f	2.70	2	0.2594	0.04	1
FAM.LEX.f:EST1	20.68	1	<0.0001	0.11	1
FAM.LEX.f:EST5	0.00	1	0.9927	0.00	1
CUMRES.f:EST1	8.74	2	0.0126	0.07	1
CUMRES.f:EST5	7.98	2	0.0185	0.07	1
EST1:EST5	103.20	1	<0.0001	0.24	1

Cuadro 6.9: X2: estadístico X2; df: grados de libertad de X2; P: p-valor; V: V de Cramer; V.df: grados de libertad de V.

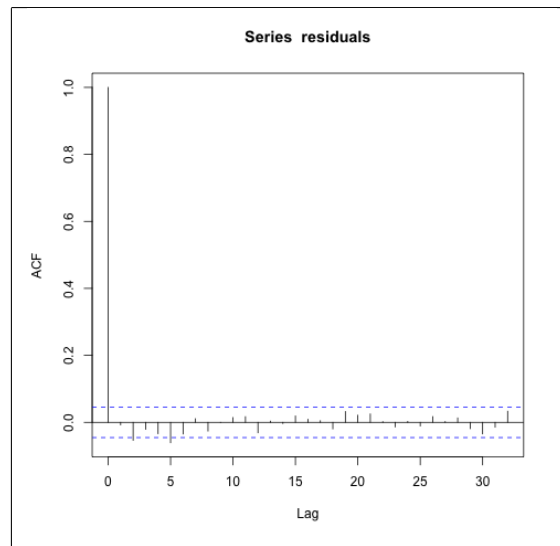


Figura 6.2: Función de Auto-correlación para el modelo logístico mixto con estructura de covarianza AR(1).

6.5.3. Modelo final: chequeo de supuestos y resultados.

El modelo condicional general (de covarianza) para la concordancia i en el grupo j (el grupo está definido como *la sesión k* anidada en *el alumno g*) [$k = 1, \dots, 12(14)$; $g = 1, \dots, 4$; $j = 1, \dots, 52$; $i = 1, \dots, 1857$] es:

$$\text{Distribuciones: } y_{ij} | v_{0ij}, v_{1ij}, w_{ij} \sim \text{Binomial}(N_{ij}, \pi_{ij}), N = 1; v_{0ij} \sim N_q(\mathbf{0}, \mathbf{G}_0); v_{1ij} \sim N_q(\mathbf{0}, \mathbf{G}), \\ q = 1, \dots, 52; \mathbf{w}_i \sim \mathbf{N}_{n_i}(\mathbf{0}, \mathbf{R}_i), \text{ donde: } w_i = w_{j1}, \dots, w_{jn_i} (n_i = 4, \dots, 81).$$

	GVIF	Df
MOD	1.48	3
CUMRES	1.01	1
MORF.f	1.50	2
FAM.LEX.f	1.05	1
Fabs.SC.f	1.41	1
EST1	1.10	1
EST5	1.34	1

Cuadro 6.10: GVIF: índice de inflación de la varianza; Df: grados de libertad de la variable (niveles menos uno)

	term	df	AIC	BIC	logLik	deviance	statistic	Chi.Df	p.value	p.corrected
1	m.2	12.000	2005.143	2071.464	-990.572	1981.143				
2	m.2.1	14.000	2001.946	2079.320	-986.973	1973.946	7.197	2.000	0.027	0.017
3	m.2.2	16.000	1993.504	2081.932	-980.752	1961.504	12.442	2.000	0.002	0.001
4	m.2.3	17.000	1995.090	2089.044	-980.545	1961.090	0.415	1.000	0.520	0.520

Cuadro 6.11: Selección de modelos: Efectos aleatorios. logLik: verosimilitud del modelo, AICc: AKAIKE corregido, delta: delta de AKAIKE, weight: pesos de AKAIKE, p.value: p-valor del LRT, p.corrected: p-valor corregido del LRT.

Matrices de varianza: $G_0 = \sigma_{v_0}^2 A$; $G_1 = \sigma_{v_1}^2 A$; $A_{UNR} = \begin{bmatrix} \sigma_{v_0}^2 & cov(v_0, v_1) \\ cov(v_0, v_1) & \sigma_{v_1}^2 \end{bmatrix}$ y además

:

$$R_i = \sigma_{\omega}^2 \begin{bmatrix} 1 & \rho & \rho^2 & \dots & \rho^{t-1} \\ \rho & 1 & \rho & \dots & \rho^{t-2} \\ \rho^2 & \rho & 1 & \dots & \rho^{t-3} \\ \dots & \dots & \dots & \dots & \dots \\ \rho^{t-1} & \rho^{t-2} & \rho^{t-3} & \dots & 1 \end{bmatrix}$$

Función de enlace: $logit = g(E[y_{ij}|v_{0ij}, v_{1ij}, w_{ij}]) = \log \left[\frac{\pi_{ij}}{1-\pi_{ij}} \right]$

Función inversa: $\pi_{ij} = g^{-1}(X\beta + Z_v v + Z_{\omega} w)$

Predictor lineal: $\eta_{ij} = \varphi_{ij} + v_{0ij} + v_{1ij} + w_{ij}$. Donde: φ_{ij} representa los efectos fijos; v_{0ij} y v_{1ij} son los efectos aleatorios «entre»; w_{ij} son los efectos aleatorios «dentro».

Modelo teórico:

$$g(E[y_{ij}|v_{0ij}, v_{1ij}, w_{ij}]) = \log \left[\frac{\pi_{ij}}{1-\pi_{ij}} \right] = (\beta_0 + v_{0i}) + (\beta_1 + v_{1i}) FAM.LEX.f + w_{ij} + \beta_2 MOD_1 + \beta_3 MOD_2 + \beta_4 MOD_3 + \beta_5 CUMRES + \beta_6 Fabs.SC.f + \beta_7 MORF.f_1 + \beta_8 MORF.f_2 + \beta_9 EST1 + \beta_{10} EST5$$

Los efectos aleatorios (condicionales) son: (i) v_{0i} : la desviación del grupo i de la ordenada al origen; (ii) v_{1i} : la desviación del grupo i de la media marginal de FAM.LEX.f; (iii) w_{ij} : efectos aleatorios «dentro» de cada grupo, con $w^T = [w_{j1}, w_{j2}, \dots, w_{jn_i}]$. Los efectos fijos (marginales) son: (i) β_0 es la media basal marginal; (ii) β_1 es el efecto de FAM.LEX.f; (iii) β_2 es el efecto de MOD, nivel 1; (iv) β_3 es el efecto de MOD, nivel 2; (v) β_4 es el efecto de MOD, nivel 3 [referencia de MOD1 es el nivel «0»]; (vi) β_5 es el efecto de CUMRES; (vii) β_6 es el efecto de Fabs.SC.f; (viii) β_7 es el efecto de MORF.f, nivel 1; (ix) β_8 es el efecto de MORF.f, nivel 2 [referencia de MORF.f es «0»]; (x) β_9 es el efecto de EST1; (xi) β_{10} es el efecto de EST5.

La matriz de varianza-covarianza de los efectos aleatorios «entre» sigue una estructura irrestricta; en cambio, aquella de los efectos aleatorios «dentro» es auto-regresiva de orden 1.

El Cuadro que sigue muestra los residuales de Pearson para el modelo elegido (sin datos agrupados). Un 2% de los residuos son mayores a ± 2 desvíos típicos, se espera que hasta un 5% de ellos superen dicha barrera.

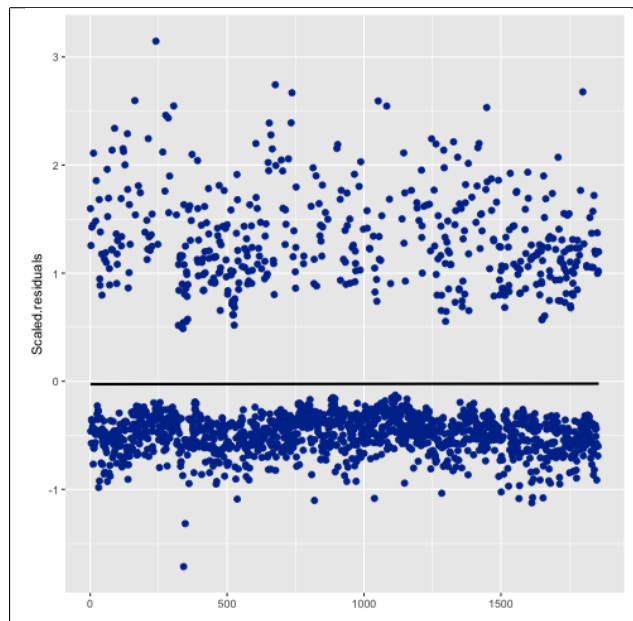


Figura 6.3: Residuos de Pearson.

En los que respecta a los supuestos distribucionales de los factores aleatorios, se aplicó el test de *Shapiro-Wilks* para detectar desvíos groseros del supuesto de normalidad. En el caso de los factores aleatorios de ordenada al origen no se detectó un apartamiento del supuesto ($W = 0,972$, $p = 0,275$). En lo que respecta a los factores aleatorios de FAM.LEX.f, el test sí rechaza el supuesto ($W = 0,953$, $p = 0,04$). Las gráficas que siguen muestran el *qq-plot* para examinar gráficamente el supuesto de normalidad de los factores aleatorios «entre». Cuánto más cerca de la línea central de igualdad entre los cuantiles empíricos y los teóricos de una distribución normal, mejor se sigue el supuesto de normalidad. Se verifican los resultados del test: los factores aleatorios de ordenada al origen se apartan menos del supuesto de normalidad que los de FAM.LEX.f.

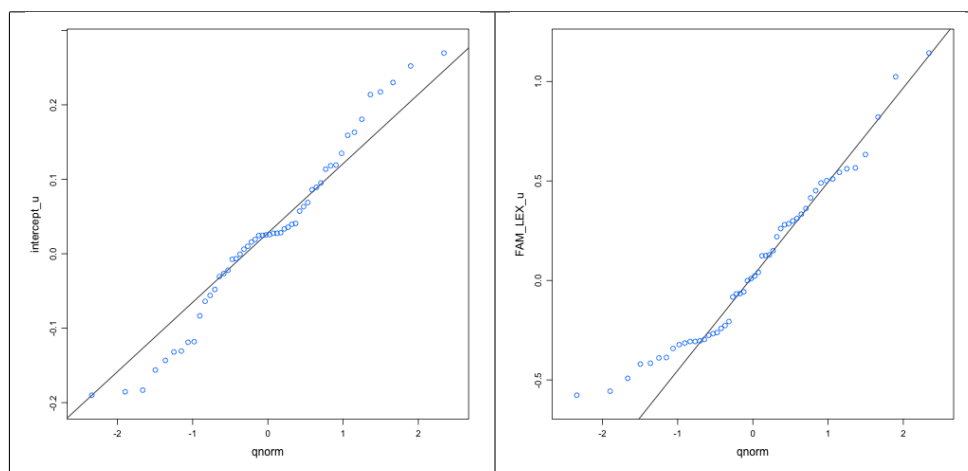


Figura 6.4: Q-Q Plot. Izquierda: efectos al. de ordenada al origen; Derecha: efectos al. de FAM.LEX.f

El Cuadro 12 muestra los resultados del ajuste.
Se reporta el modelo ajustado:

	Estimate	SE	z value	P(> z)	OR	IC_OR Lo	IC_OR Up
MOD1	-0.451	0.533	-0.847	0.397	0.637	0.224	1.81
MOD2	0.575	0.164	3.511	<0.001	1.777	1.289	2.449
MOD3	0.397	0.185	2.14	0.032	1.487	1.034	2.138
CUMRES	0.129	0.105	1.224	0.221	1.137	0.926	1.398
MORF.f1	-0.959	0.228	-4.203	<0.001	0.383	0.245	0.599
MORF.f2	-0.582	0.262	-2.22	0.026	0.559	0.334	0.934
FAM.LEX.f1	-0.603	0.18	-3.341	0.001	0.547	0.384	0.779
Fabs.SC.f1	-0.447	0.159	-2.806	0.005	0.639	0.468	0.874
EST11	-0.55	0.157	-3.5	<0.001	0.577	0.424	0.785
EST51	-0.778	0.241	-3.224	0.001	0.46	0.286	0.737
Sd (Intercept).	0.259						
Sd (FAM.LEX.f1).	0.653						
Sd (w).	0.8						
phi (ar1).	0.66						

Cuadro 6.12: Modelo binario elegido. Estimate: beta estimado; SE: Error standard de beta estimado; z.value: valor del test de Wald; P(>|z|): p-valor; OR: Odds Ratio; IC OR Lo: Intervalo de confianza de OR, cota inferior; IC OR Up: Intervalo de confianza de OR, cota superior

$$g(E[y_{ij}|v_{0ij}, v_{1ij}, w_{ij}]) = \log \left[\frac{\hat{\pi}_{ij}}{1-\hat{\pi}_{ij}} \right] = (-0,28 + v_{0i}) + (-0,6 + v_{1i}) \text{FAM.LEX.f} + w_{ij} - 0,45\text{MOD}_1 + 0,57\text{MOD}_2 + 0,39\text{MOD}_3 + 0,12\text{CUMRES} - 0,44\text{Fabs.SC.f} - 0,95\text{MORF.f}_1 - 0,58\text{MORF.f}_2 - 0,55\text{EST1} - 0,77\text{EST5}$$

Los desvíos típicos estimados de los factores aleatorios son: $\hat{\sigma}_{v_{0ij}} = 0,25$; $\hat{\sigma}_{v_{1ij}} = 0,65$; $\hat{\sigma}_{w_{ij}} = 0,8$. O sea que hubo más variabilidad entre grupos respecto de FAM.LEX.f. El coeficiente de correlación estimado para R_i resultó $\hat{\rho}_{R_i} = 0,66$; la correlación estimada en la matriz A fue $\hat{\rho}_{CS} = 0,16$ ($\widehat{cov}(\sigma_{v_{0ij}}, \sigma_{v_{1ij}}) = 0,026$), o sea que casi no hubo correlación entre efectos aleatorios «entre» (grupos).

El coeficiente de ordenada al origen varía en términos del logit como $-0,28$ más / menos la desviación del efecto aleatorio «entre» de ordenada al origen de cada grupo. Análogamente, el efecto (coeficiente) de FAM.LEX.f varía en términos del logit como $-0,6$ más / menos la desviación del efecto «entre» de FAM.LEX.f de cada grupo. El siguiente es un «forest plot» que indica la variación de los efectos aleatorios «entre» respecto de un individuo típico, es decir con $v_{0ij} = 0$ o $v_{1ij} = 0$. En términos del intercepto, los grupos (sesiones) que más se alejan de $v_{0ij} = 0$ son: 4.7, 4.11, 2.8, 2.6, 2.2 (arriba de la media marginal), 4.1, 3.3, 3.12 (debajo de la media marginal). En cambio, en términos del FAM:LEX.f, los grupos que más se alejan de $v_{1ij} = 0$ son: 4.7, 3.11, 3.12 (debajo de la media marginal), 2.1, 2.2, 2.8 (arriba de la media marginal). Por ejemplo, para β_0 , en los grupos que se ubican arriba de la media se cometen más errores de concordancia; y en los que están debajo, menos. Por ejemplo en el grupo más alejado hacia «arriba» de β_0 , el 4.11, los errores llegan al 43%; y en el más alejado hacia «abajo» (grupo 3.12), el 3%. Por otra parte, cuando está presente FAM.LEX.f, la *chance* de cometer un error de concordancia frente a no cometerlo *aumenta* en los grupos «hacia arriba» y *disminuye* en los que están «hacia abajo».

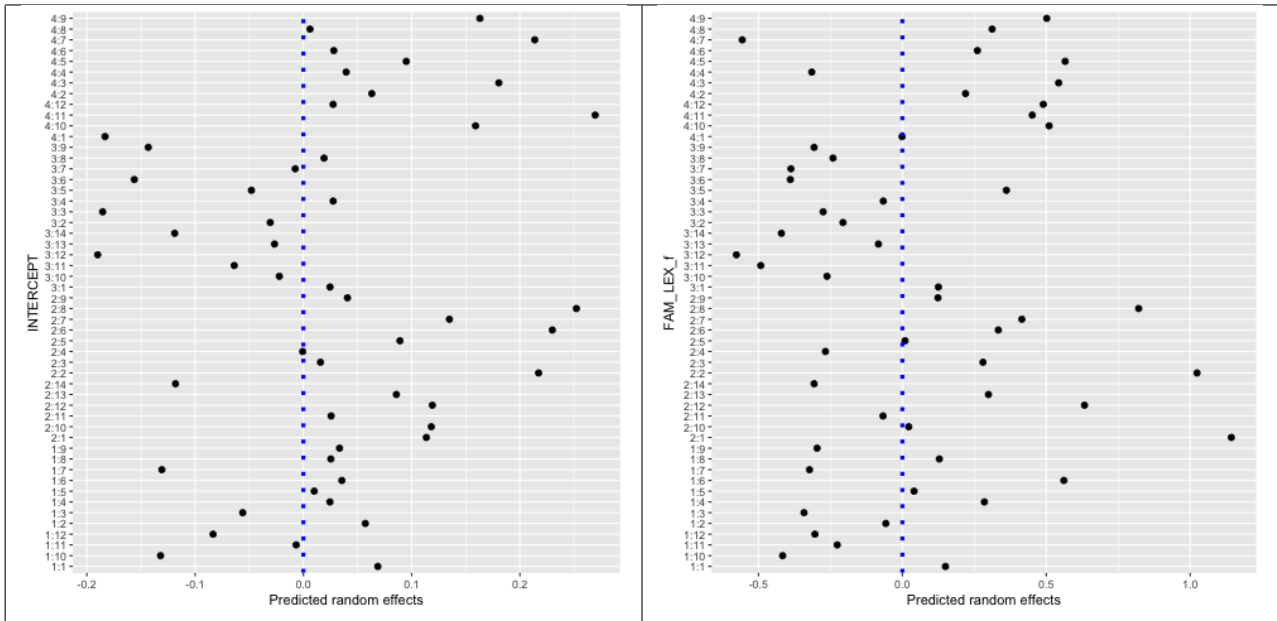


Figura 6.5: Forest Plot: Izq: efectos aleatorios «entre» de intercepto; Dcha: efectos aleatorios «entre» de FAM.LEX.f

La gráfica que sigue muestra una «trayectoria» de los factores aleatorios de ordenada al origen para cada sujeto. La línea azul es $v_{0ij} = 0$; la media de error entre los grupos (sesiones): β_0 . JAKO, con el mejor desempeño, posee la mayoría de los puntos debajo de la media; SONIA, de desempeño intermedio, tiene los efectos aleatorios por arriba (pero cerca de la media) y por debajo; por último, MIRKA y NATI ubican sus efectos aleatorios alejándose de la media. Nótese que MIRKA comenzó muy bien, muy por debajo de $v_{0ij} = 0$, o sea que podría estar al nivel de JAKO. NATI, por su parte, termina muy bien, con las última sesiones cerca de la media, un posible indicio de cambio de régimen en la adquisición.

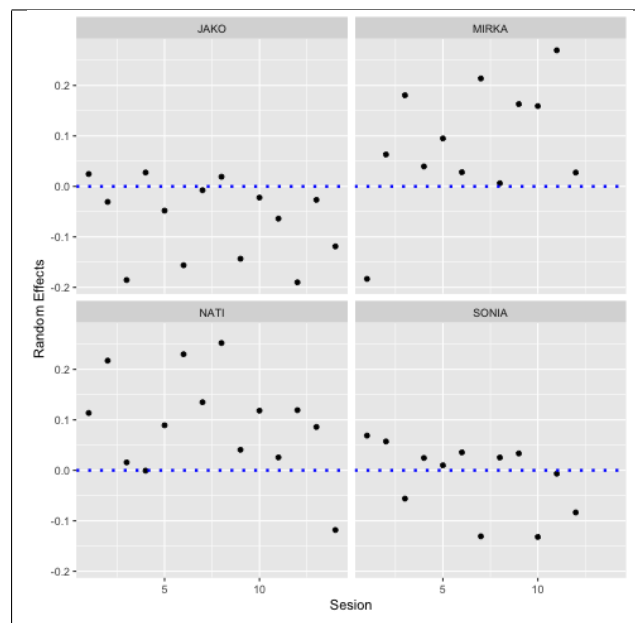


Figura 6.6: Trayectoria de efectos aleatorios de ordenada al origen, según sujeto.

El gráfico que sigue muestra los OR ($exp(\beta_i)$) para los coeficientes. Se observa que la chance de cometer un error de concordancia ($Y = 1$) respecto de no cometerlo ($Y = 0$) *augmenta* cuando se pasa de la categoría de referencia a la k -ésima categoría de la predictora para⁷: MOD(2) (78% más), MOD(3) (49% más). O sea, que dichas predictoras constituyen «factores de riesgo» de cometer un error. Por otra parte, la chance de

⁷Calculado como: $[exp(\beta_i) - 1] \times 100$.

cometer un error de concordancia ($Y = 1$) respecto de no cometerlo ($Y = 0$) *disminuye* cuando se pasa de la categoría de referencia a la k -ésima categoría de la predictora para: FAM.LEX (45% menos), MORF.f(1) (62% menos), MORF.f(2) (0.44% menos), Fabs.SC.f (36% menos), EST5 (54% menos), EST1 (42% menos). O sea, que dichas predictoras constituyen «factores de protección» contra cometer un error. Los Intervalos de Confianza (I.C.) para OR de MOD(1) y CUMRES contienen al «1», por lo tanto no son estadísticamente significativos.

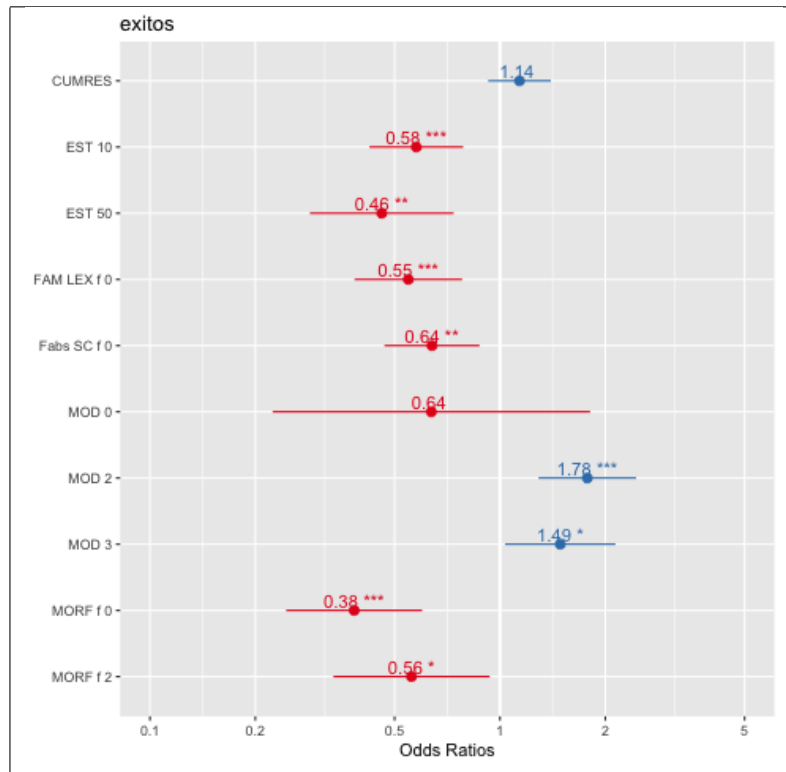


Figura 6.7: OR de los coeficientes. Azul: factores de riesgo; Rojo: factores de protección.

El Cuadro de a continuación muestra los intervalos de confianza para los contrastes entre los niveles de los efectos simples (más de dos niveles)⁸. Los intervalos que contienen al 0 no son significativos. En el caso de MOD resultó significativo el contraste MOD(0) - MOD(2) [$\widehat{logit}_{DIFF} = -0,574$; $p = 0,002$]; sin embargo MOD(0)-MOD(3) no resultó significativo con el ajuste del p-valor [$\widehat{logit}_{DIFF} = -0,396$; $p = 0,14$]. Para MORF.f, fue significativo MORF(0)-MORF(1) [$\widehat{logit}_{DIFF} = 0,955$; $p = 0,0001$] y no lo fue MORF(0)-MORF(2) [$\widehat{logit}_{DIFF} = 0,582$; $p = 0,067$]. En el apéndice se muestran los valores sobre los que se crearon las gráficas, los p-valores están ajustados por el método de *Tukey*.

⁸Por ejemplo, para la predictora MOD(3), el contraste es la diferencia entre logits: $\hat{g}(MOD(3) = 0) - \hat{g}(MOD(3) = 1) = (\beta_0 + \beta_1(0)) - (\beta_0 + \beta_1(1)) = \beta_0 + 0 - \beta_0 - 0,43 = 0 - 0,39 = -0,39$; donde \hat{g} es el logit estimado. Por esta razón, las estimaciones de los contrastes para los efectos simples son las mismas que para el Cuadro 8 con los datos del ajuste, pero con el signo cambiado. Para el contraste, resto de las predictoras se hallan fijas en el valor de referencia $x_i = 0$. Los p-valores están ajustados debido a las comparaciones múltiples.

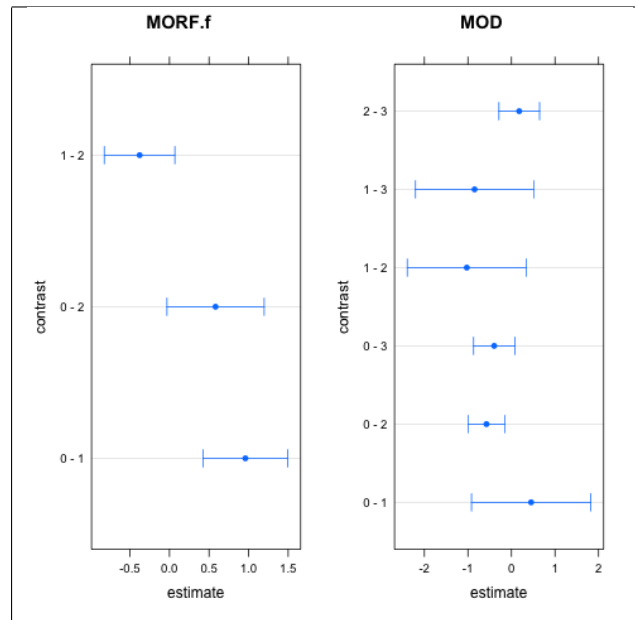


Figura 6.8: Contrastes para efectos simples.

6.6. Análisis de errores.

Se indicarán, expresadas en porcentajes, las probabilidades condicionales de error dado el nivel del predictor, por ejemplo: $p(RES.BIN = 1 | MOD = 3) = \frac{\#(RES.BIN=1 \cap MOD=3)}{\#MOD=3}$ es la probabilidad de cometer error dado que la concordancia tiene un adjetivo.

La chance de cometer un error de concordancia ($Y = 1$) respecto de no cometerlo ($Y = 0$) *aumenta* cuando:

1. Se pasa de concordancias con artículo definido a aquellas con determinantes. Si hay artículo los errores representan el 19%; en cambio si hay determinante es el 31%. En cuanto al tipo de determinante, el más frecuente es «mucho»: *muchos tiendas* [NATI, sesión 8, línea 32]; *mucha motivaciones* [MIRKA, sesión 9, línea 185]. Le sigue «todo»: *todo esos conocimiento* [MIRKA, sesión 12, línea 124]; *toda personas* [JAKO, sesión 2, línea 64]; todos ideas [NATI, sesión 12, línea 129]. Luego vienen los posesivos de 3ra. persona : *su aspecto* [JAKO, sesión 13, línea 125]; *todo mi amigos* [MIRKA, sesión 1, línea 250]. A continuación, los demostrativos: *esta historias* [NATI, sesión 8, línea 32]; *este personas* [NATI, sesión 10, línea 200]; *este chicos* [MIRKA, sesión 3, línea 227]. Y en menor medida aparecen los lemas: «otro», «poco», «mismo», «tanto».

La chance de cometer un error de concordancia ($Y = 1$) respecto de no cometerlo ($Y = 0$) *disminuye* cuando:

1. Se pasa de concordancias sin controlador con *baja* familiaridad / frecuencia léxica a aquellas con controladores de *alta* familiaridad / frecuencia léxica. Si la familiaridad / frecuencia es baja, los errores representan el 27%; cuando es alta, son el 23%. O sea que «alto» es «factor de protección» pero «bajo» es «factor de riesgo». Ejemplos de este segundo caso son: *las tasas* (it. *le tasse*, esp. *los impuestos*) [JAKO, sesión 9, línea 35]; *los bolsillo* [MIRKA, sesión 11, línea 77].
2. Se pasa de concordancias de TYPE infrecuente a aquellas de TYPE frecuente. En el primer caso el error es del orden del 30%; en el segundo, del 19%. Por ejemplo, para [Fabs.SC.f= 0]: *les primeros veces* (TYPE: *l-j-n-as-as-*es*) [SONIA, sesión 4, línea 137]; *católicos moralistes* (TYPE: *n-j-os-as*) [NATI, sesión 6, línea 232]; *los costumbres italianos* (TYPE: *l-n-j-as-es-as*) [JAKO, sesión 11, línea 51]; *todo mi amigos* (TYPE: *d-d-n-os-is-os*) [MIRKA, sesión 1, línea 250]. Con TYPE infrecuente el 54% de los casos involucra concordancias de más de dos términos.
3. Se pasa de concordancias en donde no se puede aplicar la estrategia «5», a aquellas donde es posible, o sea que si la palabra singular del italiano termina en *-e*, basta agregar *-s* a la desinencia en italiano (it.

sole > esp. *soles*). Si no se puede aplicar dicha estrategia, los errores representan el 26 %, en cambio, si se puede, representan el 19 %. Es interesante notar que, en contexto de aplicación de la estrategia, la mayoría de los errores no se deben al controlador: *mucho istituzioni* (it. sg. *istituzione*) [SONIA, sesión 4, línea 104]; los olores (it. sg. *odore*) [JAKO, sesión 1, línea 11]. Además, cuando el error sí está en el controlador, este coincide con el singular del italiano: *todos cantautore* (it. sg. *cantautore*) [MIRKA, sesión 10, línea 183]; los profesore (it. sg. *professore*) [SONIA, sesión 4, línea 104]. Estos datos parecen indicar que los hablantes probablemente tomen el singular del italiano para formar el plural del español, con palabras de raíces similares en ambas lenguas; sobrevolando la dificultad en español de tener que agregar -e- epentética (*profesor* > *profesor-e-s*).

4. Se pasa de concordancias en donde no se puede aplicar la estrategia «1», a aquellas donde es posible: si el italiano plural termina en *-i* el plural en español termina en *-os*. Esta constituye una estrategia muy común de los aprendientes italianos. Cuando se puede aplicar, los errores son del orden del 20 %; sino son del orden del 27 %. Por ejemplo: *barrios amplio* (it. pl. *grandi quartieri*) [SONIA, sesión 5, línea 241]; *su ministros* (it. pl. *i suoi ministri*) [NATI, sesión 6, línea 187]; *sus acuerdo* (it. pl. *i suoi accordi*) [JAKO, sesión 4, línea 26]; *alimentos fundamental* (it. pl. *cibi fondamentali*) [MIRKA, sesión 3, línea 144].
5. Se pasa de concordancias con terminaciones muy similares (es decir, con distancia baja en términos de operaciones de *la distancia de Levenstein*) entre español e italiano; a aquellas con terminaciones de similaridad media (distancias media). Si la similaridad es alta el error trepa al 42 %; si es media o baja el error disminuye al 23 % y 25 % respectivamente. Por tanto en la dirección *alta* > *media* / *baja*, el predictor «protege» contra errores, pero a la inversa, con *media* / *baja* > *alta*, se está más a riesgo de cometerlos. Esto va en contra del efecto esperado. Examinando los ejemplos del corpus de distancia baja (similaridad alta) con errores de concordancia, se observa, por ejemplo: *calles grandas* (it. pl. *strad-e* > esp. pl. *call-es*; it. pl. *grand-i* > esp. pl. *grand-e-s*; MORF= 2.6) [SONIA, sesión 4, línea 60]; *las investigacione* (it. pl. *l-e*; esp. pl. *l-a-s*; it. pl. *investigazion-i* > esp. pl. *investigacion-e-s*; MORF = 2.8) [NATI, sesión 9, línea 16]; *las mujeres* (it. pl. *l-e*; esp. pl. *l-a-s*; it. pl. *donn-e* > *mujer-e-s*; MORF= 2.8) [NATI, sesión 14, línea 236]; *tu amigas* (it. pl. *tu-e* > esp. pl. *tu-s*; it. pl. *amich-e* > esp. pl. *amiga-s*; MORF = 2.2) [MIRKA, sesión 2, línea 138]. En todos los casos las operaciones de edición son pocas, por ejemplo, en *strade* > *calles* (*e* > *es*) basta insertar «s» para pasar de la terminación en italiano a aquella en español; en otras se requiere además una sustitución, como en: *amiche* > *amigas* (*e* > *as*), pero la suma de las distancias en ambos términos no supera el valor 2.8. El 54 % (37) de los 68 casos de distancia baja de MORF.f con error contienen determinantes (MOD = 2) y controladores no familiares y/o no frecuentes en un 42 %; además de TYPE infrecuente en un 82 %; todos factores de riesgo. En cambio, los factores de protección EST1, EST5 representan solo el 33 % y 5 % respectivamente. No obstante, el OR de MORF.f ya está ajustado por la posible acción de estos confundidores⁹; y el efecto es significativo igualmente. Muchos de los determinantes eran posesivos de 3ra. persona plural («sus») [18/37] y la mayoría contenían errores. En italiano la construcción de posesivo cambia según la referencia sea: (i) a la 3ra. persona singular: *le sue case* («sus casas» [de el / ella]); (ii) a la 3ra. persona plural: *le loro case* («sus casas» [de ellos/ellas]). Otra posibilidad es el hecho de que veinte de los 37 determinantes modificaran controladores masculinos. En *i tuoi amici* «tus amigos», la distancia entre *tuoi* > *tus* requiere una operación de borrado y otra de sustitución. En cambio, el pasaje entre *amici* > *amigos* requiere de una sustitución más una inserción. Recuérdese que los pesos de las operaciones fueron: DELETE = 0.3; SUBSTITUTE = 0.6; INSERT = 1. Es decir que se consideró que el orden de costo de las operaciones fue DELETE < SUBSTITUTE < INSERT. Quizás «borrado» + «inserción» requiere más «costo cognitivo»; y, por ello, lo que en hipótesis parecía ser más fácil para los cuatro sujetos italianos terminó siendo en realidad más difícil. Por otra parte, entre los casos también aparecían siete plurales invariantes en italiano («moto», «radio», «video», «foto»), en estos casos los errores fueron, en su mayoría, en el controlador: *las foto* [MIRKA, sesión 7, línea 95], *los vídeo* [MIRKA, sesión 11, línea 161]. Asimismo se observaron

⁹Un confundidor es un predictor que se halla asociado a la vez con la respuesta y con otra predictora de interés. Con lo cual el efecto de esta variable sobre la respuesta queda «confundido» con el efecto del predictor.

Predictor	1-0	2-0	3-0	Comentario
MOD	↑ NO	↑ SI	↑ NO	
Fabs.SC.f	↓ SI	-	-	
MORF.f	↑ NO	↑↑NO	-	dirección opuesta
CUMRES	↑NO	↑↑NO	-	
FAM.LEX.f	↓SI	-	-	
EST1	↓SI	-	-	
EST5	↓SI	-	-	

Cuadro 6.13: Expectativas cumplidas del análisis. Expectativas cumplidas del análisis. Leyenda: (↑) = efecto aumenta; (↓) = efecto disminuye; (↑↑/↓↓) = mayor tamaño de efecto; NO = no se cumple expectativa (no significativo); SI = se cumple expectativa (significativo).

dos instancias del determinante «diferente»: *diferente personas* [JAKO, sesión 2, línea 59], *diferente postaciones* (esp. *ubicaciones*) [JAKO, sesión 8, línea 48]. En italiano «diverso» tiene cuatro formas: *divers-o Masc. Sg. / -a Fem. Sg. / -i Masc. Pl. / -e Fem. Pl.* Nótese que el singular del español coincide con la desinencia del plural femenino en italiano. En los ejemplos, los controladores son ambos femeninos. Por otro lado hay dos usos del numeral *quattrocento*, que en italiano no lleva concordancia: *quattrocento kilómetros* [JAKO, sesión 10, línea 71] y *quattrocento habitantes* [JAKO, sesión 12, línea 15]. Mírese ahora los ejemplos de distancia «media» con errores, por ejemplo: *los veneciano* (it. pl. **-i** > esp. pl. *l-o-s*; it. pl. *venezian-i* > esp. pl. *venecia-os*; MORF= 3.2) [SONIA, sesión 5, línea 293]; *personas motivada* (it. pl. *person-e*; esp. pl. *person-a-s*; it. pl. *motivati-e* > esp. pl. *motivad-a-s*; MORF = 3.2) [NATI, sesión 8, línea 391]; *lo viajes* (it. pl. **-i**; esp. pl. *l-o-s*; it. pl. *viaggi-i* > *viaj-e-s*; MORF= 3.2) [JAKO, sesión 1, línea 75]; *las costumbre* (it. pl. *l-e* > esp. pl. *l-a-s*; it. pl. *abitudine-i* > esp. pl. *costumbre-s*; MORF = 3.2) [MIRKA, sesión 4, línea 16]. La mayoría de los casos de error con [MORF.f = 1] requieren de sustituciones e inserciones, por ejemplo en *viaggi* > *viajes* basta sustituir «i» por «e» e insertar «s». Quizás no tener que usar la operación «borrar» haya hecho bajar la chance de error.

6.7. Conclusiones.

El Cuadro siguiente ilustra si se cumplieron las expectativas del estudio, para las predictoras del modelo elegido. En el caso de MOD, solo resultaron significativos los determinantes para facilitar el error. En cuanto a MORF.f, se verificó la dirección opuesta a la esperada: distancias cortas facilitaron el error y medias lo inhibieron (distancias largas quedaron al borde de significancia estadística). Como se esperaba fueron inhibidores de error: la alta frecuencia de TYPE, EST1, EST5 y los controladores familiares y/o de alta frecuencia. No hubo efecto de los errores acumulados. Se podría plantear la hipótesis, a confirmar con un experimento aleatorizado, de que los italianos aplican la estrategia de usar «-os» en español cuando la palabra plural del italiano termina en *-i*; y de agregar *-s* en español a las palabras que en el singular del italiano terminan en *-e* (*canzone* > *canciones*), lo cual ayuda a circunvalar el problema de agregar *-e*-epentética en español. La predictora GRUPO6 no influyó en la chance de cometer un error¹⁰, por lo cual el *clustering* no resultó efectivo.

¹⁰También se testeó el mismo modelo ajustado pero incluyendo solamente a GRUPO6. Ningún contraste resultó significativo (ajuste de p-valor según el método de *Dunnet*).

6.8. Apéndice de capítulo.

	cut	total	obs0	obs1	exp0	exp1
1	[0.0651,0.106]	85.00	79.00	6.00	77.08	7.92
2	(0.106,0.112]	47.00	39.00	8.00	41.80	5.20
3	(0.112,0.126]	63.00	54.00	9.00	55.34	7.66
4	(0.126,0.137]	66.00	54.00	12.00	57.18	8.82
5	(0.137,0.153]	63.00	52.00	11.00	53.86	9.14
6	(0.153,0.16]	61.00	53.00	8.00	51.44	9.56
7	(0.16,0.168]	70.00	58.00	12.00	58.58	11.42
8	(0.168,0.175]	58.00	46.00	12.00	48.06	9.94
9	(0.175,0.187]	72.00	65.00	7.00	58.86	13.14
10	(0.187,0.196]	56.00	47.00	9.00	45.26	10.74
11	(0.196,0.204]	72.00	62.00	10.00	57.53	14.47
12	(0.204,0.217]	65.00	53.00	12.00	51.21	13.79
13	(0.217,0.227]	55.00	39.00	16.00	42.75	12.25
14	(0.227,0.238]	64.00	49.00	15.00	49.07	14.93
15	(0.238,0.248]	64.00	51.00	13.00	48.48	15.52
16	(0.248,0.257]	64.00	50.00	14.00	47.80	16.20
17	(0.257,0.268]	64.00	50.00	14.00	47.21	16.79
18	(0.268,0.275]	66.00	43.00	23.00	48.05	17.95
19	(0.275,0.288]	62.00	45.00	17.00	44.55	17.45
20	(0.288,0.296]	66.00	48.00	18.00	46.65	19.35
21	(0.296,0.308]	63.00	38.00	25.00	43.98	19.02
22	(0.308,0.323]	64.00	45.00	19.00	43.79	20.21
23	(0.323,0.337]	64.00	48.00	16.00	42.89	21.11
24	(0.337,0.358]	65.00	42.00	23.00	42.34	22.66
25	(0.358,0.381]	62.00	41.00	21.00	39.18	22.82
26	(0.381,0.407]	65.00	30.00	35.00	39.41	25.59
27	(0.407,0.44]	64.00	38.00	26.00	36.80	27.20
28	(0.44,0.471]	63.00	33.00	30.00	34.32	28.68
29	(0.471,0.665]	64.00	32.00	32.00	30.53	33.47

Cuadro 6.14: Test de Hosmer Lemeshow. cut: cortes de probabilidad que definen el grupo, total: casos totales por grupo, obs0: casos con $Y=0$, obs1 = casos con $Y=1$; exp0: suma de los predichos para $Y=0$, exp1: suma de los predichos para $Y=1$

contrast	estimate	SE	df	z.ratio	p.value
0 - 1	0.9591	0.2282		4.203	0.0001
0 - 2	0.5820	0.2622		2.220	0.0679
1 - 2	-0.3770	0.1901		-1.983	0.1165

Results are averaged over the levels of: MOD, FAM.LEX.f, Fabs.SC.f, EST1, EST5

P value adjustment: tukey method for comparing a family of 3 estimates

Cuadro 6.15: efectos simples: MORF.f

contrast	estimate	SE	df	z.ratio	p.value
0 - 1	0.4510	0.5328		0.847	0.8321
0 - 2	-0.5747	0.1637		-3.511	0.0025
0 - 3	-0.3966	0.1853		-2.140	0.1404
1 - 2	-1.0257	0.5313		-1.931	0.2152
1 - 3	-0.8477	0.5290		-1.602	0.3772
2 - 3	0.1781	0.1821		0.978	0.7620

Results are averaged over the levels of: MORF.f, FAM.LEX.f, Fabs.SC.f, EST1, EST5

P value adjustment: tukey method for comparing a family of 4 estimates

Cuadro 6.16: efectos simples: MOD

	INSTANCIA	RES_CAT	ID	SESION	LINEA
43	vacacione agreables	3	1	4	10
52	calles grandas (1)	2	1	4	60
322	tu pies	3	2	1	78
335	su padres	3	2	1	218
350	mi compañeros	3	2	2	142
356	mis amigo	3	2	2	225
442	su ministros	3	2	6	187
459	la curas (1)	3	2	7	16
526	su ojos	3	2	8	263
534	las investigacione	3	2	9	16
565	el desfiles	3	2	10	65
626	mi compañeros	3	2	12	6
670	las mujeres	2	2	13	129
672	[mujeras] famosas <de los políticos>	2	2	13	129
698	las mujeres	2	2	14	236
750	diferente personas	3	3	2	59
780	tu manos	3	3	2	172
817	sus acuerdo	3	3	4	26
827	su productos	3	3	4	81
914	su condiciones	3	3	7	40
925	su pasiones (1)	3	3	7	150
934	su pensamientos	3	3	7	189
957	diferente postaciones	3	3	8	48
986	tu parámetros (2)	3	3	9	30
1034	cuatrociento kilómetros	3	3	10	71
1039	la [motos] (1)	3	3	10	179
1040	[motos] más cómoda (2)	3	3	10	179
1046	las moto (1)	3	3	10	216
1047	únicas moto (2)	3	3	10	216
1097	cuatrociento habitantes	3	3	12	15
1151	su aspectos (1)	3	3	13	125
1213	mi amigos (2)	3	4	1	250
1243	parte diferentes	3	4	1	425
1264	tu amigos	3	4	2	133
1267	tu amigas	3	4	2	138
1282	otras radio	3	4	2	249
1283	radio especializadas	3	4	2	255
1298	su tareas	3	4	3	50
1333	su libros	3	4	3	217
1358	costumbre diferentes	3	4	4	20
1383	[costumbres] diferente	3	4	4	183
1459	su instintos (1)	3	4	6	70
1475	su principios	3	4	6	197
1515	su razones	3	4	7	79
1516	las foto	3	4	7	95
1630	su juegos	3	4	9	109
1642	mis amigo (2)	3	4	9	168
1645	mucha motivaciones	3	4	9	185
1651	mi amigos	3	4	9	214
1660	su padres	3	4	9	248
1710	su textos	3	4	10	110
1734	su hijos	3	4	10	207
1764	los video	3	4	11	161

Cuadro 6.17: Instancias de distancia baja de MORF.f.

Capítulo 7

Análisis de respuesta categórica.

Se presenta el modelo mixto multinomial bayesiano para analizar la respuesta categórica y se establecen expectativas de análisis. Primero se lleva a cabo selección de modelos de acuerdo al paradigma de información con modelos multinomiales generalizados. Una vez seleccionadas las predictoras se comparan modelos bayesianos con tres estructuras de varianza diferentes (efectos aleatorios). Luego, se examinan los supuestos del modelo y se presentan los resultados. Por último se analizan los errores y se comparan resultados obtenidos con aquellos de la respuesta binaria.

7.1. Marco teórico.

Esta sección introduce los conceptos básicos de: (i) distribución multinomial, Test de χ^2 (Kateri, 2014; caps. 1 y 2); (ii) regresión logística multinomial (Hosmer, Lemeshow & Sturdivant, 2013; cap. 8); (iii) su versión mixta (Hedeker, 2006; cap. 11). Se introducen también las nociones pertinentes de estadística bayesiana para delinear el modelo logístico multinomial mixto desde dicha perspectiva (Gelman et al. ,2004; Sorensen & Gianola, 2002; cap.14).

7.1.1. Distribución multinomial.

Suponiendo un ensayo con $K \geq 2$ resultados, la probabilidad de cada resultado será π_k ; $k = 1, 2, \dots, K$; con la condición: $\sum_{i=1}^K \pi_k = 1$. Dichos resultados son los posibles niveles de una variable categórica aleatoria X con frecuencias (N_1, N_2, \dots, N_K) . Ya que $\sum_{i=1}^K N_k = n$, solamente $K - 1$ categorías resultan ser aleatorias. Entonces, la probabilidad de que ocurra una muestra de frecuencias *observadas* $n = (n_1, n_2, \dots, n_K)$; $\sum_{i=1}^K n_k = n$, será:

$$p(n_1, n_2, \dots, n_K) = P(N_1 = n_1, N_2 = n_2, \dots, N_K = n_K) = \left(\frac{n!}{n_1! n_2! \dots n_K!} \right) \pi_1^{n_1} \pi_2^{n_2} \dots \pi_K^{n_K}$$

En forma resumida, se dice que las frecuencias aleatorias de las categorías se distribuyen como multinomial de parámetros n y vector de probabilidades $\pi: (N_1, N_2, \dots, N_{K-1}) \sim \mathcal{M}(n; (\pi_1, \pi_2, \dots, \pi_{K-1}))$.

Para $K = 2$, la distribución multinomial se reduce a una binomial $\mathcal{B}(n, \pi_1)$. Además, cada categoría se distribuye marginalmente como $\mathcal{B}(n, \pi_k)$, $k = 1, 2, \dots, K$. Por lo tanto, la esperanza y la varianza de N_k , $k = 1, 2, \dots, K$, es: $E[N_k] = n\pi_k$ y $V[N_k] = n\pi_k(1 - \pi_k)$.

7.1.2. Test de chi-cuadrado.

Uno de los test más simples para encontrar una asociación entre dos variables aleatorias categóricas es el de χ^2 . Dicho test pone a prueba la hipótesis nula de independencia para las variables X (con I niveles) e Y (con J niveles): $P(X = i, Y = j) = P(X = i)P(Y = j)$; $i = 1, \dots, I, j = 1, \dots, J$. En una tabla de contingencia $I \times J$, la proporción total las filas de *cada* nivel de X forman un vector de probabilidades marginales $\pi_r = \pi_{1+}, \pi_{2+}, \dots, \pi_{I+}$, es decir, la distribución marginal de X (ignorando el nivel de Y). Análogamente el

vector de proporciones totales de las columnas forma la distribución marginal de Y (o sea, ignorando el nivel de X): $\pi_c = \pi_{+1}, \pi_{+2}, \dots, \pi_{+J}$. Por lo tanto, X e Y serán independientes si para cada combinación de niveles (celdas) se verifica: $\pi_{ij} = \pi_{i+} \pi_{+j}; i = 1, \dots, I, j = 1, \dots, J$. El número *esperado* de frecuencias será: $m_{ij} = n\pi_{i+} \pi_{+j}$ y su estimador (de máxima verosimilitud): $\hat{m}_{ij} = np_{i+} p_{+j} = n \frac{n_{i+} n_{+j}}{n^2} = \frac{n_{i+} n_{+j}}{n}; i = 1, \dots, I, j = 1, \dots, J$. El estadístico de Pearson para testear H_0 es:

$$X^2 = \sum_{i,j} \frac{(n_{ij} - \hat{m}_{ij})^2}{\hat{m}_{ij}}, \text{ cuya distribución asintótica bajo } H_0 \text{ es } \chi^2_{(I-1)(J-1)}. \text{ Son requisitos del test que el}$$

80% de los valores esperados sean mayores a 5 y ninguno sea menor a 1. Si no se cumplen, la aproximación a la distribución de χ^2 es incorrecta. Pero dicho test es global, es decir, rechaza o no la hipótesis de independencia sin especificar cuáles celdas de la tabla de contingencia se alejan más de la hipótesis nula. Para ello, es necesario examinar los residuos $e_{ij} = n_{ij} - \hat{m}_{ij}$: la diferencia entre las frecuencias estimadas y esperadas. Si los residuos *estandarizados* (e_{ij}^s) en valor absoluto son tales que $|e_{ij}^s| \geq 2$, entonces la celda es estadísticamente significativa al nivel $\alpha = 0,05$ ya que $|e_{ij}^s| > z_{0,025} = 1,96$ (asintóticamente, $e_{ij}^s \sim N(0, 1)$). Hay dos tipos de residuos a considerar:

(i) Residuos estandarizados de Pearson (corregidos):

$$e_{ij}^p = \frac{e_{ij}}{\sqrt{\hat{m}_{ij} \hat{v}_{ij}}}; \hat{v}_{ij} = \left(1 - \frac{n_{i+}}{n}\right) \left(1 - \frac{n_{+j}}{n}\right); i = 1, \dots, I, j = 1, \dots, J$$

(ii) Residuos de Devianza:

$$e_{ij}^d = \text{sign}(n_{ij} - \hat{m}_{ij}) \left[2n_{ij} \log \left(\frac{n_{ij}}{\hat{m}_{ij}} \right) \right]^{\frac{1}{2}}; i = 1, \dots, I, j = 1, \dots, J$$

7.1.3. Modelo logístico multinomial (mixto).

Para extender el modelo logístico binario al multinomial, resulta necesario establecer una categoría de referencia $Y = 0$ para la variable de respuesta categórica. Si la respuesta, para cada observación i , es $Y_i = c; c = 1, 2, \dots, C$ (la respuesta nominal para el i -ésimo individuo tiene el valor de la categoría c entre C categorías, sin contar la referencia) y su referencia $Y_i = 0$; se ajustarán C modelos logísticos binarios, cada uno con función de enlace *logit*, es decir, el logaritmo de la chance de que ocurra el evento $Y = c$ respecto de que no ocurra ($Y = 0$) dado un vector de valores de la covariable para dicho individuo. Formalmente:

$$g(E[Y_i | \mathbf{X} = \mathbf{x}_i] = \pi_i) = \text{logit}(\pi_i) = \log \left[\frac{\pi_i = P(Y = c | \mathbf{x}_i)}{1 - \pi_i = P(Y = 0 | \mathbf{x}_i)} \right] = \beta_{c0} + \beta_{c1}x_1 + \dots + \beta_{cp}x_p = \mathbf{x}_i^T \boldsymbol{\beta} = \eta_{ic}$$

para $c = 1, 2, \dots, C$ y donde η es el predictor lineal. En el caso de RES_CAT, $c = 1, 2, 3, 4$ y $Y = 0 = \text{referencia}$; por lo tanto, se ajustarán cuatro modelos. La probabilidad de que la respuesta del i -ésimo individuo tenga la categoría c dado el valor de las predictoras para dicho individuo es:

$$\pi_{ic}(\mathbf{x}_i) = P(Y = c | \mathbf{x}_i) = \frac{\exp\{\mathbf{x}_i^T \boldsymbol{\beta}\}}{1 + \sum_{k=1}^c \exp\{\mathbf{x}_i^T \boldsymbol{\beta}\}_k} = \frac{e^{\eta_{ic}}}{1 + e^{\eta_{i1}} + e^{\eta_{i2}} + \dots + e^{\eta_{ic}}} = \frac{e^{\eta_{ic}}}{1 + \sum_{k=1}^C e^{\eta_{ik}}}$$

Cada probabilidad es función de $C(p-1)$ parámetros $\boldsymbol{\beta}^T = [\boldsymbol{\beta}_1^T, \boldsymbol{\beta}_2^T, \dots, \boldsymbol{\beta}_C^T]$. Por ejemplo, si se ajustara un modelo con cuatro covariables se estimarían $(4)(4-1) = 12$ efectos fijos, y 4 coeficientes de ordenada al origen. Cada respuesta y_i es una variable aleatoria que se distribuye como multinomial $(N_1, N_2, \dots, N_{K-1}) \sim \mathcal{M}(n; (\pi_1, \pi_2, \dots, \pi_{K-1}))$, con: $K = 5$ y $\sum_{i=1}^K N_k = n; \sum_{i=1}^K \pi_k = 1$. Para construir la función de verosimilitud se transforma a la respuesta en una variable indicadora con $y_{im} = 1$ si $y_{im} = c$ o $y_{im} = 0$ sino ($m = 0, 1, 2, 3, 4$). Por ejemplo si $y_i = 0$, entonces: $y_{i0} = 1, y_{i1} = 0, y_{i2} = 0, y_{i3} = 0, y_{i4} = 0$. La función de verosimilitud es la siguiente (para N observaciones independientes):

$$\mathcal{L}(Y_i | \boldsymbol{\beta}) = \prod_{i=1}^N \prod_{m=0}^{K-1} \left(\frac{e^{\eta_{im}}}{\sum_{k=0}^{K-1} e^{\eta_{ik}}} \right)^{y_{im}} = \prod_{i=1}^N \prod_{m=0}^{K-1} (\pi_m(x_i)^{y_{im}}) = \prod_{i=1}^N [\pi_0(x_i)^{y_{i0}} \pi_1(x_i)^{y_{i1}} \pi_2(x_i)^{y_{i2}} \pi_3(x_i)^{y_{i3}} \pi_4(x_i)^{y_{i4}}]$$

Los coeficientes estimados se obtienen minimizando la función de *log-verosimilitud* respecto del vector de coeficientes β . Las ecuaciones normales se hallan tomando derivadas parciales de primer orden de $\mathcal{L}(\beta)$ con respecto a cada uno de los $(C-1)(p-1)$ parámetros. Con la matriz de derivadas segundas (evaluadas en $\hat{\beta}$) se obtiene la matriz de información, cuya inversa posee en la diagonal las varianzas, y fuera de ella a las covarianzas de $\hat{\beta}$ (ver: Hosmer, Lemeshow & Sturdivant (2013), p. 271).

El *Odds Ratio* se interpretará como en el caso binario pero contrastando la chance de $Y = c$ versus $Y = 0$ (referencia) para $x_j = a$ con respecto a $x_j = b$:

$$OR_c(x_j = a, x_j = b) = \frac{P(Y = c | x = a) / P(Y = 0 | x = a)}{P(Y = c | x = b) / P(Y = 0 | x = b)} = \exp\{\beta_{jc}\}$$

También los contrastes entre niveles de una predictora son como en el caso binomial pero teniendo en cuenta la respuesta $Y = c$; o sea: $\text{logit}(x_{c=2,1} = a) - \text{logit}(x_{c=1,1} = b) = \hat{\beta}_{c=2,1} - \hat{\beta}_{c=1,1}$; lo cual resulta equivalente a la diferencia entre el logit de $Y = 1$ versus el logit de $Y = 2$. La varianza de la diferencia entre coeficientes es: $\widehat{V}(\hat{\beta}_{c=2,1} - \hat{\beta}_{c=1,1}) = \widehat{V}(\hat{\beta}_{c=2,1}) + \widehat{V}(\hat{\beta}_{c=1,1}) - 2\widehat{Cov}(\hat{\beta}_{c=2,1}, \hat{\beta}_{c=1,1})$.

Ahora bien, el modelo mixto se obtiene agregando al predictor lineal los efectos aleatorios. A diferencia del caso binario, cada categoría $c = 1, 2, \dots, C$ tiene su propio efecto aleatorio de ordenada al origen, así como otros efectos aleatorios asociados a las predictoras. Se denota $Y_{ij} \in \{1, 2, \dots, C\}$ para representar a la categoría de la variable del individuo i anidado en el grupo j y con $Y_{ij} = 0$ a la categoría de referencia. Entonces se puede expresar la probabilidad de que la respuesta del i -ésimo individuo del grupo j -ésimo tenga la categoría c dado el valor de las predictoras para dicho individuo (y grupo), *condicionales* a los efectos aleatorios:

$$\pi_{ijc}(\mathbf{x}_{ij}) = P(Y = c | \theta = \mathbf{z}_{ij}) = \frac{\exp\{x_{ij}^T \beta + z_{ij}^T \mathbf{v}\}_c}{1 + \sum_{k=1}^C \exp\{x_{ij}^T \beta + z_{ij}^T \mathbf{v}\}_k} = \frac{e^{\eta_{ijc}}}{1 + e^{\eta_{ij1}} + e^{\eta_{ij2}} + \dots + e^{\eta_{ijC}}} = \frac{e^{\eta_{ijc}}}{1 + \sum_{k=1}^C e^{\eta_{ijk}}}$$

La función de verosimilitud es idéntica al caso no mixto, pero ahora en lugar de hacer la productoria sobre N individuos se la hace sobre los n_i individuos del j -ésimo grupo.

$$\mathcal{L}(Y_j | \theta = \mathbf{z}_{ij}) = \prod_{i=1}^{n_i} \prod_{m=0}^{K-1} \left(\frac{e^{\eta_{ijc}}}{\sum_{k=0}^{K-1} e^{\eta_{ijk}}} \right)^{y_{ijm}} = \prod_{i=1}^{n_i} [\pi_0(x_{ij})^{y_{ij0}} \pi_1(x_{ij})^{y_{ij1}} \pi_2(x_{ij})^{y_{ij2}} \pi_3(x_{ij})^{y_{ij3}} \pi_4(x_{ij})^{y_{ij4}}]$$

El *coeficiente de correlación intra clase* (ICC) mide el grado de correlación presente entre los individuos de un grupo. En términos de los individuos el modelo multinomial mixto con efecto aleatorio de ordenada al origen, más otros efectos aleatorios si los hubiere, puede expresarse como: $y_{ijc} = x_{ij}^T \beta_c + z_{ij0}^T \mathbf{v}_{0c} + \dots + z_{ijq}^T \mathbf{v}_{qc} + \varepsilon_{ijc}$. La varianza de los errores es $\sigma_\varepsilon^2 = \pi^2/3$; que es la varianza de la distribución logística estándar. Por lo tanto el ICC resulta de comparar la varianza del efecto aleatorio *de ordenada al origen* con la varianza total (la suma de las varianzas de los efectos aleatorios más la del error): $r_c = \sigma_{0c}^2 / (\sigma_{0c}^2 + \dots + \sigma_{qc}^2 + \pi^2/3)$. Recuérdese que: $\mathbf{v}_{qc} \sim N(0, \sigma_{qc}^2)$.

7.1.4. Modelo multinomial mixto bayesiano.

Considérese el problema de inferencia sobre un parámetro θ cualquiera. En estadística frecuentista se considera que es un valor fijo poblacional, aunque desconocido, que debe estimarse. Esta estimación da como resultado *un número*, con una varianza asociada que da la incertidumbre en dicha estimación. La pregunta de inferencia es: ¿Cuál es el valor poblacional de θ ? En cambio, desde la perspectiva bayesiana los parámetros se consideran variables aleatorias y, como tales, llevan asociada una distribución, que a su vez, depende de un parámetro genérico ψ . Si se quiere hacer inferencia, la pregunta bayesiana sería: ¿cuáles son los valores del parámetro θ , que explican los datos y (que están fijos)? Es decir que la respuesta es una *distribución* de dichos valores, que representa la incerteza en la estimación del parámetro θ . A

esta distribución se la denomina «posterior» y se denota: $p(\theta | y)$. Asigna diferente grado de probabilidad («creencia») a los posibles valores del parámetro θ luego de ver la evidencia. El problema de inferencia se resuelve mediante el teorema de Bayes, que reza:

$$p(\theta | y) = \frac{p(y | \theta) p(\theta | \psi)}{p(y)}$$

donde: (i) $p(y | \theta)$ es la verosimilitud de los datos y (o sea, la distribución de y variando los valores del parámetro y dejando fijo los datos), que depende de parámetro genérico θ ; (ii) $p(\theta | \psi)$ es la distribución del parámetro θ , que depende del parámetro genérico ψ . $p(y)$ es la distribución marginal de los datos: $p(y) = \int_{\theta} p(y | \theta) p(\theta) d\theta$. Como esta última no depende de θ (porque al integrar sobre θ se obtiene una constante), se la considera como constante de normalización, con lo cual se puede reemplazar «=» por « \propto » («proporcional a»): $p(\theta | y) \propto p(y | \theta) p(\theta | \psi)$. La idea es asignar, antes de ver los datos, una distribución «prior» al parámetro θ , que indica una creencia sobre los posibles valores de θ . Esta creencia es actualizada por la distribución de θ teniendo en cuenta la evidencia (datos). Si la evidencia es muy fuerte (N grande), el *prior* queda «borrado» por los datos y la posterior dependerá más de la verosimilitud que del *prior*; pero si es débil la elección del prior resulta influyente. En cualquier caso, la distribución posterior siempre es un compromiso entre verosimilitud de los datos y el *prior* del parámetro. La distribución posterior puede resumirse según: (i) la moda; (ii) la media; (iii) la mediana; (iv) el intervalo de *credibilidad* del 95%. Este último da los valores de θ que dejan a la derecha y a la izquierda una densidad de probabilidad del 0.025. Es decir que indica la probabilidad de que el parámetro esté en un intervalo de credibilidad del 95%.¹

La distribución posterior para el modelo mixto es (Sorensen & Gianola, 2002, cap.14.2):

$$p(\beta, u, \sigma_u^2 | y) = p(y | \beta, u) p(\beta) p(u | \sigma_u^2) p(\sigma_u^2 | \psi)$$

Las variables aleatorias se distribuyen como sigue ($c = 1, 2, 3, 4$):

Respuesta: $y_{ijc} | \beta, u \sim \mathcal{M}(N; (\pi_1, \pi_2, \pi_3, \pi_4))$;

Coefficientes: $\beta_{kc} | \sigma_{\beta}^2 \sim \mathcal{N}(0, \sigma_{\beta}^2 = \sigma_e^2 + \pi^2/3)$; en notación vectorial: $\beta | \sigma_{\beta}^2 \sim NMV(\mathbf{0}, \mathbf{B} \otimes \mathbf{V}_{\beta})$; $\mathbf{B}=\mathbf{I}$.

Factores aleatorios: $u_j | \sigma_u^2 \sim \mathcal{N}(0, \sigma_u^2)$; donde: $\sigma_u^2 \sim \chi_u^2(1)$ [chi-cuadrada con 1 grado de libertad]; En notación vectorial: $\mathbf{u} | \sigma_u^2 \sim NMV(\mathbf{0}, \mathbf{A} \otimes \mathbf{V}_{\mathbf{u}})$; $\mathbf{A}=\mathbf{I}$; $V \sim W(v, S)$ [Wishart].

Las matrices de varianza-covarianza $\mathbf{V}_{\mathbf{u}}$ de los factores aleatorios son diagonales heterogéneas, con una varianza para cada efecto aleatorio. Los efectos aleatorios se consideran independientes. Por ejemplo, en el modelo más simple habrá cuatro efectos aleatorios de ordenada al origen, uno por cada categoría c , para cada grupo. Entonces la matriz tendrá en su diagonal las varianzas para cada uno de ellos.

$$V_{SESSION:ID,INTERCEPT} = \begin{bmatrix} \sigma_{v_{01}}^2 & & & \\ & \sigma_{v_{02}}^2 & & \\ & & \sigma_{v_{03}}^2 & \\ & & & \sigma_{v_{04}}^2 \end{bmatrix}$$

Es necesario estimar las distribuciones posteriores de los coeficientes (β), los factores aleatorios (u) y la varianza de los factores aleatorios (σ_u^2). La varianza de los errores no es un parámetro a estimar porque no es independiente de la media para los modelos (multi/bi)nomiales. Según recomendaciones de Hadfield (2010b) se fijará la varianza de los errores en $V = \frac{1}{k} (I_{(k-1)} + J_{(k-1, k-1)})$, donde k es el número de categorías de la respuesta, I es la matriz identidad y J , una matriz de unos. Siguiendo a Gelman (2006) se usó el siguiente prior para β : $\beta_k \sim N(0, \sigma_e^2 + \pi^2/3)$. Por otra parte, se tomará la sugerencia de Villemereuil et al. (2013) [ver también: Villemereuil (2012)] de utilizar *priors expandidos* para σ_u^2 , y usando una distribución

¹En cambio, en estadística frecuentista el intervalo de confianza indica que, si se repitiera muchas veces el experimento con diferentes muestras, el 95% de la veces el verdadero parámetro poblacional caerá en el intervalo. Como no es una variable aleatoria (y por lo tanto carece de distribución) esto no es lo mismo que decir que hay una probabilidad del 95% de que el parámetro caiga dentro del intervalo.

χ^2 (que pertenece a la familia *gama*) con un grado de libertad². Generalmente estas varianzas se modelan como χ^2 inversas escaladas, pero la nueva distribución pone menos densidad de probabilidad cerca del cero, lo cual hace que la MCMC salga más fácilmente de la región del cero si se queda atascada.

La solución de la posterior $p(\beta, u, \sigma_u^2 | y)$ no es analítica explícita, sino que se hace por medios computacionales usando MCMC (Markov Chain Monte Carlo) y sus aplicaciones, como el muestreo de *Gibbs* (Blangiardo & Cameletti, 2016, cap. 4). Una serie de variables aleatorias $(X^{(0)}, X^{(1)}, \dots, X^{(t)}, \dots)$ es una cadena de markov si se satisface la condición de que la probabilidad condicional de $X^{(t)}$ dado $(X^{(0)}, X^{(1)}, \dots, X^{(t-1)})$, solo dependa de $X^{(t-1)}$ y sea independiente de las demás: $P(X^{(t)} | X^{(0)}, X^{(1)}, \dots, X^{(t-1)}) = P(X^{(t)} | X^{(t-1)})$; donde el término de la derecha indica la probabilidad de transición del estado de la cadena en $t-1$ al estado de la misma en t . El punto importante es que, satisfaciendo ciertas condiciones, la cadena *llega* («converge») a una distribución π de estados que es invariante; que no se modifica con el tiempo. En el contexto bayesiano, la distribución invariante es precisamente $p(\theta | y)$, la posterior. La idea es tomar una secuencia de valores de los parámetros de la posterior $(\theta^1, \theta^2, \dots)$ hasta que la cadena alcance la distribución invariante; que será una aproximación a la posterior. Un método para lograrlo es el muestreo de Gibbs. Supóngase una posterior $p(\theta | y)$ con vector de parámetros $\theta = (\theta_1, \dots, \theta_p)$. Para cada parámetro θ_i se obtiene la distribución posterior de dicho parámetro *condicional* al resto (o sea manteniendo el valor del resto como fijo): $p(\theta_i | \theta_{j \neq i}, y)$, para $i = 1, \dots, p$. El procedimiento consiste en muestrear valores iterativamente de todas las distribuciones condicionales hasta alcanzar la distribución estacionaria $\pi = p(\theta | y)$. La convergencia sucede luego de un (largo) número de interacciones, digamos $t > t_0$. Las interacciones hasta t_0 se descartan, y este conjunto de interacciones descartadas se conoce como periodo de *burn-in*. Sucede que en dicho periodo los valores muestreados pueden estar más correlacionados entre sí. En cambio los valores muestreados de la posterior (invariante) deben ser independientes. Por ello, un diagnóstico frecuente de convergencia de la cadena consiste en evaluar la auto-correlación r a diferentes *lags* o intervalos entre valores y verificar que $r < 0,1$. Cuanto más parámetros tenga el modelo, más se tardará en lograr la convergencia.

Por último, resulta conveniente presentar una medida bayesiana de comparación de modelos: el criterio de información de devianza [DIC] (Gelman et al., cap. 6). La devianza se define como menos dos veces el logaritmo de la verosimilitud del modelo: $D(y, \hat{\theta}) = -2 \log p(y | \hat{\theta})$, donde $\hat{\theta}$ es una estimación puntual, por ejemplo la moda o la media de la posterior del (vector de) parámetro(s). La esperanza $D_{avg}(y) = E(D(y, \theta) | y)$ puede estimarse como: $\hat{D}_{avg}(y) = \frac{1}{L} \sum_{l=1}^L D(y, \theta^{(l)})$, donde $\theta^{(l)}$ es la l -ésima muestra del parámetro, usando, por ejemplo, muestreo de Gibbs. Es decir que es un promedio de las devianzas calculadas sobre cada uno de los valores del parámetro obtenidos en la posterior. Usando estas dos fórmulas, se define el número efectivo de parámetros como: $p_D = \hat{D}_{avg}(y) - D(y, \hat{\theta})$. Entonces la medida DIC, que debe minimizarse para el mejor modelo rezará: $DIC = 2\hat{D}_{avg}(y) - D(y, \hat{\theta}) = p_D + \hat{D}_{avg}(y) = D(y, \hat{\theta}) + 2p_D$. La formulación de la derecha es similar a la medida AIC. Por ejemplo, si: $D(y, \hat{\theta}) = 54,9$; $\hat{D}_{avg}(y) = 62,6$ y $p_D = 7,7$; se calcula como: $DIC = 2(62,6) - 54,9 = 7,7 + 62,6 = 54,9 + 2(7,7) = 70,3$.

7.2. Expectativa del análisis.

Se desea encontrar las variables que influyen en la chance de que un caso de concordancia tenga el tipo de error ($Y = c$), para $c = 1, 2, 3, 4$. Recuérdese que los tipos de error se definieron como: error de género ($c = 1$); error de *-e-* epentética ($c = 2$); error de plural ($c = 3$); error mixto ($c = 4$). Se esperan los siguientes resultados de los predictores, en cuanto a la facilitar con un tamaño de efecto menor (\uparrow) / mayor ($\uparrow\uparrow$), o inhibir con un tamaño de efecto menor (\downarrow) / mayor ($\downarrow\downarrow$) errores de concordancia. Solo se pusieron en el Cuadro los predictores discretos, ya que no se involucrará a los continuos en el análisis.

Según el Cuadro, se espera que existan más *tipos* de errores con concordancias que contengan: (i) artículos indefinidos, determinantes y adjetivos; (ii) concordancias a larga distancia (la chance de error crece

²Se expande el efecto aleatorio u_i en dos componentes: $u_i = \alpha \eta_i$ con $\eta_i \sim \mathcal{N}(0; V_\eta)$; $\alpha \sim \mathcal{N}(0, V_\alpha)$; $V_\eta \sim W^{-1}(S, \nu)$ [Wishart inversa]. Implícitamente es lo mismo que definir $V_u = V_u / V_\alpha$, que se distribuye como una F de Fisher $\mathcal{F}(1, \nu)$. A medida que $\nu \rightarrow \infty$, dicha distribución se acerca a una $\chi^2_{(1)}$ (chi-cuadrado con un grado de libertad).

Predictor	Efecto	$c = 1$	$c = 2$	$c = 3$	$c = 4$
MOD	MOD(1) - MOD(0)	↑	↑	↑	↑
	MOD(2) - MOD(0)	↑	↑	↑	↑
	MOD(3) - MOD(0)	↑	↑	↑	↑
Fabs.SC.f	Fabs.SC.f(1) - Fabs.SC.f(0)	↓	↓	↓	↓
MORF.f	MORF.f(1) - MORF.f(0)	↑	↑	↑	↑
	MORF.f(2) - MORF.f(0)	↑↑	↑↑	↑↑	↑↑
CUMRES.f	CUMRES.f(1) - CUMRES.f(0)	↑	↑	↑	↑
	CUMRES.f(2) - CUMRES.f(0)	↑↑	↑↑	↑↑	↑↑
FAM.LEX.f	FAM.LEX.f(1) - FAM.LEX.f(0)	↓	↓	↓	↓
EST1	EST1(1) - EST1(0)	↓↓	↓	↓	↓
EST2	EST2(1) - EST2(0)	↓↓	↓	↓	↓
EST3	EST3(1) - EST3(0)	↑	↑	↑	↑
EST4	EST4(1) - EST4(0)	↓	↓	↓	↓
EST5	EST5(1) - EST5(0)	↓	↓↓	↓	↓
EST4	EST6(1) - EST6(0)	↓	↓	↓	↓
EST7	EST7(1) - EST7(0)	↑	↑	↑	↑
ANIM	ANIM(1) - ANIM(0)	↑	↑	↑	↑
ES	ES(1) - ES(0)	↑	↑	↑	↑
	ES(2) - ES(0)	↑↑	↑↑	↑↑	↑↑
POS	POS(1)-POS(0)	↑	↑	↑	↑
	POS(2)-POS(0)	↑↑	↑↑	↑↑	↑↑
DIS	POS(1)-POS(0)	↑	↑	↑	↑
	POS(2)-POS(0)	↑↑	↑↑	↑↑	↑↑
LDA	LDA(1)-LDA(0)	↑	↑	↑	↑
GRAMS	GRAMS(1)-GRAMS(0)	↑	↑	↑	↑
STEM.f	STEM.f(1)-STEM.f(0)	↑	↑	↑	↑
IMA.CONC.f	IMA.CONC.f(1) - IMA.CONC.f(0)	↓	↓	↓	↓
GRUPO6	GRUPO6(2)-GRUPO6(1)	↑	↑	↑	↑
GRUPO6	GRUPO6(3)-GRUPO6(1)	↑	↑	↑	↑
GRUPO6	GRUPO6(4)-GRUPO6(1)	?	?	?	?
GRUPO6	GRUPO6(5)-GRUPO6(1)	↑	↑	↑	↑
GRUPO6	GRUPO6(6)-GRUPO6(1)	?	?	?	?

Cuadro 7.1: Expectativas del análisis. Leyenda: (↑) = efecto aumenta; (↓) = efecto disminuye; (↑↑/↓↓) = mayor tamaño de efecto; ? = efecto indeterminado.

con: distancia larga / sin subordinada > distancia entre enunciados / con subordinada); (iii) *-e-* epentética, con chance de error creciente si está en un término o en ambos; (iv) concordancias de más de dos términos; (v) controlador animado; (vi) TYPE infrecuente; (vii) similaridad baja (alta distancia) entre las raíces léxicas del español y el italiano; (viii) similaridad *media* (*media* distancia) entre las terminaciones del español y el italiano, según el tamaño de efecto *baja* > *media*; (ix) cantidad *media* > *alta* de errores acumulados; (x) controlador infrecuente y/o no familiar; (xi) controlador de baja imaginabilidad y/o concretud; (xii) las «estrategias» EST3 y EST7 porque identifican casos de difícil producción para los italianos; (xiii) en GRUPO6, los grupos (5), (2), (3); el primero porque concentra casos de larga distancia, infrecuentes y sin artículos definidos; el segundo, por aglomerar casos de MORF.f alta y con *-e-* epentética en un término; el tercero, por concentrar casos de más de dos términos y con distancias MORF.f media. En los demás niveles de GRUPO6 el efecto se considera indeterminado, por eso se marca «?». Además, se espera hallar un efecto inhibitorio mayor para el error de género en EST1 y EST2 ya que las estrategias se definen según género masculino o femenino respectivamente; y también para EST5 respecto del error de *-e-* epentética ya que la estrategia ayuda a sortear la inserción de *-e-*.

7.3. Selección de modelos.

Se comenzó por un análisis univariado de cada predictora categórica contra la respuesta nominal, mediante el test de independencia χ^2 . El siguiente es un gráfico de mosaico entre las variables MORF.f y RES_CAT. El tamaño de los rectángulos es proporcional a las frecuencias observadas de las celdas (f_{ij}) en la tabla de contingencia. Los colores no grisáceos indican las celdas significativas al 5% según los residuos estandarizados de devianza, que toman como referencia la escala de la derecha. En este caso, se rechaza la hipótesis de independencia en las celdas [RES_CAT = 0, MORF.f = 0] y [RES_CAT = 3, MORF.f = 0]; el test global de X^2 es significativo ($p < 0,0001$). El resto de las gráficas se muestran en el apéndice. La tabla 1 muestra las celdas significativas (residuos de Devianza) para cada variable categórica (filas) y RES_CAT (columnas) y el p-valor del test global. Los p-valores con (*) indican que los valores esperados no cumplen con los requisitos del test. Por tanto estos resultados se excluyen. Del resto, en las variables que resultaron significativas se resaltan las siguientes asociaciones entre RES_CAT y las variables: (i) ADJ: entre errores de plural y mixtos y concordancias sin artículos definidos; (ii) ANIM: entre errores de *-e-* epentética y concordancias con controlador inanimado; (iii) entre errores mixtos y concordancias con pocos errores acumulados; (iv) Fabs.SC.f: errores de género y concordancias de TYPE frecuente; y entre errores de plural y TYPE infrecuente; (v) entre errores de *-e-* epentética y concordancias con controladores frecuentes y/o familiares; (vi) EST1: entre errores de género y mixtos y la estrategia 1; (vii) EST2: entre errores de género y la estrategia 2; (viii) EST5: entre errores con *-e-* epentética y la estrategia 5; (ix) GRAMS: entre errores de género y concordancias de más de dos términos.

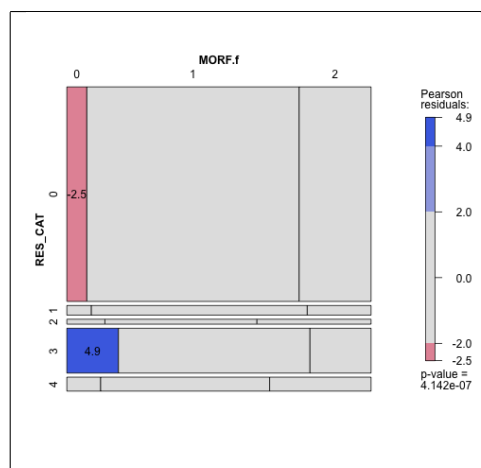


Figura 7.1: Gráfico de Mosaico para MORF.f y RES_CAT.

Variable	$P(X_{obs}^2 \geq X^2)$	0	1	2	3	4
IMA.CONC.f	0.234					
MOD	$p < 0,0001$ (*)		2		0,3	0
ADJ	$p < 0,0001$				0,1	0,1
LDA	0.021 (*)				1	
POS	0.084 (*)					
DIS	0.009 (*)				1	3
ANIM	0.001			0,1		
CUMRES.f	0.0003					0
Fabs.SC.f	$p < 0,0001$		1		0,1	
FAM.LEX.f	0.022			0,1		
ES	0.001 (*)				2	1
STEM.f	0.292					
GRAMS	0.027		0,1			
EST1	$p < 0,0001$		1			1
EST2	0.0005		1			
EST3	0.06 (*)					
EST4	0.005 (*)			1		1
EST5	0.01			1		
EST6	0.014 (*)					
EST7	0.001 (*)					1
GRUPO6	$p < 0,0001$ (*)			6	3,6	2

Cuadro 7.2: Test de X^2 : celdas significativas según residuos de devianza. (*) No cumple con los requisitos del test.

Luego se llevó a cabo selección de modelos (multinomiales generalizados) basado en medidas de información (ver cap. 6), optimizando la función de log-verosimilitud por medio de una red neuronal (utilizando el paquete *nnet* de *R* [Venables & Ripley (2002), p. 203]). Se decidió dividir el problema en dos grupos de variables, ya que utilizar todas las discretas implicaba una búsqueda exhaustiva aproximada de 4 millones de modelos. El primer grupo contenía en el modelo global a las variables predictoras: Fabs.SC.f, MORF.f, STEM.f, MOD, ES, ANIM, GRAMS, FAM.LEX.f, IMA.CONC.f, CUMRES.f, LDA. Mientras que el segundo contenía: POS, DIS, EST1, EST2, EST3, EST4, EST5, EST6, EST7, GRUPO6. Se usó *AIC* como medida de información para la selección. Los Cuadros 3 y 4 muestra la importancia relativa de las predictoras sobre 2048 modelos del grupo I, y 1024 del grupo II. Se reporta la importancia relativa de las predictoras para ambos grupos, sobre el total de modelos. Se nota que LDA, STEM.f, GRAMS, DIS, POS, EST4 poseen probabilidades debajo del 50%. A continuación, se llevó a cabo el promedio de los coeficientes en el conjunto de «confianza» de los modelos (con la regla $\frac{w(i)}{w(1)} > \frac{1}{8}$). Los Cuadros 5 y 6 muestran las predictoras promediadas que resultan significativas: Fabs.SC.f, MOD, ANIM, FAM.LEX.f, ES, MORF.f, CUMRES.f, EST1, EST2, EST5, GRUPO6. A excepción de GRAMS (y ADJ), las variables del test de X^2 aparecen todas.

7.4. Modelo Bayesiano. Resultados y chequeo de supuestos.

Se ajustaron modelos mixtos multinomiales bayesianos con la parte fija saturada con las variables seleccionadas en el apartado anterior (sin interacciones). Se utilizó el paquete *MCMCglmm* de *R* para el ajuste (Hadfield, 2010a). El Cuadro que sigue muestra la medida DIC con el objetivo de elegir la estructura aleatoria adecuada de los modelos. Se corrieron 4100000 iteraciones para cada modelo (muestreando cada 2000), con el objetivo de reducir la correlación en la distribución posterior y asegurar la convergencia de la cadena de Markov. También se incluyó TIME como efecto fijo. Los modelos comparados fueron: (i) con efecto de aleatorio de ordenada al origen; (ii) con efecto de aleatorio de ordenada al origen y FAM.LEX.f; ; (iii) con

	names	x
1	Fabs.SC.f	1.00
2	ANIM	1.00
3	MORF.f	1.00
4	CUMRES.f	1.00
5	FAM.LEX.f	1.00
6	MOD	0.99
7	ES	0.90
8	IMA.CONC.f	0.72
9	LDA	0.44
10	STEM.f	0.22
11	GRAMS	0.11

Cuadro 7.3: Importancia relativa de las predictoras: grupo 1

	names	x
1	EST5	1.00
2	EST1	0.98
3	GRUPO6	0.96
4	EST2	0.93
5	EST6	0.82
6	EST7	0.75
7	EST3	0.62
8	EST4	0.48
9	DIS	0.27
10	POS	0.06

Cuadro 7.4: Importancia relativa de las predictoras: grupo 2

efecto de aleatorio de ordenada al origen y tiempo. Se observa en el Cuadro 7 que el modelo con efecto aleatorio de intercepto para cada modelo c ($c = 1, 2, 3, 4$), más efecto aleatorio de FAM.LEX resultó ser el mejor (DIC = 2857.807). Se eligió éste. El Cuadro también incluye el ICC para cada c (se usó la moda para la estimación puntual). Nótese que las correlaciones intra clase resultan bajas.

El modelo mixto multinomial bayesiano general para la concordancia i en el grupo j (el grupo está definido como *la sesión k anidada en el alumno g*) [$k = 1, \dots, 12(14)$; $g = 1, \dots, 4$; $j = 1, \dots, 52$; $i = 1, \dots, 1857$], dado que la observación tiene la categoría $c = 1, \dots, 4$ es:

Distribuciones y Priors: $y_{ij} = c | v_{0ij}, v_{1ij} \sim \text{multinomial}(N, (\pi_{ij1} \pi_{ij2} \pi_{ij3} \pi_{ij4}))$, $N = 1$; $v_0 \sim \text{NMV}(\mathbf{0}, \mathbf{G}_0)$;

$v_1 \sim \text{NMV}(\mathbf{0}, \mathbf{G}_1)$; $\beta \sim \text{NMV}(\mathbf{0}, \mathbf{V}_\beta)$;

Matrices de varianza: $G_{\text{INTERCEPT}} = V \otimes A$; $G_{\text{FAM.LEX.f}} = V \otimes A$; $A = I$; $V = I$ (efectos aleatorios independientes).

$$V_{\text{SESION:ID,INTERCEPT}} = \begin{bmatrix} \sigma_{v_01}^2 & & & \\ & \sigma_{v_02} & & \\ & & \sigma_{v_03} & \\ & & & \sigma_{v_04} \end{bmatrix};$$

$$V_{\text{SESION:ID;FAM:LEX}} = \begin{bmatrix} \sigma_{v_1, \text{FAM.LEX.f}=0}^2 & \\ & \sigma_{v_1, \text{FAM.LEX.f}=1} \end{bmatrix}$$

Función de enlace: $\text{logit} = g(E[y_{ij} = c | v_{0ij}, v_{1ij}]) = \log \left[\frac{\pi_{ijc}}{1 - \pi_{ijc}} \right]$

Función inversa: $\pi_{ij} = g^{-1}(X\beta + Z_v v)$

	Estimate	Std. Error	z value	Pr(> z)
1((Intercept))	-3.799	0.691	5.494	0.000
1(Fabs.SC.f1)	-1.171	0.320	3.661	0.000
1(MOD2)	0.646	0.309	2.089	0.037
2((Intercept))	-3.543	0.924	3.835	0.000
2(ANIM1)	1.833	0.431	4.248	0.000
2(FAM.LEX.f1)	-1.317	0.424	3.102	0.002
3((Intercept))	-0.724	0.294	2.461	0.014
3(CUMRES.f2)	0.470	0.173	2.711	0.007
3(ES2)	-18.804	8.832	2.129	0.033
3(FAM.LEX.f1)	-0.330	0.137	2.408	0.016
3(Fabs.SC.f1)	-0.402	0.168	2.395	0.017
3(MOD2)	0.353	0.175	2.018	0.044
3(MOD3)	0.482	0.199	2.425	0.015
3(MORF.f1)	-1.090	0.231	4.720	0.000
3(MORF.f2)	-0.691	0.304	2.272	0.023
4((Intercept))	-3.841	0.598	6.420	0.000
4(CUMRES.f1)	1.114	0.363	3.067	0.002
4(CUMRES.f2)	1.297	0.365	3.554	0.000
4(FAM.LEX.f1)	-0.464	0.230	2.019	0.043
4(MOD2)	1.074	0.293	3.659	0.000

Cuadro 7.5: Promedio de los coeficientes con FULL AVERAGE ($p < 0.05$): grupo 1

Predictor lineal: $\eta_{ij} = \varphi_{ij} + v_{0ij} + v_{1ij}$. Donde: φ_{ij} representa los efectos fijos; v_{0ij} y v_{1ij} son los efectos aleatorios «entre».

Tamaño muestral efectivo: $(Interactions - burnin) / thin = (4100000 - 100000) / 2000 = 2000$.

Modelo teórico: para $c = 1, 2, 3, 4$, con $c = 0$ como categoría de referencia.

$$g(E[y_{ij} = c | v_{0ij}, v_{1ij}]) = \log \left[\frac{\pi_{ijc}}{1 - \pi_{ijc}} \right] = (\beta_{0c} + v_{0ic}) + (\beta_{1c} + v_{1ic}) FAM.LEX.f$$

$$\beta_{2c}MOD_1 + \beta_{3c}MOD_2 + \beta_{4c}MOD_3 + \beta_{5c}CUMRES.f_1 + \beta_{6c}CUMRES.f_2$$

$$\beta_{7c}Fabs.SC.f + \beta_{8c}MORF.f_1 + \beta_{9c}MORF.f_2 + \beta_{10c}EST1 + \beta_{11c}EST5 + \beta_{12c}EST2$$

$$\beta_{13c}GRUPO6_2 + \beta_{14c}GRUPO6_3 + \beta_{15c}GRUPO6_4 + \beta_{16c}GRUPO6_5 + \beta_{17c}GRUPO6_6$$

$$\beta_{18c}ES1 + \beta_{19c}ES2 + \beta_{20c}ANIM + \beta_{21c}TIME$$

	Estimate	Std. Error	z value	Pr(> z)
1((Intercept))	-2.904	0.324	8.949	0.000
1(EST11)	-1.693	0.534	3.169	0.002
2((Intercept))	-2.946	0.370	7.954	0.000
2(EST21)	-2.629	1.053	2.496	0.013
2(GRUPO66)	1.419	0.546	2.600	0.009
3((Intercept))	-1.845	0.176	10.506	0.000
3(GRUPO63)	0.544	0.167	3.256	0.001
3(GRUPO65)	0.974	0.300	3.245	0.001
4((Intercept))	-2.914	0.316	9.226	0.000
4(EST51)	-0.971	0.460	2.111	0.035
4(GRUPO62)	0.930	0.369	2.517	0.012

Cuadro 7.6: Promedio de los coeficientes con FULL AVERAGE ($p < 0.05$): grupo 2

Medida	Modelo (i)	Modelo (ii)	Modelo (iii)
DIC	2873.743	2857.807	2873.343
ICC(1) [ESS; CI(lo, up)]	0.343 [576.943; (0.165, 0.507)]	0.302 [814.635; (0.132, 0.441)]	0.31 [705.153; (0.163, 0.496)]
ICC(2) [ESS; CI(lo, up)]	0.002 [290.597; (3e-08, 0.336)]	0.002 [278.429; (1e-05, 0.315)]	0.0012 [419.289; (7e-07, 0.337)]
ICC(3) [ESS; CI(lo, up)]	0.063 [1380.019; (0.018, 0.151)]	0.0005 [1103.633; (3e-08, 0.091)]	0.0004 [1020.867; (5e-09, 0.119)]
ICC(4) [ESS; CI(lo, up)]	0.181[634.272; (0.011, 0.307)]	0.001[518.895; (5e-09, 0.18)]	0.001[515.107; (2e-07, 0.221)]

Cuadro 7.7: Modelo bayesiano: selección de estructura aleatoria. ICC(c) [c=1,2,3,4]: Coeficiente de correlación intra clase; ESS: tamaño muestral efectivo; CI(lo, up): Intervalo de credibilidad (inferior, superior).

Parámetro	Descripción
ν_{0ic}	la desviación del grupo i_c de la ordenada al origen
ν_{1ic}	la desviación del grupo i_c de la media marginal de FAM.LEX.f.
β_{0c}	la media basal marginal
β_{1c}	el efecto de FAM.LEX.f, nivel 1 (referencia: FAM.LEX.f = 0)
β_{2c}	el efecto de MOD, nivel 1 (referencia: MOD = 0)
β_{3c}	el efecto de MOD, nivel 2 (referencia: MOD = 0)
β_{4c}	el efecto de MOD, nivel 3 (referencia: MOD = 0)
β_{5c}	el efecto de CUMRES.f, nivel 1 (referencia: CUMRES.f = 0)
β_{6c}	el efecto de CUMRES, nivel 2 (referencia: CUMRES.f = 0)
β_{7c}	el efecto de Fabs.SC.f, nivel 1 (referencia: Fabs.SC.f = 0)
β_{8c}	el efecto de MORF.f, nivel 1 (referencia: MORF.f = 0)
β_{9c}	el efecto de MORF.f nivel 2 (referencia: MOD = 0)
β_{10c}	el efecto de EST1, nivel 1 (referencia: EST1 = 0)
β_{11c}	el efecto de EST5, nivel 1 (referencia: EST1 = 0)
β_{12c}	el efecto de EST2, nivel 1 (referencia: EST1 = 0)
β_{13c}	el efecto de GRUPO6, nivel 2 (referencia: GRUPO6 = 1)
β_{14c}	el efecto de GRUPO6, nivel 3 (referencia: GRUPO6 = 1)
β_{15c}	el efecto de GRUPO6, nivel 4 (referencia: GRUPO6 = 1)
β_{16c}	el efecto de GRUPO6, nivel 5 (referencia: GRUPO6 = 1)
β_{17c}	el efecto de GRUPO6, nivel 6 (referencia: GRUPO6 = 1)
β_{18c}	el efecto de ES, nivel 1 (referencia: ES = 0)
β_{19c}	el efecto de ES, nivel 2 (referencia: ES = 0)
β_{20c}	el efecto de ANIM, nivel 1 (referencia: ES = 0)
β_{21c}	el efecto de TIME

Cuadro 7.9: Parámetros del modelo logístico multinomial bayesiano.

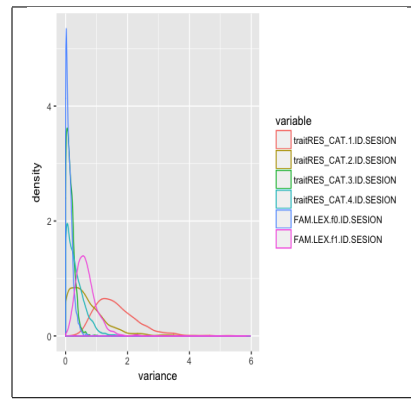


Figura 7.2: Posteriores de las varianzas.

El Cuadro 9 describe los parámetros del modelo a estimar, para *cada* modelo $c = 1, 2, 3, 4$. Hubo 22 efectos fijos a estimar por modelo. Considérense en primer lugar los diagnósticos de convergencia. En lo que respecta a las varianzas de los factores aleatorios, las correlaciones en $lags > 2000$ estaban debajo de $< 0.1^3$. Para el $lag 2000$ (equivalente al $lag 1$) la autocorrelación fue mayor a 0.1 para $r_{intercept,c,lag=2000} = [1 = 0,33; 2 = 0,6; 3 = 0,21; 4 = 0,55]$; y < 0.1 para FAM.LEX.f. En suma, la autocorrelación fue moderada ($lag 1$) para las varianzas en las categorías 2 y 4. En cuanto a los coeficientes β , la correlación como mucho rondó 0.3. Por lo tanto, se considera que hubo una convergencia razonable de las distribuciones posteriores⁴.

La gráfica de la Figura 2 ilustra las posteriores de las varianzas. Se nota que son muy asimétricas a derecha. Por lo tanto se usó la moda como medida para la estimación puntual.

Las varianzas de los factores aleatorios son los que se detallan en las matrices:

$$V_{INTERCEPT} = \begin{bmatrix} 1,139 & & & \\ & 0,015 & & \\ & & 0,002 & \\ & & & 0,004 \end{bmatrix}; V_{FAM.LEX.f} = \begin{bmatrix} 0,001 & \\ & 0,562 \end{bmatrix}.$$

Hubo $52 \times 4 = 208$ efectos aleatorios de ordenada al origen y $52 \times 2 = 104$ efectos aleatorios de FAM.LEX.f. Los desvíos típicos de los efectos aleatorios de ordenada al origen para cada categoría y para FAM.LEX.f resultaron:

$$\begin{aligned} \sigma_{0ic} &= [c1 = 1,067; c2 = 0,125; c3 = 0,052; c4 = 0,068] \\ \sigma_{1ic} &= [FAM.LEX.f(0) = 0,043; FAM.LEX.f(1) = 0,75] \end{aligned}$$

Hubo más desvío en las categorías 1 y 2 respecto de las 3 y 4. El desvío de FAM.LEX.f(0) fue casi nulo. El coeficiente de ordenada al origen de *cada* categoría varió en términos del *logit* como $\beta_{0c} = [c1 = -2,28; c2 = -3,07; c3 = -0,72; c4 = -2,77]$ más / menos la desviación del efecto aleatorio «entre» de ordenada al origen de cada grupo. Análogamente, el efecto (coeficiente) de FAM.LEX.f varió en términos del *logit* en cada categoría como:

$\beta_{1c} = [c1 = -0,46; c2 = -1,36; c3 = -0,56; c4 = -0,89]$ más / menos la desviación del efecto «entre» de FAM.LEX.f de cada grupo.

En los que atañe a los supuestos distribucionales de los factores aleatorios, se aplicó el test de *Shapiro-Wilks* para detectar desvíos groseros del supuesto de normalidad. Se tomó a la media como resumen de la posterior⁵. En cuanto a los efectos aleatorios de ordenada al origen, el supuesto no se cumplió para las categorías 1 y 2 ($p < 0,0001$) y se cumplió para las categorías 3 ($p = 0,77$) y 4 ($p = 0,45$). En los que atañe

³Como se muestrea de la posterior cada 2000 valores, las equivalencias serían: $lag 2000 = lag 1$; $lag 10000 = lag 5$; $lag 20000 = lag 10$; $lag 100000 = lag 50$.

⁴El test de estacionariedad de Heidelberg & Welch (1981) propone como hipótesis nula que los valores muestreados de la posterior provienen de una distribución estacionaria. Ningún σ^2 de efectos aleatorios o β de efectos fijos rechazó dicha hipótesis nula.

⁵La correlación de Pearson entre las estimaciones puntuales usando la media y la mediana estuvieron arriba de $r > 0,98$; con lo cual se consideró que las distribuciones eran simétricas. Por tanto, la media es un buen resumen de la distribución.

a los factores aleatorios de FAM.LEX.f se cumplió para ambos, aunque con menos evidencia para rechazar la hipótesis nula de normalidad para [FAM.LEX.f = 1] ($p = 0,054$), que para [FAM.LEX.f = 0] ($p = 0,39$). Las gráficas de la Figura 3 muestran el *qq-plot* para examinar gráficamente el supuesto de normalidad de los factores aleatorios. Cuánto más cerca de la línea central de igualdad entre los cuantiles empíricos y los teóricos de una distribución normal, mejor se sigue el supuesto de normalidad. Se verifican los resultados del test: los factores aleatorios de ordenada al origen para categorías 1 y 2; y para el nivel 1 de FAM.LEX.f se apartan más del supuesto de normalidad.

Por otra parte, las Figuras 4 y 5 ilustran las trayectorias, para cada sujeto, de los efectos aleatorios de ordenada al origen para los modelos con $c = 1, 2, 3, 4$. En lo que respecta a la categoría 1 (error de género), para JAKO y MIRKA el nivel de error se mantiene por debajo del poblacional en la mayoría de las sesiones. En cambio SONIA excede al poblacional en 6 sesiones (cayendo debajo del poblacional en las últimas tres) y NATI en 10, esta última con diferencias más pronunciadas, pero disminuyendo las diferencias positivas en las últimas sesiones. En cuanto a la categoría 2 (error de *-e-* epentética), JAKO y MIRKA se mantienen también debajo de la media de error poblacional. Sin embargo, ahora SONIA posee las diferencias más pronunciadas aunque disminuyendo hacia las sesiones finales. Por el contrario las diferencias positivas aumentan con el correr de las sesiones para NATI. En lo que atañe a la categoría 3 (error de plural), SONIA se mantiene por debajo del umbral de error poblacional (las diferencias positivas son muy pequeñas). JAKO cruza a diferencias positivas en las sesiones 1, 2, 7, 8 y 10 pero en las últimas cuatro sesiones se mantiene debajo de la media general de error. MIRKA no logra salir de diferencias positivas, aunque acercándose a la media general de error en la última sesión. Por último, NATI consigue controlar los errores de plural hacia las tres últimas sesiones. Respecto de la categoría 4 (errores mixtos) JAKO solo se ve afectado en las sesiones 1 y 4. SONIA tiene diferencias positivas debajo de 0.25 en las sesiones 2, 6 y 8. MIRKA está debajo del nivel general de error en las sesiones 1, 5, 6 para luego pasar a diferencias positivas aumentando incluso en las últimas dos sesiones. NATI solo cruza a errores debajo de la media poblacional en las sesiones 6, 9 y 14.

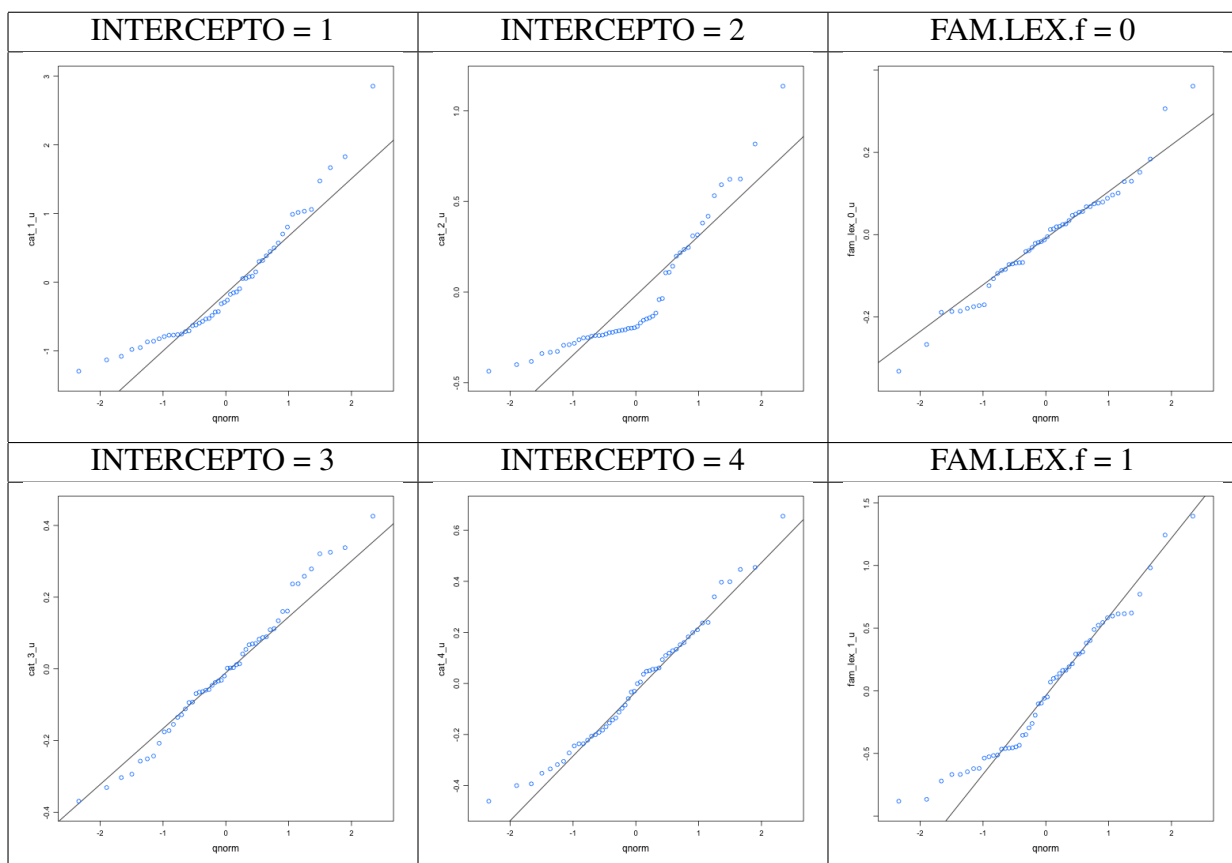


Figura 7.3: Factores Aleatorios: QQ-plots.

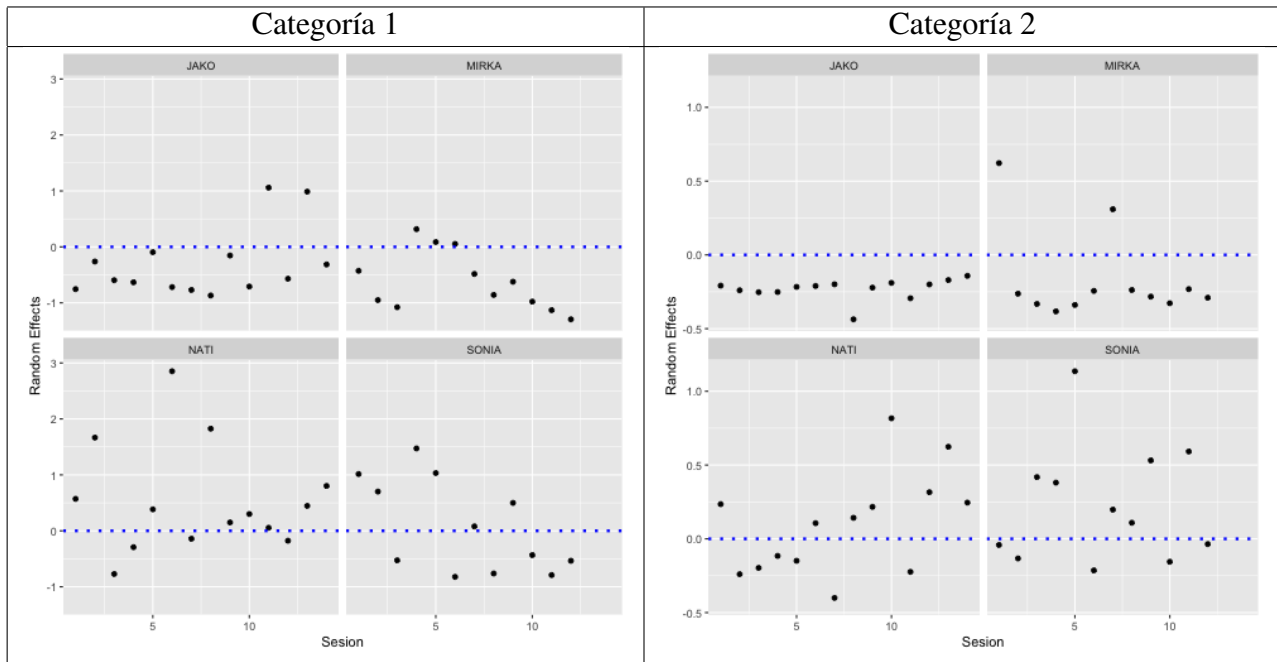


Figura 7.4: Trayectorias: Factores aleatorios de ordenada al origen. Categorías 1 y 2.

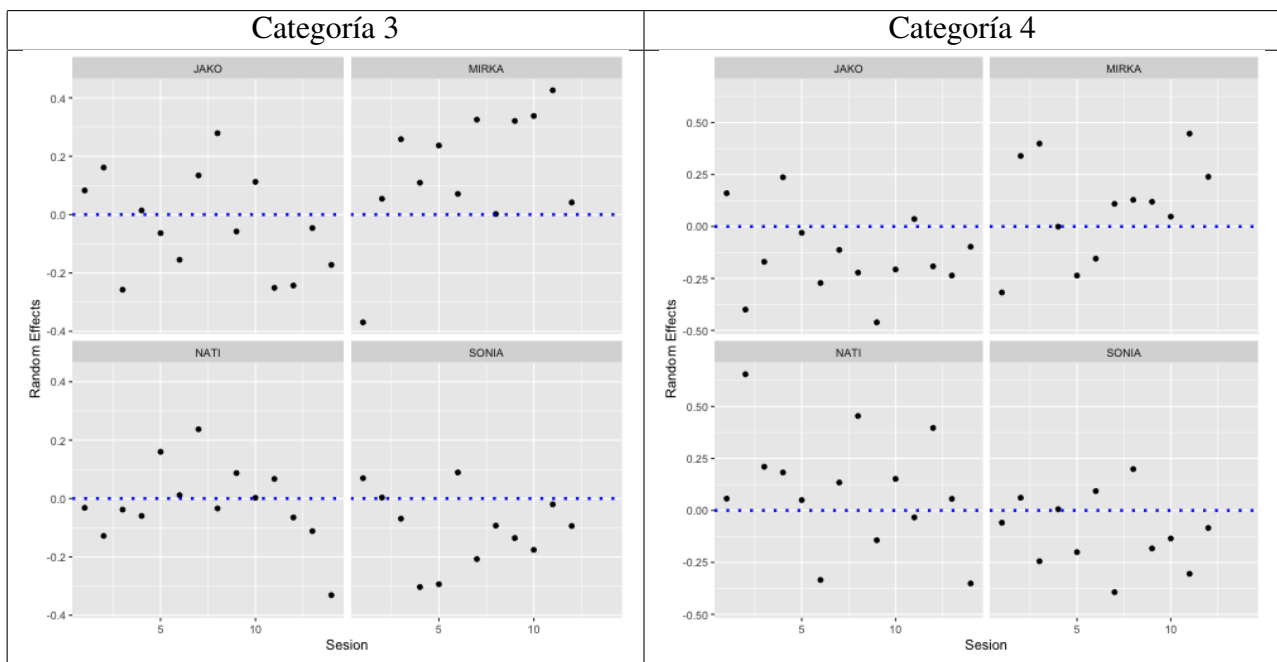


Figura 7.5: Trayectorias: Factores aleatorios de ordenada al origen. Categorías 3 y 4.

El Cuadro 10 muestra los resultados del ajuste para los coeficientes significativos (el Cuadro total se incluye en el apéndice). Se tomó la media de la posterior de los coeficientes β como medida puntual⁶. La Figura 6 es un *forest plot* con las estimaciones puntuales y el intervalo de credibilidad en escala del *logit*. Resultaron significativos (Intervalo de credibilidad no contiene al cero) los siguientes efectos (entre corchetes se interpreta en términos del *Odds Ratio*): (i) error de género (1): MOD(3) [60% menos de chance de error], Fabs.SC.f [72% menos], EST1 [83% menos]; (ii) error de *-e-* epentética (2): ANIM [250% más de chance de error], FAM.LEX.f [75% menos], EST2 [78% menos], EST5 [89% menos]; (iii) error de plural: MOD(3) [53% más de chance de error], Fabs.SC.f [35% menos], ES(2) [87% menos], MORF.f(1) [71% menos], MORF.f(2) [62% menos], FAM.LEX.f [43% menos]; (iv) errores mixtos (4): MOD(2) [141% más de chance de error], Fabs.SC.f [60% menos], FAM.LEX.f [60% menos], CUMRES.f(1) [100% más],

⁶La correlación de Pearson entre medias y modas de los coeficientes fue $r = 0,988$.

EST1 [59% menos], EST5 [85% menos]. Por lo tanto, son factores de riesgo de cometer un error ANIM en categoría 2; MOD(3) en categoría 3; MOD(2) y CUMRES.f(1) en categoría 4. Los demás protegen contra el error. Notar que el tamaño del efecto (distancia a la línea vertical del cero, en el *forest plot*) es muy pequeño para Fabs.SC.f [categoría 3], MOD(3) [categoría 3], MOD(3) [categoría 1]. Si bien el p-valor MCMC de GRUPO6(5) es 0.036; el intervalo de credibilidad es $(-2,366; 0,050)$, como incluye al cero no se considera dicho predictor.

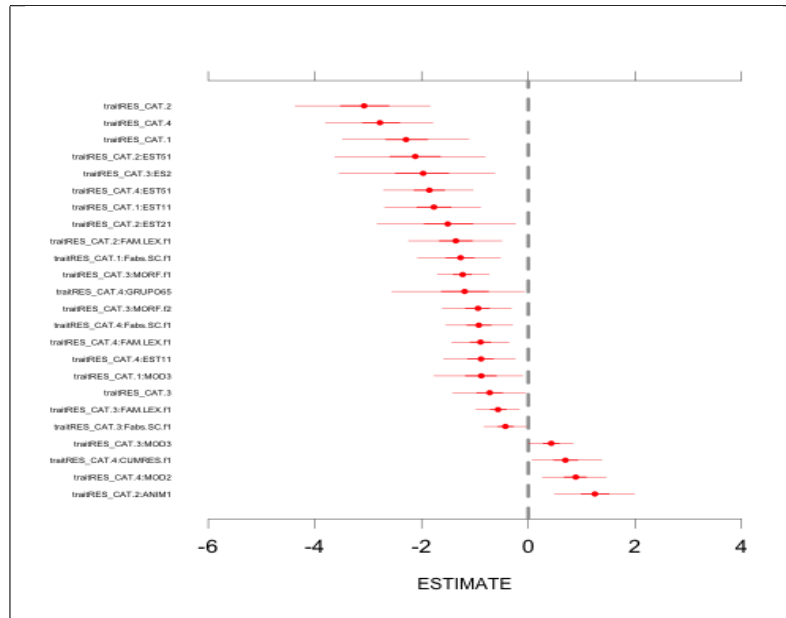


Figura 7.6: Forest Plot: estimaciones puntuales de efectos simples e intervalos de credibilidad. Escala del *logit*.

7.5. Análisis de errores.

Se discriminará según la categoría de error. Se indicarán entre paréntesis las probabilidades condicionales de error de tipo $c = 1, 2, 3, 4$, dado el nivel del predictor (usando la escala del porcentaje), por ejemplo: $p(RES.CAT = 1 | MOD = 3)$ es la probabilidad de cometer error de género dado que la concordancia tiene un adjetivo.

7.5.1. Error de género vs. no error.

La chance de cometer un error de concordancia ($Y = 1$) respecto de no cometerlo ($Y = 0$) *disminuye* cuando:

1. Se pasa de concordancias con artículo definido a aquellas con adjetivos. Si había artículo los errores fueron 24 (2,7%); si había adjetivo fueron 10 (2,1%). Ejemplos: *primeros veces* [SONIA, sesión 4, línea 137]; *turistas italianos* [NATI, sesión 6, línea 21]; *películas estúpidos* [NATI, sesión 8, línea 75]; *costumbres italianos* [JAKO, sesión 11, línea 51]. En el caso de «turist-a», el plural italiano es «turist-i», el error analoga «-i» a «-os» en español. La palabra «película» es masculina en italiano: «il film»; por tanto es un error de mala especificación del rasgo de género en la base léxica. Por otra parte «abitudine» es sustantivo femenino pero hace plural en *-i*. JAKO podría estar aplicando la estrategia EST1 [*-i > os*] en el adjetivo. Notar que la dirección del efecto fue *contraria* a la esperada: menos errores con adjetivos que con artículos definidos. Sin embargo, también en el caso de los artículos se observan casos que tienen que ver con los controladores y no con el modificador: (i) plurales italianos en «-i» a los que se les aplica «-os» en español: *i turisti* (masc. pl.) > *los turistas* [NATI, sesión 6, línea 21]; *le religioni* (fem. pl.) > *los religiones* [NATI, sesión 6, línea 76]; *i protagonisti* (masc. pl.)

	mean	l.CI	u.CI	eff	pMCMC	OR	l.CI.OR	u.CI.OR
traitRES_CAT.1	-2.281	-3.443	-1.081	2000.000	p<0.001	0.102	0.032	0.339
traitRES_CAT.2	-3.071	-4.420	-1.890	1572.478	p<0.001	0.046	0.012	0.151
traitRES_CAT.3	-0.724	-1.432	-0.085	2000.000	0.037	0.485	0.239	0.919
traitRES_CAT.4	-2.772	-3.801	-1.786	1594.264	p<0.001	0.063	0.022	0.168
traitRES_CAT.4:MOD2	0.883	0.315	1.500	1596.517	p<0.001	2.419	1.370	4.480
traitRES_CAT.1:MOD3	-0.898	-1.814	-0.163	1173.143	0.032	0.407	0.163	0.850
traitRES_CAT.3:MOD3	0.431	0.007	0.840	1738.849	0.045	1.538	1.007	2.316
traitRES_CAT.1:Fabs.SC.f1	-1.277	-2.036	-0.491	1596.650	p<0.001	0.279	0.130	0.612
traitRES_CAT.3:Fabs.SC.f1	-0.424	-0.799	-0.012	1851.367	0.035	0.654	0.450	0.988
traitRES_CAT.4:Fabs.SC.f1	-0.927	-1.485	-0.250	1530.948	p<0.001	0.396	0.226	0.778
traitRES_CAT.2:ANIM1	1.254	0.521	2.008	1190.987	p<0.001	3.503	1.684	7.445
traitRES_CAT.3:ES2	-2.007	-3.532	-0.624	1601.730	p<0.001	0.134	0.029	0.536
traitRES_CAT.3:MORF.f1	-1.233	-1.714	-0.746	2000.000	p<0.001	0.291	0.180	0.474
traitRES_CAT.3:MORF.f2	-0.949	-1.557	-0.287	1647.699	p<0.001	0.387	0.211	0.751
traitRES_CAT.2:FAM.LEX.f1	-1.362	-2.201	-0.475	1098.049	p<0.001	0.256	0.111	0.622
traitRES_CAT.3:FAM.LEX.f1	-0.563	-0.953	-0.141	2000.000	p<0.001	0.570	0.386	0.868
traitRES_CAT.4:FAM.LEX.f1	-0.897	-1.432	-0.360	1533.184	p<0.001	0.408	0.239	0.698
traitRES_CAT.4:CUMRES.f1	0.704	0.025	1.322	1653.229	0.026	2.021	1.026	3.749
traitRES_CAT.1:EST11	-1.773	-2.720	-0.924	914.828	p<0.001	0.170	0.066	0.397
traitRES_CAT.4:EST11	-0.893	-1.539	-0.229	1130.814	p<0.001	0.410	0.215	0.796
traitRES_CAT.2:EST21	-1.510	-2.773	-0.198	672.114	0.021	0.221	0.062	0.820
traitRES_CAT.2:EST51	-2.139	-3.540	-0.729	856.514	p<0.001	0.118	0.029	0.483
traitRES_CAT.4:EST51	-1.855	-2.756	-1.088	1372.963	p<0.001	0.156	0.064	0.337
traitRES_CAT.4:GRUPO65	-1.207	-2.366	0.050	1367.219	0.036	0.299	0.094	1.051

Cuadro 7.10: ajuste del modelo (efectos con $p < 0.05$). mean = media de la posterior, l CI = Int. de credibilidad del 95 por ciento, extremo inferior, u CI = Int. de credibilidad del 95 por ciento, extremo superior; eff = tamaño muestral efectivo; pMCMC = p valor; OR = $\exp(\text{mean})$, l CI OR = $\exp(\text{l CI})$, u CI OR = $\exp(\text{u CI})$

> *los protagonistas* [NATI, sesión 8, línea 194]; (ii) plurales irregulares del italiano en *-a*: *le ossa* (masc. pl.) > *las huesos* [MIRKA, sesión 4, línea 75]; *le uova* (masc. pl.) > *las huevos* [MIRKA, sesión 6, línea 278] (en ambos casos por analogar *le > las*); (iii) géneros diferentes en ambas lenguas: *i film* (masc. pl.); otros casos irregulares: *gli analfabeti* (masc. pl.) [masc. sg.: *l'analfabeta*] > *los analfabetas*. En suma, el caso del error de género parece estar más asociado al controlador que al tipo de modificador.

- Se pasa de concordancias con TYPE infrecuente a aquellas con TYPE frecuente. Si hay TYPE infrecuente los errores fueron 44 (4,3%); en cambio si es frecuente fueron 10 (2,1%). Ejemplos: *muchos corrientes* [SONIA, sesión 5, línea 89]; *los empresas* [NATI, sesión 10, línea 114]; *otros posibilidades* [JAKO, sesión 13, línea 51]; *los situaciones* [MIRKA, sesión 5, línea 153].
- Se pasa de concordancias donde no se puede aplicar la estrategia 1 (poner plural en *-os* si el plural italiano termina en *-i*) a aquellas en donde se puede aplicar. Si no es aplicable los errores fueron 57 (4,3%); en cambio si lo es fueron 5 (0,9%). Ejemplos: *barrios nuevas* [SONIA, sesión 7, línea 296]; *muchas monumentos* [NATI, sesión 11, línea 215]; *los datas* [JAKO, sesión 13, línea 18]; *las ministerios* [MIRKA, sesión 9, línea 231].

7.5.2. Error de *-e-* epentética vs. no error.

La chance de cometer un error de concordancia ($Y = 1$) respecto de no cometerlo ($Y = 0$) disminuye cuando:

1. Se pasa de concordancias con controlador no familiar y/o infrecuente a aquellas con controlador familiar y/o frecuente. En el primer caso fueron 24 (2,4%); en cambio en el segundo fueron 10 (0,9%). Ejemplos: *calles grandas* [SONIA, sesión 4, línea 60]; *las mujeres* [NATI, sesión 13, línea 129]; *pocos trenos* [MIRKA, sesión 1, línea 158].
2. Se pasa de concordancias donde no se puede aplicar la estrategia 2 (poner plural en *-as* si el plural italiano termina en *-e*) a aquellas en donde se puede aplicar. Si no es aplicable los errores fueron 31 (2,2%); en cambio si lo es no hubo ninguno.
3. Se pasa de concordancias donde no se puede aplicar la estrategia 5 (si palabra singular del italiano termina en *-e*, agregar *-s* a la desinencia en italiano) a aquellas en donde se puede aplicar. Si no es aplicable los errores fueron 32 (1,9%); en cambio si lo es no hubo ninguno.

La chance de cometer un error de concordancia ($Y = 1$) respecto de no cometerlo ($Y = 0$) *aumenta* cuando se pasa de concordancias con controlador no animado a aquellas con controlador animado. Si era inanimado los errores fueron 12 (0,9%); en cambio si era animado fueron 20 (3,8%). Ejemplos: *los alemanos* [SONIA, sesión 3, línea 178]; *los ruses* [NATI, sesión 8, línea 188]. Observando las 20 concordancias en cuestión, 7 de ellas repiten el error con «tedesco» [masc. sing.] (alemán), los italianos transfieren el singular a «alemano» en español y forman el plural a partir de éste: «alemanos», en lugar de insertar *-e*⁷. Otros dos casos son con la palabra «modelo» que en italiano es «modella» [fem. sing.]; y su plural «modelle»; la concordancia formada es «las *modeles*», agregando *-s-* al plural del italiano. Luego se usa tres veces «mujer», cuya contraparte italiano es «*donna*» [fem. sing.]. El error es formar el singular en español como «mujera» y agregarle *-s-*: «*las mujeres*». NATI hablaba también francés como segunda lengua extranjera. Tres de sus producciones: «católicos moralistes» [NATI, sesión 6, línea 232], «los soldates» [NATI, sesión 11, línea 85], «los ruses» [NATI, sesión 8, línea 188] parecen tener que ver con cognados del francés: «*catholiques moralistes*»; «*les russes*»; «*les soldates*».

7.5.3. Error de plural vs. no error.

La chance de cometer un error de concordancia ($Y = 1$) respecto de no cometerlo ($Y = 0$) *disminuye* cuando:

1. Se pasa de concordancias con TYPE infrecuente a aquellas con TYPE frecuente. Si hay TYPE infrecuente los errores fueron 192 (19%); en cambio si es frecuente fueron 97 (11,4%). Ejemplos: *niño solos* [SONIA, sesión 6, línea 243]; *esta historias* [NATI, sesión 8, línea 32]; *muchísimas tienda* [JAKO, sesión 2, línea 43]; *los animale* [MIRKA, sesión 3, línea 32].
2. Se pasa de concordancias sin *-e-* epentética en ningún término de la concordancia a aquellas con *-e-* epentética en ambos términos. En el primer caso los errores fueron 217 (16,2%); en cambio en el segundo no hubo ninguno. La dirección del efecto fue contraria a la esperada: no hubo errores para ES(2), que se supuso como el nivel de la variable de mayor dificultad. Por tanto, la presencia de *-e-* epentética no pone dificultades a los cuatro aprendientes.
3. Se pasa de concordancias con controlador no familiar y/o infrecuente a aquellas con controlador familiar y/o frecuente. En el primer caso los errores fueron 156 (16,7%); en cambio en el segundo hubo 133 (14,3%). Ejemplos: *los profesore* [SONIA, sesión 9, línea 225]; *vinos tinto* [NATI, sesión 1, línea 128]; *lo viajes* [JAKO, sesión 1, línea 75]; *su juegos* [MIRKA, sesión 9, línea 109].
4. Se pasa de concordancias con similaridad alta (distancia baja) de terminaciones entre español e italiano a aquellas con similaridad media (distancia media). Si la similaridad era alta los errores fueron 49 (30,8%); en cambio si era similaridad media hubo 182 (14,4%). Ejemplos: *la herbas* [SONIA, sesión 6, línea 78]; *mucha lluvias* [NATI, sesión 5, línea 71]; *cartas escrita* [JAKO, sesión 2, línea 172]; *los pollo* [MIRKA, sesión 3, línea 34].

⁷Idéntico error se produce con *il treno* > los *trenos (trenes).

5. Se pasa de concordancias con similaridad alta (distancia baja) de terminaciones entre español e italiano a aquellas con similaridad alta (distancia baja). Si la similaridad era alta los errores fueron 49 (30,8 %); en cambio si era similaridad media hubo 58 (13,1 %). Ejemplos: *grande ciudades* [SONIA, sesión 2, línea 12]; *lo jueces* [NATI, sesión 7, línea 60]; *la comunicaciones* [JAKO, sesión 8, línea 70]; *programa musicales* [MIRKA, sesión 10, línea 143].

La chance de cometer un error de concordancia ($Y = 1$) respecto de no cometerlo ($Y = 0$) *aumenta* cuando se pasa de concordancias con artículo definido a aquellas con adjetivo. Si estaba presente el artículo definido los errores fueron 103 (11,8 %); en cambio si había adjetivo fueron 96 (20,7 %). Ejemplos: *días ocupado* [SONIA, sesión 8, línea 369]; *solucione secretas* [NATI, sesión 7, línea 109], *diferentes fuente* [JAKO, sesión 7, línea 186]; *platos típico* [MIRKA, sesión 4, línea 51].

Notar que en el caso de MORF.f no solo la dirección del efecto fue contraria sino también la magnitud del efecto: las concordancias de distancia alta (similaridad baja) tuvieron efecto negativo, aunque menor que en el caso de la distancia media. Para [MORF.f = 1, CAT = 3] se podría sugerir que, al igual que en el caso de la respuesta binaria, el hecho de tener que utilizar pocas operaciones de «borrado» pueda bajar la chance de cometer error. En el caso de [MORF.f = 2, CAT = 3] el efecto resultó protector aunque menos inhibitorio de error. En estos casos había ocho instancias en las que se debían aplicar más operaciones de «inserción», que, en la construcción del índice se supuso como la operación de mayor costo. Por ejemplo: *città > ciudades* requiere insertar «d-e-s», con lo cual se aumenta la dificultad. Aunque el OR del predictor fue ajustado por el efecto de EST5, obsérvese que 25 de los 58 casos (43,1 %) eran oportunidades de aplicar la estrategia 5, que es factor de protección⁸.

7.5.4. Error mixto vs. no error.

La chance de cometer un error de concordancia ($Y = 1$) respecto de no cometerlo ($Y = 0$) *disminuye* cuando:

1. Se pasa de concordancias con TYPE infrecuente a aquellas con TYPE frecuente. Si hay TYPE infrecuente los errores fueron 60 (5,9 %); en cambio si es frecuente fueron 30 (3,5 %). Ejemplos: *los vacaciones* [SONIA, sesión 4, línea 10]; *historia románticos* [NATI, sesión 3, línea 159]; *la sitios* [JAKO, sesión 2, línea 102]; *las imagen* [MIRKA, sesión 10, línea 136].
2. Se pasa de concordancias con controlador no familiar y/o infrecuente a aquellas con controlador familiar y/o frecuente. En el primer caso los errores fueron 51 (5,4 %); en cambio en el segundo hubo 39 (4,2 %). Ejemplos: *les joven* [SONIA, sesión 2, línea 144]; *este empresa* [NATI, sesión 10, línea 206]; *animales común* [JAKO, sesión 4, línea 72]; *las mujer* [MIRKA, sesión 10, línea 65].
3. Se pasa de concordancias donde no se puede aplicar la estrategia 1 a aquellas en donde se puede aplicar. Si no es aplicable los errores fueron 77 (5,9 %); en cambio si lo es fueron 13 (2,3 %). Ejemplos: *este daños* [NATI, sesión 12, línea 104]; *procesos mental* [JAKO, sesión 5, línea 199]; *todo lo genere* [MIRKA, sesión 2, línea 246].
4. Se pasa de concordancias donde no se puede aplicar la estrategia 5 a aquellas en donde se puede aplicar. Si no es aplicable los errores fueron 82 (6,2 %); en cambio si lo es fueron 8 (1,4 %). Ejemplos: *los profesor* [SONIA, sesión 8, línea 386]; *inundaciones peligroso* [NATI, sesión 5, línea 162]; *estas función* [JAKO, sesión 8, línea 78]; *seres trabajador* [MIRKA, sesión 3, línea 48]. Las palabras que involucran la estrategia fueron: it. sg. *istituzione* > esp. pl. *instituciones* (instancia: «mucho *instituciones*»); it. sg. *professore* > esp. pl. *profesores* (instancia: «los *profesor*»); it. sg. *inondazione* > esp. pl. *inundación* (instancia: «*inundaciones* peligroso»); it. sg. *animale* > esp. pl. *animales* (instancia: «*animales* común»); it. sg. *funzione* > esp. pl. *funciones* (instancia: «*estas función*»); it. sg. *essere* > esp. pl. *seres*; it. sg. *lavoratore* > esp. pl. *trabajadores* (instancia: «*seres trabajador*»). En los ejemplos la estrategia no se aplica para *professore*, *funzione*, *lavoratore*; en las otras concordancias el error está en el término que no involucra la estrategia.

⁸Recordar que en el capítulo anterior se constató alta asociación entre MORF.f y EST1 / EST5.

La chance de cometer un error de concordancia ($Y = 1$) respecto de no cometerlo ($Y = 0$) *aumenta* cuando:

1. Se pasa de concordancias con artículo definido a aquellas con determinante. Si estaba presente el artículo definido los errores fueron 25 (2,8%); en cambio si había adjetivo fueron 34 (6,9%). Ejemplos: *estes filósofos* [SONIA, sesión 5, línea 66]; *mucho discusiones* [NATI, sesión 12, línea 120], *estas función* [JAKO, sesión 8, línea 78]; *este chicos* [MIRKA, sesión 3, línea 227].
2. Se pasa de concordancias con hasta 2 errores acumulados (sin contar el error actual) a aquellas con hasta 7 errores acumulados. En el primer caso los errores fueron 10 (1,9%); en cambio en el segundo fueron 39 (5,7%). Ejemplos: *cuánta vez* [SONIA, sesión 6, línea 294]; *muchas personaje* [NATI, sesión 8, línea 32], *tan hombres* [JAKO, sesión 1, línea 68]; *grande direttore* [MIRKA, sesión 12, línea 124]. O sea, solamente el tipo de error más complicado fue sensible a la acumulación de errores previos a la instancia.

7.6. Conclusiones.

El Cuadro que sigue muestra las expectativas cumplidas del análisis. Resultaron conformes a las expectativas los efectos (respecto del nivel de referencia) de: MOD(2) en error mixto [factor de riesgo]; Fabs.SC.f(1) en errores de *-e-* epentética, plural, mixto [factor de protección]; CUMRES.f(1) en error mixto [factor de riesgo]; FAM.LEX.f en errores de género, plural y mixto [factor de protección]; EST1 en error de género y mixto [factor de protección]; EST2 en error de *-e-* epentética [factor de protección]; EST5 en error de *-e-* epentética y mixto [factor de protección]; ANIM en error de *-e-* epentética [factor de riesgo]. Además EST1 obtuvo un efecto fuerte en el error de género y EST5 también con respecto al error de *-e-* epentética. EST2 no obtuvo efecto asociado al error de género. Por otra parte, no resultaron conformes a las expectativas los efectos (respecto del nivel de referencia) de: MOD(3) en error de género [factor de protección y no de riesgo]; MORF.f(1) y MORF.(3) en error de plural [factor de protección y no de riesgo]; ES(2) en error de plural [factor de protección y no de riesgo]. Asimismo, ES(2) obtuvo una magnitud de efecto fuerte pero en la dirección opuesta; y la magnitud del efecto resultó en dirección opuesta para MORF.f(2) < MORF.f(1).

Predictor	Efecto	$c = 1$	$c = 2$	$c = 3$	$c = 4$	Comentario
MOD	MOD(1) - MOD(0)	↑ NO	↑ NO	↑ NO	↑ NO	
	MOD(2) - MOD(0)	↑ NO	↑ NO	↑ NO	↑ SI	
	MOD(3) - MOD(0)	↑ NO	↑ NO	↑ SI	↑ NO	$c = 1$, dirección opuesta (↓)
Fabs.SC.f	Fabs.SC.f(1) - Fabs.SC.f(0)	↓ NO	↓ SI	↓ SI	↓ SI	
MORF.f	MORF.f(1) - MORF.f(0)	↑ NO	↑ NO	↑ NO	↑ NO	$c = 3$, dirección y tamaño de efecto opuestas (↓↓)
	MORF.f(2) - MORF.f(0)	↑↑ NO	↑↑ NO	↑↑ NO	↑↑ NO	$c = 3$, dirección y tamaño de efecto opuestas (↓)
CUMRES.f	CUMRES.f(1) - CUMRES.f(0)	↑ NO	↑ NO	↑ NO	↑ SI	
	CUMRES.f(2) - CUMRES.f(0)	↑↑ NO	↑↑ NO	↑↑ NO	↑↑ NO	
FAM.LEX.f	FAM.LEX.f(1) - FAM.LEX.f(0)	↓ SI	↓ NO	↓ SI	↓ SI	
EST1	EST1(1) - EST1(0)	↓↓ SI	↓ NO	↓ NO	↓ SI	
EST2	EST2(1) - EST2(0)	↓↓ NO	↓ SI	↓ NO	↓ NO	
EST5	EST5(1) - EST5(0)	↓ NO	↓↓ SI	↓ NO	↓ SI	
ANIM	ANIM(1) - ANIM(0)	↑ NO	↑ SI	↑ NO	↑ NO	
ES	ES(1) - ES(0)	↑ NO	↑ NO	↑ NO	↑ NO	
	ES(2) - ES(0)	↑↑ NO	↑↑ NO	↑↑ NO	↑↑ NO	$c = 3$, dirección y tamaño de efecto opuestas (↓↓)

Cuadro 7.11: Expectativas cumplidas del análisis. Leyenda: (↑) = efecto aumenta; (↓) = efecto disminuye; (↑↑/↓↓) = mayor tamaño de efecto; NO = no se cumple expectativa (cero está dentro del intervalo de credibilidad); SI = se cumple expectativa (cero está fuera del intervalo de credibilidad).

El Cuadro que sigue compara los resultados de las respuestas binaria y categórica. MOD fue factor de riesgo para ambas respuestas en el caso del error mixto pero resultó factor de protección para el caso del

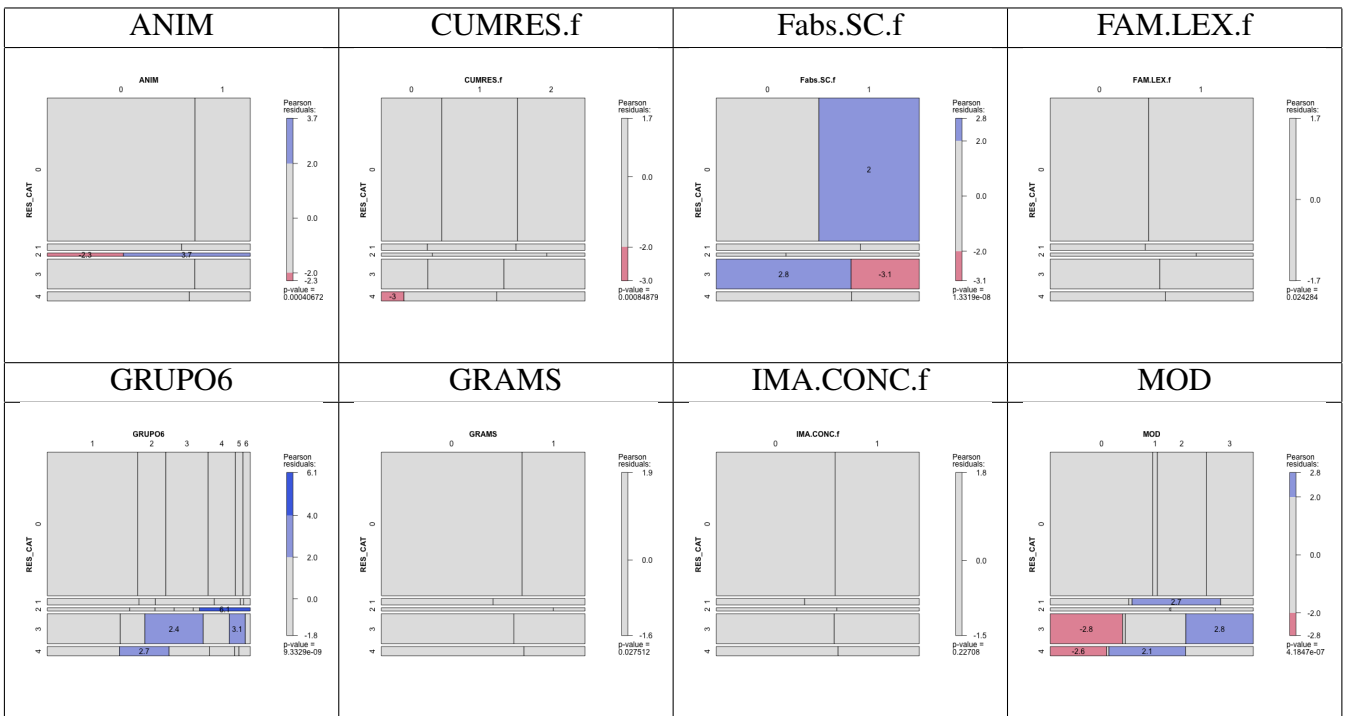
error de género. No obstante, el análisis de las instancias reveló que los errores no se debían al tipo de modificador. Por lo tanto, el efecto con dirección opuesta merece más investigación. Fabs.SC.f y FAM.LEX.f fueron en ambas respuestas factores de protección y para tres de los cuatro tipos de errores. CUMRES.f solamente resultó significativo para la respuesta categórica en errores mixtos. Las estrategias EST1 y EST5 fueron factores de protección en las dos respuestas. EST5 tuvo un efecto mayor en el caso de la respuesta categórica. Por otra parte, la estrategia EST2 solamente resultó factor de protección para errores de *-e-epentética*. En lo que atañe a la similitud entre las terminaciones entre Español e Italiano, MORF.f(1) [similitud media] tuvo mayor magnitud de efecto [de protección] que MORF.f(2) [similitud baja], pero este último no resultó significativo en el caso de la respuesta binaria. En ambos casos el efecto observado resultó contrario al esperado. Se especuló con la posible dificultad asociada a la operación de «borrado». Los controladores *animados* resultaron ser factores de riesgo de error. Por último, en ES(2) [*-e-epentética* en ambos términos de la concordancia] no verificó ningún error de plural (efecto fuerte); por tanto, la presencia de *-e-epentética* no pone dificultades a los cuatro aprendientes.

Predictor	BINARIA	CATEGÓRICA	ERROR
MOD	MOD(2) - MOD(0) [↑]	MOD(2) - MOD(0) [↑] MOD(3) - MOD(0) [↓]	mixto género
Fabs.SC.f	Fabs.SC.f(1) - Fabs.SC.f(0) [↓]	Fabs.SC.f(1) - Fabs.SC.f(0) [↓]	<i>-e-epentética</i> , plural, mixto
MORF.f	MORF.f(1) - MORF.f(0) [↓↓]	MORF.f(1) - MORF.f(0) [↓↓] MORF.f(2) - MORF.f(0) [↓]	plural plural
FAM.LEX.f	FAM.LEX.f(1) - FAM.LEX.f(0) [↓]	FAM.LEX.f(1) - FAM.LEX.f(0) [↓]	género, plural, mixto
CUMRES.f	-	CUMRES.f(1) - CUMRES.f(0) [↑]	mixto
EST1	EST1(1) - EST1(0) [↓]	EST1(1) - EST1(0) [↓]	género, mixto
EST5	EST5(1) - EST5(0) [↓]	EST5(1) - EST5(0) [↓↓]	<i>-e-epentética</i> , mixto
EST2	-	EST2(1) - EST2(0) [↓]	<i>-e-epentética</i>
ANIM	-	ANIM(1) - ANIM(0) [↑]	<i>-e-epentética</i>
ES	-	ES(2) - ES(0) [↓↓]	plural

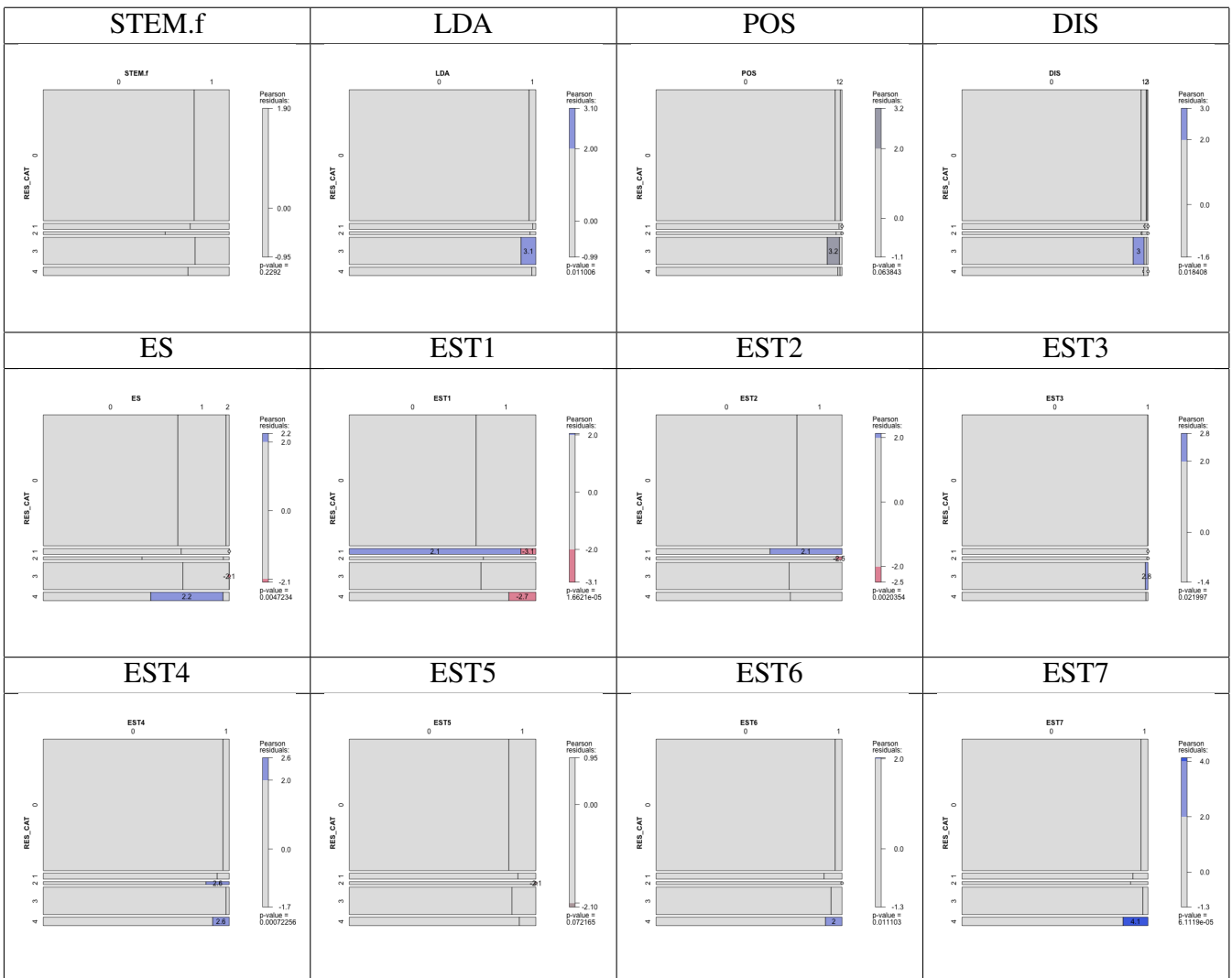
Cuadro 7.12: Comparación de efectos para respuestas binaria y categórica. Leyenda: (↑) = efecto aumenta; (↓) = efecto disminuye; (↑↑/↓↓) = mayor tamaño de efecto.

7.7. Apéndice de capítulo.

7.7.1. Gráficas de mosaico



(a) Mosaic Plot: predictor contra respuesta nominal.



(b) Mosaic Plot: predictor contra respuesta nominal.

Figura 7.7: Gráficas de mosaico.

7.7.2. Ajuste del modelo multinomial bayesiano.

Cuadro 7.13: ajuste del modelo (todos los predictores). mean = media de la posterior, l CI = Int. de credibilidad del 95 por ciento, extremo inferior, u CI = Int. de credibilidad del 95 por ciento, extremo superior; eff = tamaño muestral efectivo; pMCMC = p valor; OR = exp(mean), l CI OR = exp(l CI), u CI OR = exp(u CI)

	mean	l.CI	u.CI	eff	pMCMC	OR	l.CI.OR	u.CI.OR
traitRES_CAT.1	-2.281	-3.443	-1.081	2000.000	0.001	0.102	0.032	0.339
traitRES_CAT.2	-3.071	-4.420	-1.890	1572.478	0.001	0.046	0.012	0.151
traitRES_CAT.3	-0.724	-1.432	-0.085	2000.000	0.037	0.485	0.239	0.919
traitRES_CAT.4	-2.772	-3.801	-1.786	1594.264	0.001	0.063	0.022	0.168
traitRES_CAT.1:TIME	-0.012	-0.044	0.019	1332.941	0.458	0.988	0.957	1.020
traitRES_CAT.2:TIME	-0.007	-0.043	0.029	1093.957	0.734	0.993	0.958	1.029
traitRES_CAT.3:TIME	0.009	-0.004	0.024	1870.334	0.184	1.009	0.996	1.024
traitRES_CAT.4:TIME	0.009	-0.012	0.029	1874.825	0.384	1.009	0.989	1.030
traitRES_CAT.1:MOD1	-0.169	-1.743	1.276	2000.000	0.847	0.845	0.175	3.582
traitRES_CAT.2:MOD1	-0.583	-2.443	1.412	1713.370	0.569	0.558	0.087	4.104
traitRES_CAT.3:MOD1	-0.489	-1.645	0.532	1687.632	0.382	0.613	0.193	1.702
traitRES_CAT.4:MOD1	-0.556	-1.936	1.012	1609.775	0.475	0.574	0.144	2.750
traitRES_CAT.1:MOD2	0.347	-0.287	0.935	1812.128	0.293	1.415	0.751	2.546
traitRES_CAT.2:MOD2	0.194	-0.755	1.124	1276.952	0.644	1.215	0.470	3.077
traitRES_CAT.3:MOD2	0.328	-0.032	0.706	2038.619	0.089	1.388	0.968	2.025
traitRES_CAT.4:MOD2	0.883	0.315	1.500	1596.517	0.001	2.419	1.370	4.480
traitRES_CAT.1:MOD3	-0.898	-1.814	-0.163	1173.143	0.032	0.407	0.163	0.850
traitRES_CAT.2:MOD3	-0.259	-1.443	0.758	1208.766	0.683	0.772	0.236	2.135
traitRES_CAT.3:MOD3	0.431	0.007	0.840	1738.849	0.045	1.538	1.007	2.316
traitRES_CAT.4:MOD3	0.472	-0.157	1.163	1655.423	0.149	1.603	0.855	3.201
traitRES_CAT.1:Fabs.SC.f1	-1.277	-2.036	-0.491	1596.650	0.001	0.279	0.130	0.612
traitRES_CAT.2:Fabs.SC.f1	0.071	-0.799	1.099	1278.310	0.900	1.073	0.450	3.002
traitRES_CAT.3:Fabs.SC.f1	-0.424	-0.799	-0.012	1851.367	0.035	0.654	0.450	0.988
traitRES_CAT.4:Fabs.SC.f1	-0.927	-1.485	-0.250	1530.948	0.003	0.396	0.226	0.778
traitRES_CAT.1:ANIM1	-0.121	-0.772	0.543	1447.283	0.738	0.886	0.462	1.721
traitRES_CAT.2:ANIM1	1.254	0.521	2.008	1190.987	0.003	3.503	1.684	7.445
traitRES_CAT.3:ANIM1	0.132	-0.192	0.467	1809.376	0.436	1.141	0.826	1.596
traitRES_CAT.4:ANIM1	0.519	-0.026	1.031	1673.390	0.063	1.680	0.974	2.803
traitRES_CAT.1:ES1	0.541	-0.647	1.677	1797.495	0.382	1.718	0.524	5.351
traitRES_CAT.2:ES1	0.066	-1.274	1.383	1680.962	0.933	1.068	0.280	3.989
traitRES_CAT.3:ES1	-0.308	-1.049	0.379	2000.000	0.426	0.735	0.350	1.461
traitRES_CAT.4:ES1	0.482	-0.519	1.454	1799.522	0.355	1.620	0.595	4.281
traitRES_CAT.1:ES2	-0.672	-2.717	1.237	1867.843	0.525	0.511	0.066	3.445
traitRES_CAT.2:ES2	0.111	-1.610	1.923	1505.873	0.872	1.117	0.200	6.842
traitRES_CAT.3:ES2	-2.007	-3.532	-0.624	1601.730	0.004	0.134	0.029	0.536
traitRES_CAT.4:ES2	0.046	-1.304	1.484	1548.543	0.952	1.047	0.272	4.409
traitRES_CAT.1:MORF.f1	-0.191	-1.150	0.779	2000.000	0.667	0.826	0.317	2.178
traitRES_CAT.2:MORF.f1	-0.962	-2.058	0.274	1537.687	0.100	0.382	0.128	1.315
traitRES_CAT.3:MORF.f1	-1.233	-1.714	-0.746	2000.000	0.001	0.291	0.180	0.474
traitRES_CAT.4:MORF.f1	-0.630	-1.342	0.097	1733.746	0.091	0.533	0.261	1.102
traitRES_CAT.1:MORF.f2	-0.342	-1.359	0.757	1867.880	0.540	0.710	0.257	2.132
traitRES_CAT.2:MORF.f2	-0.086	-1.293	1.039	1725.062	0.870	0.917	0.274	2.826
traitRES_CAT.3:MORF.f2	-0.949	-1.557	-0.287	1647.699	0.005	0.387	0.211	0.751
traitRES_CAT.4:MORF.f2	-0.497	-1.299	0.374	1724.647	0.264	0.608	0.273	1.454

traitRES_CAT.1:FAM.LEX.f1	-0.464	-1.105	0.173	1544.699	0.163	0.629	0.331	1.189
traitRES_CAT.2:FAM.LEX.f1	-1.362	-2.201	-0.475	1098.049	0.001	0.256	0.111	0.622
traitRES_CAT.3:FAM.LEX.f1	-0.563	-0.953	-0.141	2000.000	0.004	0.570	0.386	0.868
traitRES_CAT.4:FAM.LEX.f1	-0.897	-1.432	-0.360	1533.184	0.001	0.408	0.239	0.698
traitRES_CAT.1:CUMRES.f1	0.137	-0.573	0.862	1863.065	0.736	1.147	0.564	2.369
traitRES_CAT.2:CUMRES.f1	0.404	-0.443	1.269	1052.338	0.351	1.498	0.642	3.558
traitRES_CAT.3:CUMRES.f1	-0.038	-0.449	0.373	2000.000	0.855	0.962	0.638	1.452
traitRES_CAT.4:CUMRES.f1	0.704	0.025	1.322	1653.229	0.026	2.021	1.026	3.749
traitRES_CAT.1:CUMRES.f2	0.046	-1.058	1.130	2000.000	0.951	1.047	0.347	3.097
traitRES_CAT.2:CUMRES.f2	-0.510	-1.849	0.874	1403.991	0.468	0.601	0.157	2.396
traitRES_CAT.3:CUMRES.f2	-0.222	-0.834	0.433	2000.000	0.467	0.801	0.434	1.542
traitRES_CAT.4:CUMRES.f2	0.503	-0.379	1.401	1616.208	0.296	1.653	0.685	4.060
traitRES_CAT.1:EST11	-1.773	-2.720	-0.924	914.828	0.001	0.170	0.066	0.397
traitRES_CAT.2:EST11	-0.444	-1.288	0.497	1306.433	0.347	0.641	0.276	1.644
traitRES_CAT.3:EST11	-0.077	-0.476	0.314	2000.000	0.691	0.926	0.621	1.368
traitRES_CAT.4:EST11	-0.893	-1.539	-0.229	1130.814	0.008	0.410	0.215	0.796
traitRES_CAT.1:EST21	0.118	-0.641	0.818	1760.448	0.755	1.125	0.527	2.267
traitRES_CAT.2:EST21	-1.510	-2.773	-0.198	672.114	0.021	0.221	0.062	0.820
traitRES_CAT.3:EST21	0.272	-0.124	0.672	2000.000	0.205	1.313	0.884	1.959
traitRES_CAT.4:EST21	0.102	-0.481	0.722	1711.588	0.740	1.107	0.618	2.058
traitRES_CAT.1:EST51	-0.865	-2.059	0.175	1556.623	0.120	0.421	0.128	1.191
traitRES_CAT.2:EST51	-2.139	-3.540	-0.729	856.514	0.001	0.118	0.029	0.483
traitRES_CAT.3:EST51	-0.107	-0.715	0.427	2000.000	0.730	0.899	0.489	1.533
traitRES_CAT.4:EST51	-1.855	-2.756	-1.088	1372.963	0.001	0.156	0.064	0.337
traitRES_CAT.1:GRUPO62	-0.603	-1.757	0.401	1510.496	0.275	0.547	0.173	1.493
traitRES_CAT.2:GRUPO62	-0.520	-1.809	0.640	1549.728	0.396	0.594	0.164	1.897
traitRES_CAT.3:GRUPO62	0.170	-0.412	0.781	2000.000	0.565	1.186	0.663	2.183
traitRES_CAT.4:GRUPO62	0.664	-0.083	1.472	1576.769	0.100	1.943	0.920	4.357
traitRES_CAT.1:GRUPO63	-0.139	-0.987	0.661	1723.035	0.756	0.871	0.373	1.937
traitRES_CAT.2:GRUPO63	-0.504	-1.695	0.654	1029.938	0.405	0.604	0.184	1.923
traitRES_CAT.3:GRUPO63	0.074	-0.342	0.488	1647.641	0.747	1.077	0.711	1.629
traitRES_CAT.4:GRUPO63	-0.488	-1.150	0.204	1587.010	0.163	0.614	0.317	1.226
traitRES_CAT.1:GRUPO64	-0.218	-1.119	0.687	1710.223	0.661	0.804	0.327	1.987
traitRES_CAT.2:GRUPO64	-0.278	-1.364	0.945	1299.757	0.659	0.757	0.256	2.573
traitRES_CAT.3:GRUPO64	-0.008	-0.448	0.499	1813.817	0.970	0.992	0.639	1.647
traitRES_CAT.4:GRUPO64	-0.067	-0.854	0.638	1650.738	0.864	0.935	0.426	1.893
traitRES_CAT.1:GRUPO65	-0.586	-2.041	0.842	1737.677	0.435	0.557	0.130	2.321
traitRES_CAT.2:GRUPO65	-0.591	-2.236	0.893	1723.644	0.466	0.554	0.107	2.443
traitRES_CAT.3:GRUPO65	0.519	-0.169	1.146	2000.000	0.117	1.681	0.845	3.145
traitRES_CAT.4:GRUPO65	-1.207	-2.366	0.050	1367.219	0.036	0.299	0.094	1.051
traitRES_CAT.1:GRUPO66	-0.242	-1.545	1.177	1369.080	0.763	0.785	0.213	3.246
traitRES_CAT.2:GRUPO66	0.752	-0.351	2.051	1864.481	0.237	2.121	0.704	7.777
traitRES_CAT.3:GRUPO66	-0.251	-1.139	0.675	2000.000	0.605	0.778	0.320	1.964
traitRES_CAT.4:GRUPO66	0.138	-0.959	1.388	1836.897	0.797	1.148	0.383	4.005

Capítulo 8

Análisis de respuesta de tiempo hasta el error

En este capítulo se examina la respuesta «tiempo hasta que ocurre un evento», donde «evento» significa el error de concordancia. Se utiliza el marco teórico del análisis de datos de sobrevivencia [Machin et al, 2006; Klein & Moeschberger, 2005; Moore, 2016; Tableman & Kin, 2005]. Se comienza detallando los pasos efectuados de procesamiento junto con algunos conceptos básicos del análisis de sobrevivencia. Luego se examina un modelo de eventos múltiples y otro de riesgos competitivos. Por último, se resumen los resultados obtenidos.

8.1. Nociones de análisis de datos de sobrevivencia.

En los datos de sobrevivencia, la respuesta es una variable aleatoria discreta o continua positiva. Además, las observaciones pueden estar censuradas a derecha o a izquierda si el evento en cuestión no se observa durante el periodo de estudio (ya sea porque pudo haber ocurrido [censura izquierda] o porque podrá ocurrir después de finalizado el estudio [censura derecha]). Si representamos a τ como a la variable aleatoria de tiempo hasta el evento (observado o no), y v como la variable (aleatoria o no) de tiempo hasta la censura, lo que observamos es $T = \min(\tau, v)$, o sea o bien el tiempo en que ocurre el evento desde el inicio del estudio o entrada del individuo al estudio; o bien el tiempo de censura, porque finaliza el estudio sin observarse el evento o porque el individuo sale del estudio en algún punto dado. Además, se define la variable indicadora $\delta = I(\tau < v)$, que toma el valor *uno* si el evento es observado o *cero*, si no lo es (censura). En el presente estudio se establece censura de tipo I, en la cual, los tiempos de censura se hallan pre-establecidos¹. Los «individuos» que sufren el evento «error» son las instancias de concordancia y el tiempo de censura es el tiempo de finalización de las entrevistas de cada sujeto. Se creó la variable «tiempo» numerando las concordancias de cada sujeto. Así por ejemplo, las concordancias van de 1 a 312 para SONIA, 390 para NATI, 464 para JAKO y hasta 691 para MIRKA. Estos fueron los puntos de censura para las instancias de cada sujeto, si el evento error no se observaba: $T = v$. Si se observaba, $T = \tau$, el tiempo de ocurrencia del error según la variable definida.

Por ejemplo, el Cuadro que sigue ilustra los primeros datos de SONIA. La variable «STATUS» indica si la observación fue censurada con «0» u observada con «1». Si hay censura la variable «TIEMPO» tiene como valor al máximo de los datos (312). El individuo «muchos personas» tiene un tiempo de espera hasta el evento de una unidad y el individuo «mucho pensadores», de tres unidades. En cambio, para el individuo «muchos pueblos» no se observó nunca el evento error y está censurada. Se consideró que todos los individuos / instancias comienzan en tiempo «0».

¹La censura puede ser aleatoria, por ejemplo, en la industria se sigue el desempeño de una máquina hasta que un porcentaje de estas falla.

INSTANCIA	ESPAÑOL	STATUS	TIEMPO	... (otras variables)
muchos personas	muchas personas	1	1	...
muchos pueblos	muchos pueblos	0	312	...
mucho pensadores	muchos pensadores	1	3	...
otros autores	otros autores	0	312	...

Cuadro 8.1: Ejemplo de datos de sobrevivencia.

8.1.1. Función de sobrevivencia y de riesgo.

La función de sobrevivencia indica la probabilidad de que un individuo «sobreviva» más allá de un cierto tiempo t (o sea que no experimente el evento hasta t): $S(t) = P(T > t) = 1 - F(t)$, $t \in \mathbb{R} \geq 0$; donde $F(t) = P(T \leq t)$ es la función acumulada, es decir, la probabilidad de que el individuo viva menos o igual que t . La función de sobrevivencia es una curva decreciente que vale 1 en $S(t=0)$ [«todos sobreviven»] y 0 en $S(t=\infty)$ [«nadie sobrevive»].

La función de riesgo (o tasa de riesgo instantánea) indica la probabilidad de sobrevivir un intervalo Δ corto de tiempo adicional sabiendo que el individuo sobrevivió hasta el tiempo t . Muestra el riesgo de experimentar el evento en cada instante y puede tener cualquier forma.

$$h(t) = \lim_{\Delta \rightarrow 0} \frac{P(t < T < t + \Delta \mid T > t)}{\Delta}$$

Si bien esta función es más fácil de interpretar, la primera es más sencilla de estimar. Una forma de estimar $S(t)$ a partir de datos muestrales, $\hat{S}(t)$, es por medio del método de Kaplan-Meier. Como los tiempos de sobrevivencia se consideran independientes, el estimador puntual es el producto de las probabilidades (condicionales) de sobrevivir más que un tiempo t sabiendo que se sobrevivió hasta $t-1$: $P(T > t) = P_1 P_2 \dots P_{t-1}$; donde cada probabilidad se define según la regla del producto, por ejemplo: $P_{t=2} = P(T > 2 \mid T > 1)P(T > 1)$, y se estima como: $\hat{p}_{t_i} = \frac{n-d_i}{n_i} = 1 - \frac{d_i}{n_i}$. En la expresión, n_i es el número de sujetos es riesgo en el tiempo t_i y d_i es el número de individuos que «mueren» en ese tiempo t_i . En resumen:

$$\hat{S}(t) = \prod_{t_i < t} \left(1 - \frac{d_i}{n_i}\right)$$

La varianza del estimador es: $\widehat{Var}(\log[-\log \hat{S}(t)]) = \frac{1}{[\log(\hat{S}(t))]^2} \sum_{t_i < t} \frac{d_i}{n_i(n_i - d_i)}$;

y su desvío típico: $\widehat{ES}(\log[-\log \hat{S}(t)]) = \sqrt{\widehat{Var}(\hat{S}(t))}$.

El intervalo de confianza puntual con nivel aproximado del 95% en el intervalo $[0, 1]$ rezará:

$$\log[-\log \hat{S}(t)] \pm 1,96 \times \widehat{ES}(\log[-\log \hat{S}(t)])$$

La función de riesgo $h(t)$ puede calcularse a partir de la función de supervivencia como la derivada de su logaritmo, cambiando el signo: $h(t) = -[\ln(\hat{S}(t))]'$. Ahora bien, la función de riesgo acumulada es la integral de $h(t)$ entre 0 y t : $H(t) = \int_0^t h(u) du$; ya que la integral cancela la derivada: $H(t) = -\ln(S(t))$. Con la derivada de esta función de riesgo acumulada se obtiene la tasa instantánea de riesgo. Es decir que la pendiente de su estimador $\hat{H}(t) = -\ln(\hat{S}(t))$ resulta ser un estimador de la tasa de riesgo para cada tiempo t .

8.1.2. El modelo semi-paramétrico de Cox.

Defínase primero una función de riesgo «promedio» para el individuo típico denotada por $h_0(t)$. Luego, se puede especificar la función de riesgo para un individuo en particular, denotada por $h(t)$, y relacionarla con aquella promedio mediante un ratio de riesgo o *hazard ratio* (HR):

$$HR : \frac{h(t)}{h_0(t)} = \psi$$

Obsérvese que se ha escrito ψ como una constante que *no depende* del tiempo, es decir que, no obstante $h(t)$ y $h_0(t)$ si lo hagan, su ratio se mantiene igual a lo largo del tiempo. Este es el supuesto de HR proporcional. También se puede escribir la expresión anterior como: $h(t) = h_0(t) \psi$, o sea que la función de riesgo de un individuo particular, $h(t)$, se define como el factor ψ que multiplica a la función de riesgo del individuo «promedio», $h_0(t)$. Si no hay covariables en el modelo, entonces $\psi = \exp(0) = 1$; en cambio, si existieran, $\psi = \exp(\mathbf{x}^T \boldsymbol{\beta})$. El modelo de Cox proporcional para el i -ésimo individuo, dadas las covariables, se escribe pues como:

$$h_i(t | \mathbf{x}) = h_0(t) \exp(\mathbf{x}_i^T \boldsymbol{\beta}) = h_0 \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)$$

Por ejemplo, si tuviéramos una sola covariable binaria, $x = \{0, 1\}$, que compara un grupo control con otro tratado, el *hazard* del grupo control sería $h_C(t) = h_0(t) \exp(\beta_1 (x = 0)) = h_0(t)$ y el del grupo tratado: $h_T(t) = h_0(t) \exp(\beta_1 (x = 1))$. El *hazard ratio* del grupo tratado relativo al control se escribe del modo siguiente:

$$HR : \frac{h_T(t)}{h_C(t)} = \frac{h_0(t) \exp(\beta_1)}{h_0(t)} = \exp(\beta_1) = \psi$$

Nótese que solamente la función $h_0(t)$ depende del tiempo, pero se cancela, por lo tanto el HR resulta proporcional. Entonces, el exponencial de los coeficientes del modelo indican el HR de la variable j -ésima, ajustada por las demás variables. Es el factor ψ por el cual se multiplica la función de riesgo si la j -ésima variable aumenta una unidad, manteniendo constantes los valores de las otras covariables. Se tienen las siguientes situaciones: (i) $\psi > 1$, entonces $h_T(t) > h_C(t)$, el riesgo basal *aumenta* en proporción de ψ cuando se usa el tratamiento en lugar del control; (ii) $\psi < 1$, entonces $h_T(t) < h_C(t)$, el riesgo basal *disminuye* en proporción de ψ cuando se usa el tratamiento en lugar del control; (iii) $\psi = 1$, entonces $h_T(t) = h_C(t)$, el riesgo basal no cambia.

El logaritmo del hazard es una función lineal de las variables explicativas:

$$\ln[h_i(t | \mathbf{x})] = \ln[h_0(t)] \mathbf{x}_i^T \boldsymbol{\beta} = \ln[h_0(t)] (\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)$$

Donde: \mathbf{x} es el vector de covariables para el individuo i -ésimo y $\boldsymbol{\beta}$, el vector de coeficientes. Cada predictor actúa de forma *multiplicativa* sobre el *hazard basal*. Además $h_0(t)$ puede asumir cualquier forma. Los coeficientes betas se estiman por máxima (log)verosimilitud parcial. Denótese el *hazard ratio* (HR) del i -ésimo individuo como $\psi_i = \exp(\mathbf{x}_i^T \boldsymbol{\beta})$ y al conjunto de riesgo R_j como aquel conjunto de individuos en riesgo de experimentar el evento en el tiempo j -ésimo. Como los tiempos de falla se presumen independientes, la verosimilitud es un producto de términos, uno para cada tiempo de falla. En cada término, el denominador es la suma de los riesgos de sufrir el evento de los $k = 1, \dots, K$ individuos presentes en el conjunto R_j . El riesgo del k -ésimo individuo es: $h_k(t_j) = h_0(t_j) \psi_k$. El numerador es el riesgo del individuo de R_j que sufrió efectivamente el evento en el tiempo j -ésimo: $h_i(t_j) = h_0(t_j) \psi_i$. Formalmente:

$$\mathcal{L}(\boldsymbol{\beta}) = \prod_{j=1}^D \frac{h_0(t_j) \psi_i}{\sum_{k \in R_j} h_0(t_j) \psi_k} = \prod_{j=1}^D \frac{\psi_i}{\sum_{k \in R_j} \psi_k}$$

Donde D es el índice máximo de los tiempos de falla (ordenados). Se denomina verosimilitud «parcial» porque solo se toman en cuenta los tiempos de falla, y no aquellos en los que se producen censuras. Nótese que la *hazard* basal $h_0(t_j)$ se cancela, por lo cual el modelo de Cox *no* cuenta con un coeficiente de ordenada al origen. Tomando logaritmo, derivando con respecto a cada coeficiente beta e igualando a cero se obtienen las ecuaciones de *score* que luego se resuelven mediante optimización numérica para obtener los coeficientes estimados. Para sacar la varianza de cada coeficiente es preciso calcular la información observada: $I(\hat{\boldsymbol{\beta}}) = -\log \mathcal{L}^{(2)}(\hat{\boldsymbol{\beta}})$ (menos la segunda derivada de la log-verosimilitud parcial evaluada en $\hat{\boldsymbol{\beta}}$), la cual es una medida de la curvatura de la función de verosimilitud en $\hat{\boldsymbol{\beta}}$. A mayor curvatura,

más información y por ende menos varianza. A menor curvatura, menos información y, por ende, más varianza. Como la relación es inversa, la varianza de $\hat{\beta}$ será $\widehat{Var}(\hat{\beta}) = I(\hat{\beta})^{-1}$ y su error típico: $SE(\hat{\beta}) = \sqrt{I(\hat{\beta})^{-1}}$. Este último se usa para construir un test z como $Z = \hat{\beta}/SE(\hat{\beta})$ que rechazará $H_0 : \beta = 0$ si $|Z| > z_{\alpha/2}$ o $Z^2 > \chi_{\alpha, df=1}^2$.

Por último, cabe mencionar que el modelo puede estratificarse. En tal caso habrá un *hazard* basal (posiblemente de forma diferente) para *cada* estrato:

$$h_i(t | \mathbf{x}, V = j) = h_{0j}(t) \exp(\mathbf{x}_i^T \beta) = h_{0j} \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p)$$

Nótese que en la expresión anterior los coeficientes de una variable no dependen del estrato j . El *HR* de las variables es el mismo para todos los estratos. En cambio, se puede pensar que alguna variable (o todas) posee(n) *HR* diferentes según el estrato. Por ejemplo, en el modelo que sigue la variable X_2 interactúa con el nivel de estrato.

$$h_i(t | \mathbf{x}, V = j) = h_{0j}(t) \exp(\mathbf{x}_i^T \beta) = h_{0j} \exp(\beta_1 x_1 + \beta_2 x_2 + \beta_{2:V} x_2 : V)$$

Si el modelo tuviera todos los coeficientes diferentes para cada estrato, las estimaciones serían similares a tomar los individuos de cada estrato por separado, y ajustar un modelo por estrato.

8.1.3. Residuos.

Los residuos permiten evaluar: (i) el supuesto de proporcionalidad; (ii) posibles valores atípicos; (iii) posibles valores atípicos que son además influyentes. El primer tipo a considerar es el de los residuos *martingala*. Básicamente son, para cada individuo, la diferencia entre el indicador de censura $\delta_i = I(\tau < \nu) = \{0 : \text{censura}, 1 : \text{evento}\}$ y el valor esperado de dicho indicador bajo el modelo de *hazards* proporcionales. Reza: $m_i = \delta_i - \hat{H}_0(t_i) \exp(x_i^T \beta)$, donde: $H_0(t_i) = \sum_{t_j \leq t_i} h_0(t_j)$ es la *hazard* basal acumulada. Estos residuos suman 0 y tienen rango $[-\infty, 1]$ pero con distribución asimétrica. En cambio, los residuos de *devianza* poseen distribución simétrica y de esperanza cero. Su fórmula es:

$$d_i = \text{sign}(m_i) \{-2[m_i + \delta_i \log(\delta_i - m_i)]\}^{\frac{1}{2}}$$

Donde m_i es el residuo *martingala* y δ_i , el indicador de censura. La suma de estos residuos al cuadrado es el valor del test de cociente de verosimilitud entre el modelo con las covariables y el modelo reducido nulo. Residuales de devianza en valor absoluto > 2 indican valores atípicos, con una tolerancia del 5%. Los residuos de *martingala* se usan para examinar la forma funcional de las variables continuas, es decir, el supuesto de linealidad.

Los residuos «delta betas» (*dfbetas*) marcan cuáles individuos ejercen mayor influencia sobre la estimación de los coeficientes. Hay una serie de residuos para *cada* coeficiente estimado. Para el i -ésimo individuo se calcula la diferencia entre el valor estimado del coeficiente usando todas las observaciones y su valor estimado excluyendo al individuo en cuestión. Si la diferencia es mayor al 10% se considera un individuo influyente. Los residuos se pueden estandarizar dividiéndolos por el error típico del coeficiente. En este caso, se consideran influyentes aquellos con desvíos superiores a ± 2 .

Por último es necesario mencionar a los residuos de *Schoenfeld*. *Cada* coeficiente tiene su serie de residuos. Los residuos de *Schoenfeld* son los términos individuales de la ecuación de score para cada coeficiente. Se definen como la diferencia entre el valor de la covariable x_i para el i -ésimo individuo y su valor esperado, $E[x_i]$. Formalmente:

$$\hat{r}_i = x_i - E[x_i] = x_i - \left[\sum_{k \in R_j} x_k p(\beta, x_k) \right]; p(\beta, x_k) = \frac{\exp(x_k^T \beta)}{\sum_{k \in R_j} \exp(x_k^T \beta)}$$

Obsérvese que $E[x_i]$ es la suma de los valores de la covariable para los individuos dentro del conjunto de riesgo en el tiempo j -ésimo multiplicado por un peso. Dicho peso es la probabilidad de seleccionar

un individuo particular del conjunto de riesgo en el tiempo j -ésimo. Los residuos escalados son: $r_j^* = r_j \times d \times \text{var}(\hat{\beta})$; $d = \sum(\delta = 1)$; los residuos multiplicados por el total de eventos, multiplicados por la varianza del coeficiente.

Si el supuesto de proporcionalidad se cumple, los residuos deberían ubicarse aleatoriamente en torno al cero. Ahora bien, decir que el $HR : \exp(\hat{\beta})$ no es proporcional es lo mismo que decir que el coeficiente depende del tiempo: $HR : \exp(\hat{\beta}(t))$. Grambsch and Therneau (1994) consideraron definir el coeficiente dependiente del tiempo como: $\beta(t) = \beta + \theta g(t)$, donde $g(t)$ es una función del tiempo previamente definida para modelar la dependencia temporal. Si el HR resulta ser función del tiempo, entonces la esperanza de los residuos escalados es aproximadamente: $E[r_i^*] \approx \beta + \beta(t)$. Por lo tanto el coeficiente $\beta(t)$ se podría estimar como: $\hat{\beta}(t) = r_i^* + \beta$. Si se grafica $\hat{\beta}(t)$ versus los tiempos de falla ordenados, se revela la forma funcional de la dependencia temporal. Además es posible testear que $\theta = 0$, ajustando una recta al gráfico y evaluando la significatividad estadística de la pendiente. Si se rechaza, el coeficiente en cuestión depende del tiempo. Otro modo, más simple, de solucionar el problema de la no proporcionalidad consiste en estratificar según una partición de los datos en intervalos temporales disjuntos y estimar coeficientes específicos para cada estrato.

8.2. Expectativa de análisis.

Se desea encontrar las variables que influyen en el *riesgo* de que una concordancia tenga un error ($\delta = 1$) de modo *recurrente*; o bien en el *riesgo* de que una concordancia tenga un error de un tipo determinado. Se esperan los siguientes resultados de los predictores, en cuanto a *aumentar* con menor (\uparrow) / mayor ($\uparrow\uparrow$) tamaño de efecto, o *disminuir* (\downarrow) el riesgo de cometer errores de concordancia. Se usaron las siguientes predictoras discretas: Fabs.SC.f, MORF.f, STEM.f, MOD, ES, ANIM, GRAMS, FAM.LEX.f, IMA.CONC.f, LDA, EST1, EST2, EST3, EST4, EST5, EST6, EST7.

Predictor	Efecto	dirección / tamaño
MOD	MOD(1) - MOD(0)	\uparrow
	MOD(2) - MOD(0)	\uparrow
	MOD(3) - MOD(0)	\uparrow
Fabs.SC.f	Fabs.SC.f(1) - Fabs.SC.f(0)	\downarrow
MORF.f	MORF.f(1) - MORF.f(0)	\uparrow
	MORF.f(2) - MORF.f(0)	$\uparrow\uparrow$
FAM.LEX.f	FAM.LEX.f(1) - FAM.LEX.f(0)	\downarrow
EST1	EST1(1) - EST1(0)	\downarrow
EST2	EST2(1) - EST2(0)	\downarrow
EST3	EST3(1) - EST3(0)	\uparrow
EST4	EST4(1) - EST4(0)	\downarrow
EST5	EST5(1) - EST5(0)	\downarrow
EST6	EST6(1) - EST6(0)	\downarrow
EST7	EST7(1) - EST7(0)	\uparrow
ANIM	ANIM(1) - ANIM(0)	\uparrow
ES	ES(1) - ES(0)	\uparrow
	ES(2) - ES(0)	$\uparrow\uparrow$
LDA	LDA(1)-LDA(0)	\uparrow
GRAMS	GRAMS(1)-GRAMS(0)	\uparrow
STEM.f	STEM.f(1)-STEM.f(0)	\uparrow
IMA.CONC.f	IMA.CONC.f(1) - IMA.CONC.f(0)	\downarrow

Cuadro 8.2: Expectativas del análisis. Leyenda: (\uparrow) = riesgo aumenta; (\downarrow) = riesgo disminuye; ($\uparrow\uparrow/\downarrow\downarrow$) = mayor tamaño de efecto.

Según el Cuadro, se espera que el riesgo *aumente* con: (i) artículos indefinidos, determinantes y adjetivos; (ii) concordancias a larga distancia (el riesgo de error crece con: distancia larga / sin subordinada > distancia entre enunciados / con subordinada); (iii) *-e-* epentética, con riesgo de error creciente si está en un término o en ambos; (iv) concordancias de más de dos términos; (v) controlador animado; (vi) TYPE infrecuente; (vii) similaridad baja (alta distancia) entre las raíces léxicas del español y el italiano; (viii) similaridad media (media distancia) entre las terminaciones del español y el italiano, según el tamaño de efecto *baja > media*; (ix) controlador infrecuente y/o no familiar; (x) controlador de baja imaginabilidad y/o concretud; (xi) las «estrategias» EST3 y EST7 porque identifican casos de difícil producción para los italianos.

8.3. El modelo de eventos múltiples.

8.3.1. Pre-procesamiento y marco teórico.

Los modelos de esta sección tienen que ver con eventos que pueden repetirse. Las instancias se consideraron TOKENS de un TYPE, o sea que el [TYPE = «los profesores»] puede aparecer como [TOKEN = «lo profesores», «los profesor», ...]. Una concordancia / individuo (TYPE) puede registrar un error en el curso de seguimiento del sujeto o bien varios errores. Para reflejar esto, los datos se dispusieron de manera de que cada fila representara un TYPE y la variable TIEMPO se reorganizó para expresar los intervalos entre los cuales se producían los errores. Por ejemplo:

ID	TYPE	TOKEN	STATUS	T_START	T_STOP	... (otras variables)
1	barrios amplios	barrios amplio	1	0	115	...
1	barrios amplios	–	0	115	312	...
2	los profesores	los profesor	1	0	212	...
2	los profesores	los profesore	1	212	240	...
2	los profesores	–	0	240	312	...
3	alemanes fieles	alemanes fideles	0	0	312	...

Cuadro 8.3: Ejemplo de datos de sobrevivencia: modelo de eventos repetidos.

En este caso el primer TYPE, «barrios amplios», tiene dos intervalos: $(0, 315]$ y $(315, 312]$; el primero desde el cero hasta el tiempo 315, donde se produce el error (STATUS = 1) y el segundo desde 316 a 312, siendo este último el tiempo de censura (STATUS = 0). El TYPE «los profesores» registra dos TOKEN, el primero es un error en el tiempo 212 (intervalo $(0, 212]$) y el segundo con error en el tiempo 240 (intervalo $(240, 312]$); en el último periodo no se registra error, por tanto, va de 241 a 312, el tiempo de censura. Por último, el TYPE «alemanes fieles» no registra error², por tanto, el intervalo va del cero hasta el tiempo de censura³. Es necesario notar que cada TYPE tiene al menos un TOKEN. Por ejemplo, «los profesores» registra para SONIA ocho instancias, de las cuales solamente tres son errores, que es el tiempo que se registra. En cuanto a la relación entre TYPE y TOKEN, 247 TYPES de 1152 tienen entre 2 y 32 TOKENS, y 127 solamente 2. Todas las covariables se asumen *fijas* en el tiempo. O sea que, por ejemplo, una covariable específica *repite* sus valores a lo largo de los eventos dentro de un mismo TYPE. En total hubo 1813 observaciones y 473 eventos de error.

Siguiendo a Hosmer, Lemeshow & May (2008, cap. 9.2), se definieron los siguientes modelos:

Andersen-Gill [AG].

²«fideles» es un error por no conocer la base léxica, por eso no se consideró error de concordancia [It. *fedele*, «fiel»].

³También es posible utilizar el tiempo desde que el último evento ocurrió. En tal caso, en el Cuadro 2 se crea una variable TG («TIME GAP») = T_STOP - T_START. Entonces los intervalos comienzan ahora en cero y terminan en TG. Por ejemplo el sujeto 1 tendrá: $TG = (115 - 0, 312 - 115)$, y los intervalos serán: $(0, 115]$, $(0, 197]$. No se usará esta escala de tiempo.

El modelo estima una *hazard* basal común para todos los eventos y también coeficientes globales, que no dependen de los eventos. Se supone independencia entre los eventos dentro de cada sujeto⁴. La diferencia con el modelo clásico de Cox expuesto en 8.1.2 reside en que un sujeto contribuye al conjunto de riesgo para un tiempo de evento *j*-ésimo específico (R_j) siempre y cuando el sujeto se halle bajo observación en el tiempo que el evento ocurre. Esto quiere decir que los individuos / concordancias (índice «ID» en Cuadro 3) se encuentran en el conjunto de riesgo desde el tiempo 0 hasta el tiempo de censura (312, para SONIA). En cambio en el modelo clásico el individuo deja el conjunto de riesgo cuando el evento se produce. El modelo se escribe como sigue:

$$h_{ik}(t | \mathbf{x}) = Y_{ik}(t) h_0(t) \exp(\mathbf{x}_i^T \boldsymbol{\beta})$$

donde $h_{ik}(t | \mathbf{x})$ representa la función de riesgo de error del *k*-ésimo evento dentro del *i*-ésimo individuo (TYPE) en el tiempo *t*; $h_0(t)$ es el *hazard* basal común para todos los eventos; \mathbf{x}_i es el vector de covariables para el *i*-ésimo individuo; $\boldsymbol{\beta}$ es el vector de coeficientes y $Y_{ik} = \{0, 1\}$ indica cuándo el *i*-ésimo individuo se halla bajo observación⁵.

El modelo podría incluir estratos (que no sean eventos), y por ende, habría una *hazard* basal por estrato:

$$h_{ik}(t | \mathbf{x}, V = j) = Y_{ik}(t) h_{0j}(t) \exp(\mathbf{x}_i^T \boldsymbol{\beta})$$

Prentice, Williams and Peterson [PWP].

El modelo estratifica por evento; por lo tanto, hay una *hazard* basal diferente para cada estrato. Todos los individuos están en riesgo para el primer estrato pero solo aquellos con un evento en el estrato precedente permanecen en riesgo en el estrato siguiente. Por ejemplo, en el Cuadro 3, el individuo 2 no estará en riesgo de sufrir el segundo evento (error) hasta que no haya sufrido el primer evento. Se pueden estimar coeficientes específicos para cada estrato / evento. En la práctica se deben limitar la cantidad de eventos para estratificar; ya que, de no hacerse, la cantidad de individuos en riesgo para los estratos posteriores se vuelve muy pequeña y las estimaciones no son confiables. El modelo se escribe como el anterior pero con un índice adicional en el *hazard* basal, indicando que es diferente para cada estrato *k*-ésimo.

$$h_{ik}(t | \mathbf{x}, V = k) = Y_{ik}(t) h_{0k}(t) \exp(\mathbf{x}_i^T \boldsymbol{\beta})$$

El supuesto de eventos (errores) independientes dentro de cada individuo / TYPE resulta irreal. Por lo tanto, para dar cuenta de la posible correlación dentro de cada TYPE, se utiliza un estimador «sándwich» para la varianza de los coeficientes. Dicho estimador fue propuesto por Lin & Wei (1989) : $\hat{R}(\boldsymbol{\beta}) = \widehat{Var}(\hat{\boldsymbol{\beta}}) [\hat{\mathbf{L}}^T \hat{\mathbf{L}}] \widehat{Var}(\hat{\boldsymbol{\beta}})$, donde $\widehat{Var}(\hat{\boldsymbol{\beta}})$ es la varianza del coeficiente estimado basado en la inversa de la información observada mencionada en 1.2 y $\hat{\mathbf{L}}$ es una matriz $n \times p$ con *n* filas como individuos y *p* covariables, y cuyas filas contienen residuos de *score* (ver: Hosmer, Lemeshow & May (2008), cap. 6). Lo importante es que se trata de una varianza que ajusta por datos agrupados (varios eventos en individuos).

Modelo de fragilidad compartida («shared frailty») [Frailty].

El modelo asume que los sujetos (TOKEN) pueden estar expuestos a diferentes niveles de riesgo, ser unos más (menos) «frágiles» que otros, debido al efecto de covariables no observadas. Si los sujetos (concordancias TOKEN) que forman parte un grupo (concordancia TYPE) comparten el mismo nivel de fragilidad, el modelo se denomina de «fragilidad compartida». Llámese u_i a la fragilidad del TYPE *i*-ésimo, que modela estos efectos de covariables no observadas. Dichos u_i son entonces similares a efectos aleatorios de

⁴Es similar a un proceso de conteo *Poisson*, en el cual, el número de eventos de los intervalos disjuntos de tiempo son independientes entre sí.

⁵Bajo el modelo clásico $Y_{ik} = 0$ luego de que el evento se produce y el individuo deja el conjunto de riesgo. En cambio aquí $Y_{ik} = 1$, haciendo que los individuos permanezcan en el conjunto de riesgo.

un modelo mixto. El modelo incluye un coeficiente «fragilidad» que debería resultar significativo. La distribución de los efectos aleatorios se asume *Gamma* con esperanza $E[\mathbf{u}] = 1$ y varianza $Var(\mathbf{u}) = \theta$:

$$g(u_i, \theta) = \frac{u_i^{\frac{1}{\theta}-1} e^{-\frac{u_i}{\theta}}}{\Gamma(\frac{1}{\theta}) \theta^{\frac{1}{\theta}}}$$

Si un sujeto (TOKEN) de un grupo (TYPE) tiene $u_i > 1$, entonces es «frágil», es decir, con más riesgo de sufrir el evento «error». En cambio un sujeto con $u_i < 1$ será más «fuerte», con menos riesgo de sufrir dicho evento. La correlación intra-clase se define como: $IC = \theta / (2 + \theta)$. El modelo se escribe como riesgo del k -ésimo TOKEN del grupo (TYPE) i -ésimo *condicional* a las covariables y al efecto aleatorio u_i del i -ésimo grupo (TYPE):

$$h_{ik}(t | \mathbf{x}, u_i) = h_0(t) \exp(\mathbf{x}_{ik}^T \beta + u_i); u_i \sim \text{Gamma}(\theta)$$

Además, podría incluir estratos, en cuyo caso se agregaría un índice a la *hazard* basal para el j -ésimo estrato:

$$h_{ik}(t | \mathbf{x}, u_i, V = j) = h_{0j}(t) \exp(\mathbf{x}_{ik}^T \beta + u_i); u_i \sim \text{Gamma}(\theta)$$

8.3.2. Selección de modelos.

A modo exploratorio, la gráfica siguiente muestra la *hazard* basal estratificada por alumno. Las curvas tienen diferente longitud porque la censura fue diferente para cada alumno. MIRKA (ID = 4) y NATI (ID = 2) perforan el 75% de probabilidad de sobrevivencia (no sufrir «error»)⁶. El mejor es, por supuesto, JAKO (ID = 3), que mantiene alta la probabilidad de no sufrir errores; seguido por SONIA (ID = 1). A ojo de buen cubero, no parece haber diferencias entre las curvas de riesgo de JAKO y MIRKA; pero sí entre JAKO y SONIA / NATI. Por otro lado, también se observan diferencias entre las curvas de SONIA y NATI, al menos a partir de $t = 200$. Además, existen diferencias entre las curvas de MIRKA y SONIA / NATI. La probabilidad de sobrevivencia baja con el tiempo, pero dicha probabilidad parece caer más rápido en NATI y SONIA para los primeros 200 tiempos. Es decir que NATI comete errores más aceleradamente que MIRKA, lo cual es esperable ya que MIRKA es nivel C1 y NATI, B1. El test de *log-rank* entre los niveles de ID resulta significativo ($p < 0,0001$). Lo dicho sugiere ajustar modelos que incluyan *hazards* basales diferentes por alumno.

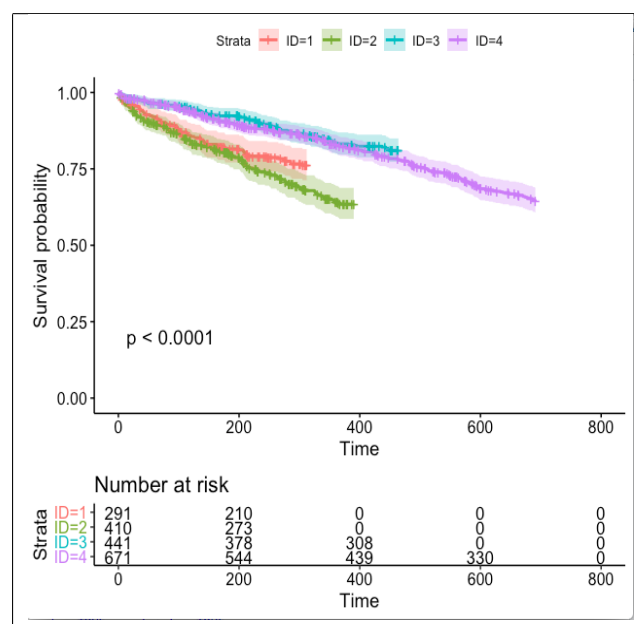


Figura 8.1: Hazards basales por alumno. 1 = SONIA, 2 = NATI, 3 = JAKO, 4 = MIRKA.

⁶Las probabilidades de sobrevivencia son superiores al 60% porque el tiempo de censura fue, por alumno, igual para todos los TYPE. Esto evitó que salieran del conjunto de riesgo para tiempos tardíos.

Se ajustaron pues los siguientes modelos:

1. AG: modelo AG con varianza «sándwich» por TYPE.
2. AG-STRATA: modelo AG con varianza «sándwich» por TYPE y estratos por alumno: $V = 1, 2, 3, 4$.
3. PWP: modelo PWP con estratos por eventos, reducidos a 3 eventos como máximo para evitar problemas de estimación ($V = 1, 2, 3$). varianza «sándwich» por TYPE.
4. Frailty: modelo de «shared frailty» con distribución gamma de efectos aleatorios por TYPE.
5. Frailty-STRATA: modelo de «shared frailty» con distribución gamma de efectos aleatorios por TYPE; y estratos por alumno: $V = 1, 2, 3, 4$.

El Cuadro siguiente muestra la comparación de los modelos según las medidas de información AIC, delta y pesos de AKAIKE. Los dos mejores modelos son los que incluyen estratificación por alumno. El de «fragilidad compartida» es ligeramente mejor que el de AG. Sin embargo, en el primero el coeficiente de fragilidad no fue significativo ($p = 0,37$) y la varianza de los u_i fue muy baja ($\theta = 0,021$). Por lo tanto, se eligió el modelo AG estratificado por alumno, con varianza «sándwich» por individuo (TYPE).

	logLik	AIC	delta	weight
model.frailty.strata	-2727.688	5516.660	0.000	0.667
model.AG.strata	-2738.023	5518.046	1.387	0.333
model.PWP	-2993.258	6028.516	511.856	0.000
model.frailty	-3245.910	6556.913	1040.253	0.000
model.AG	-3258.391	6558.782	1042.123	0.000

Cuadro 8.4: Selección de modelos. logLik: verosimilitud del modelo, AIC: AKAIKE, delta: delta de AKAIKE, weight: pesos de AKAIKE

Se ajustó el modelo elegido con todas las predictoras mencionadas. Fueron $2^{17} = 131072$ modelos, jerarquizados mediante la medida de información AIC (como: $\frac{n}{p} = \frac{1813}{17} \approx 106 > 40$, no se usó la versión AICc corregida por tamaño muestral). Luego se examinó la frecuencia de las predictoras en el conjunto completo de modelos, que da un panorama de la incerteza por la selección. A continuación se redujo la cantidad de modelos al subconjunto «de confianza» con la regla $\frac{W(i)}{W(1)} > \frac{1}{8}$. Sobre dicho subconjunto se llevó a cabo un promedio de coeficientes con la varianza calculada con «full average». El Cuadro 5 muestra que las variables con porcentaje de elección arriba del 80% son: ANIM, EST1, EST5, MORF.f. Las mismas variables son las que resultan significativas en los coeficientes promediados (Cuadro 6).

8.3.3. Modelo final: chequeo de supuestos y resultados.

El modelo AG estratificado por alumno, con varianza «sándwich» por individuo (TYPE) con las predictoras ANIM, EST1, EST5, MORF.f, se escribe como:

$$h_{k(i)}(t | \mathbf{x}, V = j) = Y_{k(i)}(t) h_{0j}(t) \exp(\beta_1 \text{ANIM} + \beta_2 \text{EST1} + \beta_3 \text{EST5} + \beta_4 \text{MORF.f1} + \beta_5 \text{MORF.f2})$$

$$j = 1, \dots, 4; i = 1, \dots, 473; k = 1, \dots, 15; k(i) = 1, \dots, 1813; t = 1, \dots, 691$$

donde $h_{ik}(t | \mathbf{x})$ representa la función de riesgo de error del k-ésimo evento ($k = 1, \dots, 15$) anidado en el i-ésimo individuo (TYPE) [$i = 1, \dots, 473$] en el tiempo t ($t = 1, \dots, 691$); $h_{0j}(t)$ es el hazard basal específico para cada alumno ($j = 1, \dots, 4$); y $Y_{ik} = \{0, 1\}$ indica cuándo el i(k)-ésimo individuo se halla bajo observación. Los efectos fijos (marginales) son: (i) β_1 es el efecto de ANIM; (ii) β_2 es el efecto de EST1; (iii) β_3 es el efecto de EST5; (v) β_4 es el efecto de MORF.f, nivel 1; (vi) β_5 es el efecto de MORF.f, nivel 2 [referencia de MORF.f es «0»].

	names	x
1	ANIM	1.00
2	EST1	0.99
3	EST5	0.98
4	MORF.f	0.89
5	ES	0.74
6	IMA.CONC.f	0.59
7	FAM.LEX.f	0.57
8	EST4	0.50
9	MOD	0.45
10	LDA	0.44
11	STEM.f	0.39
12	EST3	0.39
13	Fabs.SC.f	0.37
14	EST7	0.34
15	EST2	0.32
16	EST6	0.30
17	GRAMS	0.29

Cuadro 8.5: Importancia Relativa de las predictoras.

La Figura 2, a la izquierda, muestra los residuos «dfbetas» escalados. Se observa que ningún gráfico supera dos desvíos en valor absoluto. Por lo tanto, se concluye que no hay observaciones influyentes. Por otra parte, en la derecha se muestran los residuos de devianza. El 5,6% supera en valor absoluto los dos desvíos (103 observaciones). Está en el borde del cinco por ciento esperado de valores atípicos.

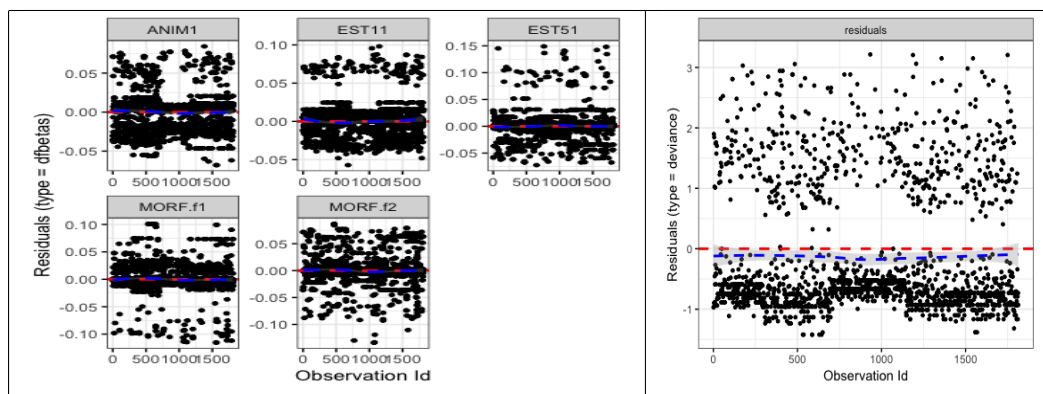


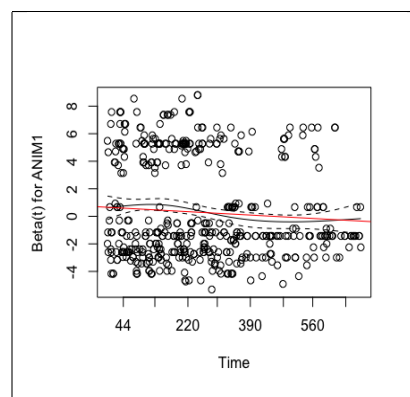
Figura 8.2: Residuos: Izquierda: DFBETAS escalados; Derecha: residuos de devianza.

El Cuadro 7 muestra los p-valores del test de $\theta = 0$ para $\beta(t) = \beta + \theta g(t)$, donde $g(t)$ es una función del tiempo definida según: (i) $g(t) = \text{rank}(t)$; (ii) $g(t) = \hat{S}_{km}(t)$; (iii) $g(t) = t$; (iv) $g(t) = \log(t)$. Según todas las funciones del tiempo, ANIM resulta significativo. Por lo tanto, no se cumple el supuesto de *hazards* proporcionales para dicha variable. La Figura 3 muestra los residuos de *Schoenfeld* versus el tiempo. La recta roja es $\beta(t)$. Se observa que el coeficiente tiene una leve pendiente negativa a medida que pasa el tiempo.

	Estimate	Std. Error	z value	Pr(> z)
ANIM1	0.412	0.115	3.596	0.000
ES1	0.244	0.250	0.976	0.329
ES2	-0.604	0.568	1.063	0.288
EST11	-0.495	0.118	4.198	0.000
EST41	-0.332	0.355	0.936	0.349
EST51	-0.598	0.185	3.224	0.001
FAM.LEX.f1	-0.107	0.108	0.989	0.323
IMA.CONC.f1	-0.114	0.117	0.978	0.328
MORF.f1	-0.519	0.174	2.987	0.003
MORF.f2	-0.491	0.279	1.760	0.078
STEM.f1	0.037	0.086	0.426	0.670
LDA1	-0.087	0.174	0.501	0.617
EST71	0.033	0.121	0.275	0.783
MOD1	-0.194	0.372	0.521	0.602
MOD2	0.046	0.096	0.478	0.633
MOD3	0.000	0.074	0.004	0.997
EST31	0.109	0.302	0.360	0.719
EST21	0.009	0.052	0.177	0.859
Fabs.SC.f1	-0.015	0.058	0.264	0.792
GRAMS1	0.003	0.032	0.097	0.923
EST61	0.007	0.063	0.104	0.917

Cuadro 8.6: Promedio de los coeficientes con FULL AVERAGE.

	df	rank.chi2	rank.p	km.chi2	km.p	time.chi2	time.p	log.time.chi2	log.time.p
MORF.f1	1.0000	0.1430	0.7054	0.0087	0.9257	0.0002	0.9881	0.2011	0.6538
MORF.f2	1.0000	0.7136	0.3983	0.5088	0.4757	0.5284	0.4673	1.1555	0.2824
ANIM1	1.0000	11.4895	0.0007	11.0026	0.0009	11.0548	0.0009	7.2430	0.0071
EST11	1.0000	0.4291	0.5125	0.2645	0.6070	0.3078	0.5790	1.1808	0.2772
EST51	1.0000	1.5289	0.2163	1.3443	0.2463	1.3744	0.2411	1.4356	0.2309
GLOBAL	5.0000	11.8179	0.0374	11.2691	0.0463	11.2809	0.0461	7.7530	0.1704

Cuadro 8.7: Test para $\theta = 0$, según diferentes funciones del tiempo: estadístico de chi cuadrado y p valorFigura 8.3: Residuos de Schoenfeld versus el tiempo. La recta roja es $\beta(t)$.

Para acomodar esto, el modelo final incluyó el coeficiente de ANIM escrito como $\beta(t) = \beta + \theta g(t)$, donde $g(t) = t$, es decir dependiendo linealmente del tiempo. Se escribió como:

$$h_{k(i)}(t | \mathbf{x}, V = j) = Y_{k(i)}(t) h_{0j}(t) \exp([\beta_0 + \beta_1 (g(t) = t)] ANIM + \beta_2 EST1 + \beta_3 EST5 + \beta_4 MORF.f1 + \beta_5 MORF.f2)$$

$$j = 1, \dots, 4; i = 1, \dots, 473; k = 1, \dots, 15; k(i) = 1, \dots, 1813; t = 1, \dots, 691$$

El Cuadro 7 muestra los resultados del ajuste.

	coef	exp.coef.	se.coef.	robust.se	z	Pr...z..	lower..95	upper..95
MORF.f1	-0.398	0.672	0.135	0.103	-3.873	0.000	0.549	0.822
MORF.f2	-0.256	0.774	0.163	0.145	-1.769	0.077	0.582	1.028
ANIM1	0.658	1.930	0.176	0.196	3.357	0.001	1.315	2.834
tt(ANIM1)	-0.001	0.999	0.001	0.001	-2.412	0.016	0.997	1.000
EST11	-0.401	0.670	0.113	0.098	-4.077	0.000	0.552	0.812
EST51	-0.396	0.673	0.169	0.155	-2.550	0.011	0.496	0.912

Cuadro 8.8: Modelo de eventos múltiples. coef: betas estimados, exp(coef): hazard ratios, se(coef): error típico de betas estimados, robust se: error típico de beta con estimador sandwich, z: coef / robust se, Pr(z): p-valor, lower 95: extremo izquierdo de intervalo de confianza de 95 por ciento para hazard ratio, upper 95: extremo derecho de intervalo de confianza de 95 por ciento para hazard ratio.

En primer lugar se observa que el coeficiente $\beta_1 (g(t) = t)$ de ANIM fue significativo, confirmando la dependencia temporal del coeficiente de ANIM. Con respecto al riesgo de recurrencia de error (columna «exp.coef»), se observa que dicho riesgo *aumenta* para las TYPE de concordancia con controlador animado en un 93 %. Por otra parte el riesgo *disminuye* para las TYPE de concordancia: (i) con distancia media [similaridad media] entre las terminaciones de español e italiano (MORF.f1) en un 32,8 %; (ii) en donde se puede aplicar la estrategia EST1 en un 33 %; (iii) en donde es posible aplicar la estrategia EST5 en un 32,7 %. Si bien el riesgo de recurrencia bajó en un 22 % para TYPE de concordancia con distancia alta [similaridad baja] entre las terminaciones de español e italiano (MORF.f2), dicha baja no resultó significativa.

8.4. El modelo de riesgos competitivos.

En este modelo se considera que hay k eventos (tipos de errores) posibles que compiten entre sí, de los cuales solo uno es observado y los demás están censurados. En este esquema *cada* individuo / instancia (sin considerar su TYPE) se organizó como en el Cuadro 1 pero cuadruplicando cada fila como se muestra en el Cuadro 9:

ID	INSTANCIA	ESPAÑOL	DE	A	TRANS	TIEMPO	STATUS	... (otras variables)
1	mucho pensadores	muchos pensadores	1	2	1	3	0	...
1	mucho pensadores	muchos pensadores	1	3	2	3	0	...
1	mucho pensadores	muchos pensadores	1	4	3	3	1	...
1	mucho pensadores	muchos pensadores	1	5	4	3	0	...

Cuadro 8.9: Ejemplo de datos de sobrevivencia: modelo de riesgos competitivos.

El modelo de riesgos competitivos se entiende como si fuera una cadena de estados que comienza en el estado 1 y de allí se puede avanzar a cualquiera de los otros k estados, aquí los $k = 2, \dots, 5$ representan los estados de cada tipo de error. El cuadro ilustra la organización de los datos para la tercera instancia del cuadro 1. Se establece que se parte del estado 1 (columna «DE») y se va hacia cualquiera de los cuatro estados posibles (columna «A»). Cada posible transición se numera en la columna «TRANS». La columna «TIEMPO» establece el tiempo del evento o de la censura, repetido para cada tipo de transición. Luego, la columna «STATUS» tiene un «1» en la transición donde se produce el evento y «0» en las demás (si no hubiera evento, hay «0» en todas las filas). O sea que, si hay evento, las otras causas posibles del error

se hallan «censuradas», indicando que en ese instante de tiempo podrían ocurrir en lugar de la observada, aunque no se observen.

Para el caso de los riesgos competitivos, es preciso definir la función de *incidencia* acumulada para una causa específica como sigue (Putter et. al., 2007; Moore, 2016): $I_k(t) = P(T \leq t, C = k) = \int_0^t h_k(u) S(u) du$. Especifica la probabilidad acumulada para un individuo / concordancia de «morir» («sufrir error») por la causa k-ésima («tipo de error»). Es una función creciente pero, en el límite, alcanza la probabilidad de «muerte» por esa causa en particular; o sea que siempre está debajo de 1 ($I_k(\infty) = P(C = k)$). La función de riesgo instantánea (*hazard*) para una causa específica se define como la probabilidad de sobrevivir (no morir de la k-ésima causa) un intervalo Δ corto de tiempo adicional sabiendo que el individuo sobrevivió hasta el tiempo t .

$$h_k(t) = \lim_{\Delta \rightarrow 0} \frac{P(t < T < t + \Delta, C = k | T > t)}{\Delta}$$

Si se suman los *hazards* puntuales de todas las causas de «muerte» se obtiene la *hazard* total mencionada en 8.1.1: $h(t) = \sum_{k=1}^K h_k(t)$. Supóngase que hay D tiempos de falla ordenados en los que pueden ocurrir fallas de cualquier causa: $0 < t_1 < t_2 < \dots < t_D$. Se puede estimar $h_k(t)$ como $\hat{h}_k(t_i) = d_{ik}/n_i$, el número de eventos de la k-ésima causa en el i-ésimo tiempo dividido por el número de individuos en riesgo en ese tiempo i . La suma sobre todas las *hazards* específicas da la *hazard* total: $\hat{h}_k(t_i) = \left(\sum_{k=1}^K d_{ik} \right) / n_i$. Ahora bien, la función de incidencia acumulada estimada indica la probabilidad (acumulada) de «morir» de la k-ésima causa en el tiempo j-ésimo:

$$\hat{I}_k(t) = P(T \leq t, C = k) = \sum_{t_i \leq t} \hat{p}_k(t_i) = \sum_{t_i \leq t} \hat{S}(t_{i-1}) \hat{h}_k(t_i)$$

Es decir, simplemente se suman todas las probabilidades de «morir» hasta el tiempo t_i , donde cada probabilidad es el producto de la función de sobrevivencia puntual para el tiempo anterior t_{i-1} y la *hazard* puntual específica para la causa k-ésima en el tiempo t_i ; que resulta también igual a uno menos la función de sobrevivencia total de «morir» de cualquier causa, para cada punto t_i .

$$\hat{S}(t) = \prod_{t_i < t} \left(1 - \frac{d_i}{n_i} \right) = \prod_{t_i < t} \left(1 - \sum_{k=1}^K \hat{h}_k(t_i) \right)$$

El modelo proporcional de Cox para riesgos competitivos se escribe como:

$$h_i(t | \mathbf{x}, C = k) = h_{0k}(t) \exp(\mathbf{x}_k^T \boldsymbol{\beta}_k) = h_{0k} \exp(\beta_{1k} x_{1k} + \beta_{2k} x_{2k} + \dots + \beta_{pk} x_{pk})$$

Donde: h_{0k} es la *hazard* basal para la causa k-ésima, \mathbf{x}_k es el vector de covariables para la k-ésima causa y $\boldsymbol{\beta}_k$, su respectivo vector de coeficientes. O sea que el modelo admite que *HR* de las covariables pueda ser específico para cada causa de «muerte» (error).

8.4.1. Selección de modelos.

En primera instancia se muestran, en la Figura 4, las curvas de incidencia acumulada para los diferentes tipos de error. La probabilidad, para cada punto t_i , de que una concordancia sufra un error del tipo $k = 1, \dots, 4$ está representada como la distancia entre las curvas. Por supuesto, cuanto más tiempo pasa, más probabilidad hay de sufrir el evento «error». Se observa que la probabilidad acumulada de plural se incrementa sostenidamente hasta acumular aproximadamente 20% de probabilidad de error. Los otros tipos de error acumulan como máximo aproximadamente: 6,6% (error mixto), 3,5% (error de género), 1,7% (error de -e- epentética). Por otra parte, la Figura 5 ilustra las incidencias acumuladas de los tipos de error por alumno (1 = SONIA, 2 = NATI, 3 = JAKO, 4 = MIRKA). Las probabilidades acumuladas del error de género y de -e- epentética son mayores en SONIA y NATI que en JAKO y MIRKA; las de los errores

mixtos son mayores en MIRKA y NATI que en SONIA y JAKO. Por otro lado, la probabilidad acumulada de error de plural sube más pronunciadamente en MIRKA y NATI.

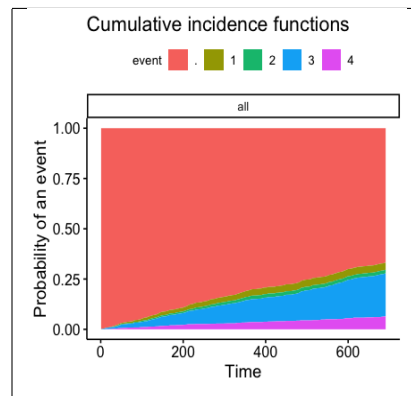


Figura 8.4: Incidencia acumulada de tipos de error: 1 = error de género; 2 = error de -e- epentética; 3 = error de plural; 4 = error mixto; EVENT = sin error (censurados).

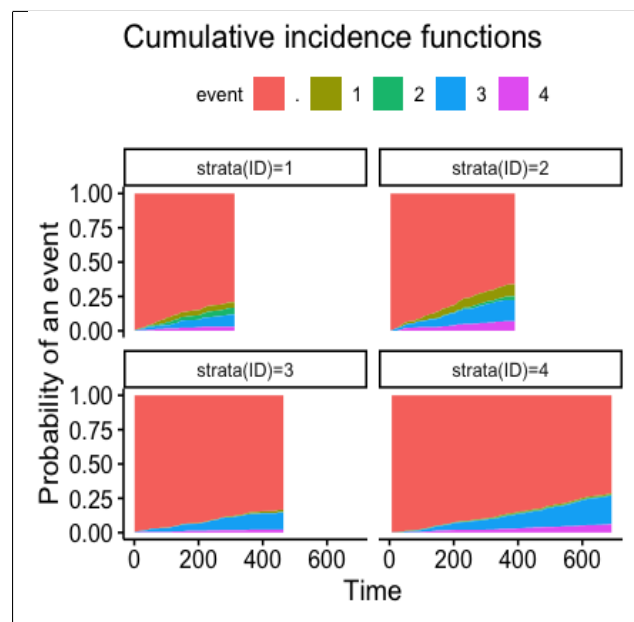


Figura 8.5: Incidencia acumulada de tipos de error por alumno (ID): 1 = SONIA; 2 = NATI; 3 = JAKO; 4 = MIRKA.

Ya que las concordancias están anidadas en sesiones (SESIÓN) y estas en los alumnos (ID), se consideraron datos agrupados en 52 *clusters* dados por la cruce entre ID:SESIÓN. Se presume que las concordancias dentro de cada uno de dichos grupos se hallan más correlacionadas entre sí que con aquellas de otros grupos. Para dar cuenta de esta correlación se ajustó un modelo con varianza «sándwich» agrupando por ID:SESIÓN y otro de «fragilidad compartida» con efectos aleatorios de ID:SESIÓN. La medida de información de Akaike arrojó $AIC = 6628,656$ para el segundo modelo y $AIC = 6703,988$ para el primero; con lo cual se eligió el modelo de «fragilidad compartida». Además, para el modelo elegido la varianza (de la distribución *Gamma*) de los efectos aleatorios fue $\theta = 0,29$ y la correlación intra grupo $IC = \theta / (2 + \theta) = 0,29 / (2 + 0,29) \approx 0,12$.

Se ajustó el modelo elegido con todas las predictoras. Fueron $2^{17} = 131072$ modelos, jerarquizados mediante la medida de información AIC (como: $\frac{n}{p} = \frac{7428}{17} \approx 436 > 40$, no se usó la versión AICc corregida por tamaño muestral). Luego se examinó la frecuencia de las predictoras en el conjunto completo de modelos, que da un panorama de la incerteza por la selección. A continuación se redujo la cantidad de modelos al subconjunto «de confianza» con la regla $\frac{W(i)}{W(1)} > \frac{1}{8}$. Sobre dicho subconjunto se llevó a cabo un promedio de

coeficientes con la varianza calculada con «full average». El Cuadro 10 muestra que las variables con porcentaje de elección arriba del 80 % son: ANIM, EST1, EST5, MORF.f, Fabs.SC.f, FAM.LEX.f, MOD, ES. A excepción de ES, la demás predictoras resultan significativas en los coeficientes promediados (Cuadro 11).

	names	x
1	frailty(ID.SESION)	1.00
2	strata(trans)	1.00
3	MOD	1.00
4	EST5	1.00
5	MORF.f	0.99
6	FAM.LEX.f	0.99
7	EST1	0.97
8	Fabs.SC.f	0.93
9	ANIM	0.93
10	ES	0.83
11	EST4	0.70
12	STEM.f	0.56
13	EST3	0.52
14	EST7	0.44
15	EST6	0.43
16	IMA.CONC.f	0.36
17	EST2	0.36
18	GRAMS	0.35
19	LDA	0.28

Cuadro 8.10: Importancia Relativa de las predictoras.

8.4.2. Modelo final: chequeo de supuestos y resultados.

Se compararon mediante *AIC* modelos de fragilidad compartida con y sin coeficientes específicos para cada causa de error; resultando en: $AIC(con) = 6595,025$; $AIC(sin) = 6626,958$. Conforme a que el *AIC* del modelo con coeficientes específicos fue menor, el modelo final incluyó los coeficientes específicos que resultaron significativos para cada tipo de error (datos completos en el Apéndice). El modelo de «fragilidad compartida» se escribió como:

$$\begin{aligned}
 h_{ik}(t | \mathbf{x}, u_i, C = j) = & h_{0j}(t) \exp(\beta_1 Fabs.SC.f(1),1 + \beta_2 Fabs.SC.f(1),2 + \beta_3 Fabs.SC.f(1),3 + \\
 & \beta_4 MORF.f(1),2 + \beta_5 MORF.f(1),3 + \beta_6 MORF.f(2),3 + \beta_7 MOD(2),3 + \beta_8 MOD(2),4 + \beta_9 MOD(3),3 + \\
 & \beta_{10} MOD(3),4 + \beta_{11} ANIM(1),2 + \beta_{12} ANIM(1),4 + \beta_{13} FAM.LEX.f(1),2 + \beta_{14} FAM.LEX.f(1),3 + \\
 & \beta_{15} FAM.LEX.f(1),4 + \beta_{16} EST1(1),1 + \beta_{17} EST1(1),4 + \beta_{18} EST5(1),4 + u_i) \\
 & j = 1, \dots, 4; i = 1, \dots, 52; k = 1, \dots, 7428; t = 1, \dots, 691; u_i \sim Gamma(\theta)
 \end{aligned}$$

donde $h_{ik}(t | \mathbf{x}, u_i, C = j)$ representa la función de riesgo de error de la k -ésima concordancia ($k = 1857 \times 4 = 7428$) dentro del i -ésimo grupo estratificado según el j -ésimo tipo de error ($j = 1, \dots, 4$) en el tiempo t ($t = 1, \dots, 691$); $h_{0j}(t)$ es el *hazard* basal específico para cada tipo de error ($j = 1, \dots, 4$); y u_i , el i -ésimo factor aleatorio del grupo ID:SESIÓN ($i = 1, \dots, 52$). Los efectos fijos específicos para una determinada causa de error se detallan a continuación:

	Estimate	Std. Error	z value	Pr(> z)
ANIM1	0.352	0.117	3.009	0.003
ES1	0.316	0.262	1.207	0.227
ES2	-0.694	0.571	1.216	0.224
EST11	-0.433	0.127	3.408	0.001
EST41	-0.534	0.369	1.446	0.148
EST51	-0.786	0.200	3.921	0.000
FAM.LEX.f1	-0.347	0.100	3.466	0.001
Fabs.SC.f1	-0.332	0.129	2.568	0.010
MOD1	-0.419	0.429	0.977	0.329
MOD2	0.518	0.126	4.128	0.000
MOD3	0.370	0.143	2.589	0.010
MORF.f1	-0.719	0.183	3.931	0.000
MORF.f2	-0.667	0.293	2.275	0.023
STEM.f1	0.108	0.129	0.837	0.402
EST31	0.312	0.475	0.655	0.512
EST71	0.118	0.214	0.550	0.582
EST61	0.071	0.159	0.445	0.656
GRAMS1	-0.027	0.076	0.351	0.726
IMA.CONC.f1	-0.022	0.069	0.312	0.755
EST21	0.017	0.074	0.230	0.818
LDA1	0.000	0.084	0.006	0.995

Cuadro 8.11: Promedio de los coeficientes con FULL AVERAGE.

Coefficiente	Descripción	Tipo de Error	Coefficiente	Descripción	Tipo de Error
β_1	Efecto de Fabs.SC.f, nivel 1 (ref=0)	Género	β_{10}	Efecto de MOD, nivel 3 (ref=0)	Mixto
β_2	Efecto de Fabs.SC.f, nivel 1 (ref=0)	-e- epentética	β_{11}	Efecto de ANIM, nivel 1 (ref=0)	-e- epentética
β_3	Efecto de Fabs.SC.f, nivel 1 (ref=0)	Plural	β_{12}	Efecto de ANIM, nivel 1 (ref=0)	Mixto
β_4	Efecto de MORF.f, nivel 1	-e- epentética	β_{13}	Efecto de FAM.LEX.f, nivel 1 (ref=0)	-e- epentética
β_5	Efecto de MORF.f, nivel 1 (ref=0)	Plural	β_{14}	Efecto de FAM.LEX.f, nivel 1 (ref=0)	Plural
β_6	Efecto de MORF.f, nivel 2 (ref=0)	Plural	β_{15}	Efecto de FAM.LEX.f, nivel 1 (ref=0)	Mixto
β_7	Efecto de MOD, nivel 2 (ref=0)	Plural	β_{16}	Efecto de EST1, nivel 1 (ref=0)	Género
β_8	Efecto de MOD, nivel 2 (ref=0)	Mixto	β_{17}	Efecto de EST1, nivel 1 (ref=0)	Mixto
β_9	Efecto de MOD, nivel 3 (ref=0)	Plural	β_{17}	Efecto de EST5, nivel 1 (ref=0)	Mixto

El Cuadro 12 muestra los p-valores del test de $\theta = 0$ para $\beta(t) = \beta + \theta g(t)$, donde $g(t)$ es una función del tiempo definida según: (i) $g(t) = rank(t)$; (ii) $g(t) = \hat{S}_{km}(t)$; (iii) $g(t) = t$; (iv) $g(t) = \log(t)$. Se observa que Fabs.SC.f(1).3, FAM.LEX.f(1).3 y EST1(1).1 resultan significativas. Por lo tanto, no se cumple el supuesto de *hazards* proporcionales para dichas variables.

Para acomodar esto, se ajustó el modelo estratificando para estas variables según $V = \{1 : t \leq 300; 2 : t > 300\}$. Es decir que para Fabs.SC.f(1).3, FAM.LEX.f(1).3 y EST1(1).1 hubo un coeficiente específico para cada intervalo de datos. El modelo final fue el siguiente:

$$\begin{aligned}
h_{ik}(t | \mathbf{x}, u_i, C = j, V = w) = & h_{0j}(t) \exp(\beta_1 Fabs.SC.f(1),1 + \beta_2 Fabs.SC.f(1),2 + \beta_{3,w=1} Fabs.SC.f(1),3 + \\
& \beta_{4,w=2} Fabs.SC.f(1),3 + \beta_5 MORF.f(1),2 + \beta_6 MORF.f(1),3 + \beta_7 MORF.f(2),3 + \\
& \beta_8 MOD(2),3 + \beta_9 MOD(2),4 + \beta_{10} MOD(3),3 + \beta_{11} MOD(3),4 + \\
& \beta_{12} ANIM(1),2 + \beta_{13} ANIM(1),4 + \beta_{14} FAM.LEX.f(1),2 + \beta_{15,w=1} FAM.LEX.f(1),3 + \\
& \beta_{16,w=2} FAM.LEX.f(1),3 + \beta_{17} FAM.LEX.f(1),4 + \beta_{18,w=1} EST1(1),1 + \beta_{19,w=2} EST1(1),1 + \\
& \beta_{20} EST1(1),4 + \beta_{21} EST5(1),4 + u_i) \\
& j = 1, \dots, 4; i = 1, \dots, 52; k = 1, \dots, 13648; t = 1, \dots, 691; w = 1, 2; u_i \sim Gamma(\theta)
\end{aligned}$$

Ahora, según el test de $\theta = 0$ para $\beta(t) = \beta + \theta g(t)$, solamente FAM.LEX.f1.3 en el grupo $V = 1$ resultó

	df	rank.chi2	rank.p	km.chi2	km.p	t.chi2	t.p	log.t.chi2	log.t.p
Fabs.SC.f1.1	1.0000	0.1424	0.7059	0.1719	0.6785	0.2095	0.6471	0.5122	0.4742
Fabs.SC.f1.2	1.0000	0.0874	0.7675	0.0487	0.8254	0.0590	0.8081	0.5978	0.4394
Fabs.SC.f1.3	1.0000	3.4814	0.0621	4.7143	0.0299	4.9946	0.0254	1.8575	0.1729
MORF.f1.2	1.0000	0.1431	0.7052	0.0605	0.8057	0.0558	0.8132	0.7727	0.3794
MORF.f1.3	1.0000	0.0212	0.8842	0.2167	0.6416	0.2568	0.6124	0.0219	0.8823
MORF.f2.3	1.0000	0.2631	0.6080	0.3644	0.5461	0.3385	0.5607	0.1632	0.6862
MOD2.3	1.0000	0.9175	0.3381	0.9863	0.3206	1.2790	0.2581	0.7758	0.3784
MOD2.4	1.0000	0.1414	0.7069	0.0188	0.8910	0.0436	0.8345	0.4572	0.4989
MOD3.3	1.0000	1.2978	0.2546	1.0419	0.3074	0.9624	0.3266	0.9200	0.3375
MOD3.4	1.0000	1.1326	0.2872	0.8331	0.3614	0.8269	0.3632	2.0319	0.1540
ANIM1.2	1.0000	0.0004	0.9840	0.0000	0.9960	0.0002	0.9888	0.4288	0.5126
ANIM1.4	1.0000	1.0977	0.2948	1.4705	0.2253	1.3641	0.2428	1.0714	0.3006
FAM.LEX.f1.2	1.0000	0.1446	0.7037	0.0515	0.8204	0.0587	0.8086	1.2557	0.2625
FAM.LEX.f1.3	1.0000	6.5752	0.0103	4.8466	0.0277	4.2661	0.0389	7.1557	0.0075
FAM.LEX.f1.4	1.0000	0.1442	0.7041	0.0051	0.9432	0.0000	0.9976	0.8679	0.3515
EST11.1	1.0000	5.4959	0.0191	4.2131	0.0401	4.0634	0.0438	4.4186	0.0355
EST11.4	1.0000	0.0042	0.9483	0.1776	0.6734	0.1269	0.7217	0.2670	0.6054
EST51.4	1.0000	0.6369	0.4248	0.5995	0.4388	0.5602	0.4542	0.0697	0.7917
GLOBAL	18.0000	21.5100	0.2545	19.5436	0.3591	19.3426	0.3710	22.9307	0.1933

Cuadro 8.12: Test para $\theta = 0$, según diferentes funciones del tiempo: estadístico de chi cuadrado y p valor.

significativo ($p = 0,025$) en $g(t) = \log(t)$. La situación mejoró mucho, teniendo en cuenta que en el Cuadro 12 el coeficiente era significativo en todas las funciones del tiempo y con $p = 0,0075$ específicamente para la logarítmica. La Figura 6 muestra los residuos «dfbetas» escalados. Se observa que ningún gráfico supera dos desvíos en valor absoluto. Por lo tanto, se concluye que no hay observaciones influyentes. Por otra parte, en la Figura 7 se muestran los residuos de devianza. El 3,7% supera en valor absoluto los dos desvíos (513 observaciones de 13648). Está dentro del cinco por ciento esperado de valores atípicos.

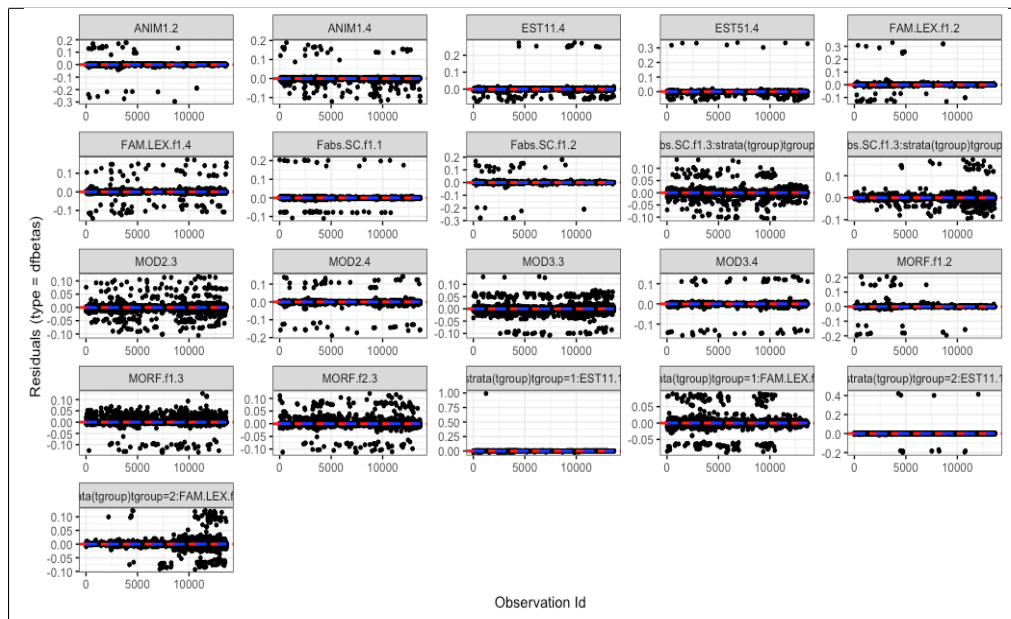


Figura 8.6: Residuos DFBETA escalados, para cada coeficiente.

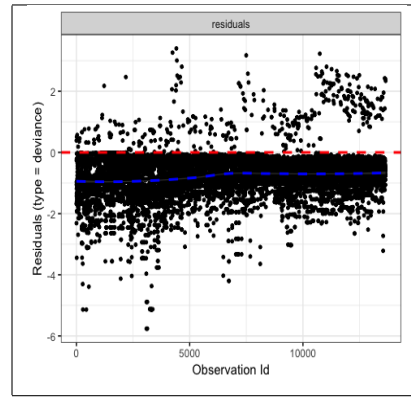


Figura 8.7: Residuos de devianza.

La Figura 8 muestra un *forest plot* con los *hazard ratios* ($HR = \exp(\beta)$) para cada coeficiente. El Cuadro 13 expresa los HR, sus intervalos de confianza de nivel de 95 %, el tamaño del efecto ($|1 - HR| \times 100$) y su dirección («UP»: factor de riesgo, «DOWN»: factor de protección). Se observa que no resultan significativos (el «uno» está contenido en el intervalo de confianza): Fabs.SC.f1.2, ANIM1.4, Fabs.SC.f1.3:group=1, FAM.LEX.f1.3:group=1, EST11.1:group=2. El riesgo de error sube para MOD(3) [error mixto] y ANIM(1) [error de -e- epentética] (pero obsérvese el largo intervalo de confianza, que hace que la estimación sea incierta), que son factores de riesgo. Para el error de género las concordancias de TYPE más frecuente [Fabs.SC.f(1)] y aquellas donde se aplica la estrategia 1 [EST1(1), $t < 300$] bajan el riesgo de error. Respecto del error de -e- epentética, el riesgo baja las concordancias con terminaciones de similaridad media con el italiano [MORF.f(1)], aquellas con controlador familiar y/o frecuente [FAM.LEX.f(1)] y lo sube el controlador animado [ANIM(1)]. En lo que atañe al error de plural, bajan el riesgo de error las concordancias con terminaciones de similaridad media [MORF.f(1)] y baja [MORF.f(2)] con el italiano, los determinantes [MOD(2)], los adjetivos [MOD(3)], las concordancias de TYPE más frecuente en el periodo $t > 300$ [Fabs.SC.f(1):group=2], y los controladores familiares y/o frecuentes en igual periodo [FAM.LEX.f(1):group=2]. En lo referente al error mixto, sube el riesgo en presencia de determinantes [MOD(2)] y adjetivos [MOD(3)], controladores familiares y/o frecuentes [FAM.LEX.f(1)], y los contextos donde es aplicable primera [EST1(1)] y la quinta [EST5(1)] estrategia.

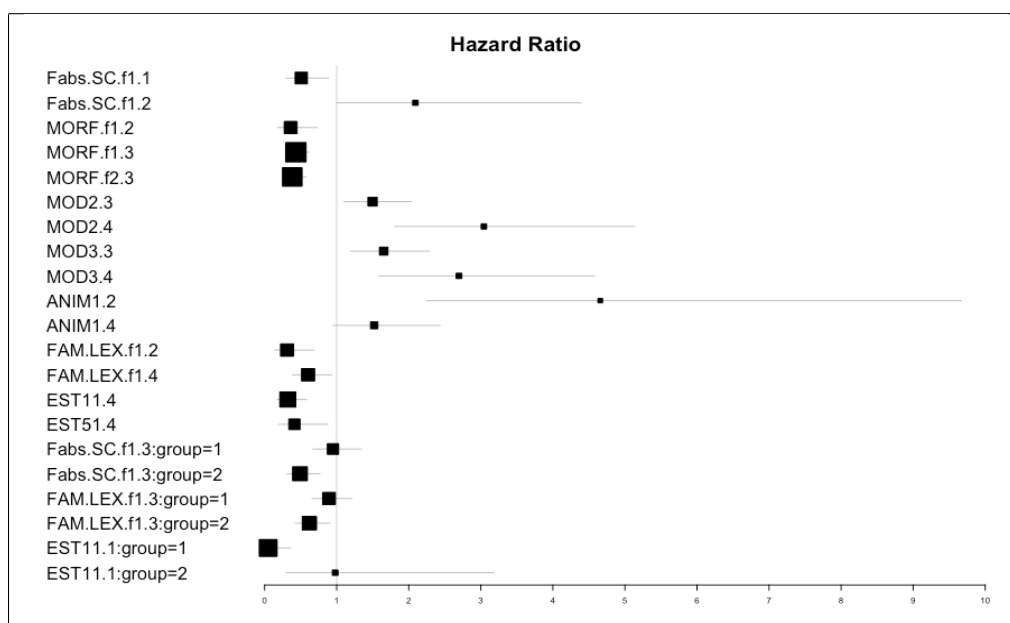


Figura 8.8: Forest Plot: HR para el modelo final.

	exp.coef.	lower..95	upper..95	size	dir
Fabs.SC.f1.1	0.508	0.293	0.882	49.181	DOWN
Fabs.SC.f1.2	2.093	0.998	4.389	109.256	UP
MORF.f1.2	0.361	0.178	0.730	63.911	DOWN
MORF.f1.3	0.433	0.310	0.606	56.682	DOWN
MORF.f2.3	0.384	0.258	0.572	61.624	DOWN
MOD2.3	1.498	1.104	2.032	49.773	DOWN
MOD2.4	3.043	1.804	5.134	204.283	UP
MOD3.3	1.652	1.193	2.287	65.179	DOWN
MOD3.4	2.697	1.588	4.580	169.674	UP
ANIM1.2	4.657	2.244	9.665	365.698	UP
ANIM1.4	1.520	0.948	2.436	52.008	DOWN
FAM.LEX.f1.2	0.311	0.142	0.682	68.859	DOWN
FAM.LEX.f1.4	0.603	0.391	0.931	39.704	DOWN
EST11.4	0.321	0.176	0.584	67.919	DOWN
EST51.4	0.413	0.196	0.870	58.692	DOWN
Fabs.SC.f1.3:group=1	0.949	0.671	1.341	5.121	DOWN
Fabs.SC.f1.3:group=2	0.491	0.314	0.767	50.909	DOWN
FAM.LEX.f1.3:group=1	0.893	0.658	1.213	10.680	DOWN
FAM.LEX.f1.3:group=2	0.621	0.427	0.904	37.867	DOWN
EST11.1:group=1	0.049	0.007	0.358	95.069	DOWN
EST11.1:group=2	0.979	0.301	3.187	2.116	DOWN

Cuadro 8.13: Modelo de riesgos competitivos. exp(coef): hazard ratios, size: tamaño del efecto en porcentaje, dir: dirección del efecto, lower 95: extremo izquierdo de intervalo de confianza de 95 por ciento para hazard ratio, upper 95: extremo derecho de intervalo de confianza de 95 por ciento para hazard ratio.

8.5. Conclusiones.

El Cuadro 14 compara los resultados de ambos modelos con las expectativas del análisis. Las casillas vacías indican efectos no significativos. Se encontró efecto en la dirección inversa a la esperada en MORF.f (riesgo bajó en lugar de subir). El riesgo también bajo en lugar de subir con el error de plural en determinantes y adjetivos pero coincidió con lo esperado en los errores mixtos.

Predictor	Efecto	dirección / tamaño esperado	Eventos Múltiples	Riesgos Competitivos (tipo de error)
MOD	MOD(1) - MOD(0)	↑		
	MOD(2) - MOD(0)	↑		↓ (plural), ↑ (mixto)
	MOD(3) - MOD(0)	↑		↓ (plural), ↑ (mixto)
Fabs.SC.f	Fabs.SC.f(1) - Fabs.SC.f(0)	↓		↓ (género, plural)
MORF.f	MORF.f(1) - MORF.f(0)	↑	↓	↓ (-e- epentética, plural)
	MORF.f(2) - MORF.f(0)	↑↑		↓↓ (plural)
FAM.LEX.f	FAM.LEX.f(1) - FAM.LEX.f(0)	↓		↓ (-e- epentética, mixto, plural: $t > 300$)
EST1	EST1(1) - EST1(0)	↓	↓	↓ (género: $t < 300$, mixto)
EST2	EST2(1) - EST2(0)	↓		
EST3	EST3(1) - EST3(0)	↑		
EST4	EST4(1) - EST4(0)	↓		
EST5	EST5(1) - EST5(0)	↓	↓	↓(mixto)
EST6	EST6(1) - EST6(0)	↓		
EST7	EST7(1) - EST7(0)	↑		
ANIM	ANIM(1) - ANIM(0)	↑	↑	↑ (plural)
ES	ES(1) - ES(0)	↑		
	ES(2) - ES(0)	↑↑		
LDA	LDA(1)-LDA(0)	↑		
GRAMS	GRAMS(1)-GRAMS(0)	↑		
STEM.f	STEM.f(1)-STEM.f(0)	↑		
IMA.CONC.f	IMA.CONC.f(1) - IMA.CONC.f(0)	↓		

Cuadro 8.14: Expectativas del análisis. Leyenda: (↑) = riesgo aumenta; (↓) = riesgo disminuye; (↑↑/↓↓) = mayor tamaño de efecto.

8.6. Apéndice de capítulo.

	coef	se.coef.	Chisq	DF	p	exp.coef.	lower..95	upper..95
Fabs.SC.f1.1	-0.972	0.303	10.248	1.000	0.001	0.378	0.209	0.686
Fabs.SC.f1.2	1.056	0.524	4.066	1.000	0.044	2.876	1.030	8.031
Fabs.SC.f1.3	-0.315	0.152	4.325	1.000	0.038	0.729	0.542	0.982
Fabs.SC.f1.4	-0.258	0.269	0.924	1.000	0.336	0.772	0.456	1.308
MORF.f1.1	0.278	0.480	0.335	1.000	0.563	1.320	0.515	3.384
MORF.f1.2	-1.379	0.612	5.075	1.000	0.024	0.252	0.076	0.836
MORF.f1.3	-0.837	0.173	23.510	1.000	0.000	0.433	0.309	0.608
MORF.f1.4	-0.440	0.355	1.529	1.000	0.216	0.644	0.321	1.293
MORF.f2.1	0.245	0.566	0.187	1.000	0.665	1.278	0.421	3.873
MORF.f2.2	-0.502	0.634	0.626	1.000	0.429	0.605	0.175	2.098
MORF.f2.3	-0.907	0.218	17.247	1.000	0.000	0.404	0.263	0.619
MORF.f2.4	0.370	0.389	0.903	1.000	0.342	1.447	0.675	3.100
MOD1.1	-0.505	1.036	0.237	1.000	0.626	0.604	0.079	4.601
MOD1.3	-0.322	0.523	0.380	1.000	0.538	0.725	0.260	2.018
MOD1.4	-0.294	1.039	0.080	1.000	0.777	0.745	0.097	5.709
MOD2.1	0.551	0.294	3.514	1.000	0.061	1.734	0.975	3.085
MOD2.2	0.373	0.507	0.540	1.000	0.463	1.451	0.537	3.921
MOD2.3	0.409	0.160	6.568	1.000	0.010	1.505	1.101	2.058
MOD2.4	1.062	0.281	14.304	1.000	0.000	2.892	1.668	5.013
MOD3.1	-0.691	0.404	2.921	1.000	0.087	0.501	0.227	1.107
MOD3.2	0.481	0.647	0.553	1.000	0.457	1.618	0.455	5.751
MOD3.3	0.487	0.171	8.108	1.000	0.004	1.627	1.164	2.275
MOD3.4	0.730	0.324	5.064	1.000	0.024	2.076	1.099	3.921
ANIM1.1	0.470	0.277	2.889	1.000	0.089	1.600	0.931	2.753
ANIM1.2	1.654	0.392	17.844	1.000	0.000	5.229	2.427	11.266
ANIM1.3	0.089	0.140	0.403	1.000	0.526	1.093	0.830	1.439
ANIM1.4	0.491	0.242	4.100	1.000	0.043	1.633	1.016	2.626
FAM.LEX.f1.1	-0.110	0.263	0.175	1.000	0.676	0.896	0.535	1.501
FAM.LEX.f1.2	-1.295	0.425	9.278	1.000	0.002	0.274	0.119	0.630
FAM.LEX.f1.3	-0.271	0.125	4.705	1.000	0.030	0.762	0.597	0.974
FAM.LEX.f1.4	-0.510	0.222	5.274	1.000	0.022	0.601	0.389	0.928
EST11.1	-1.721	0.474	13.174	1.000	0.000	0.179	0.071	0.453
EST11.2	-0.118	0.454	0.068	1.000	0.795	0.888	0.365	2.165
EST11.3	-0.097	0.138	0.489	1.000	0.484	0.908	0.692	1.191
EST11.4	-0.996	0.310	10.329	1.000	0.001	0.369	0.201	0.678
EST51.1	-0.792	0.507	2.438	1.000	0.118	0.453	0.168	1.224
EST51.3	-0.155	0.217	0.516	1.000	0.473	0.856	0.560	1.309
EST51.4	-1.388	0.409	11.495	1.000	0.001	0.250	0.112	0.557

Cuadro 8.15: Modelo de riesgos competitivos. coef: betas estimados, exp(coef): hazard ratios, se(coef): error típico de betas estimados, robust se: error típico de beta con estimador sandwich, z: coef / robust se, Pr(z): p-valor, lower 95: extremo izquierdo de intervalo de confianza de 95 por ciento para hazard ratio, upper 95: extremo derecho de intervalo de confianza de 95 por ciento para hazard ratio.

Capítulo 9

Simulación con sistemas dinámicos.

9.1. Introducción.

En este capítulo se llevan a cabo tres simulaciones basadas en modelos de sistemas dinámicos, utilizando los resultados de los capítulos 7 y 8. La primera se inspira en un modelo de *Lokta-Volterra* para competición entre dos especies, en el marco de la ecología de poblaciones. La segunda, en la dinámica de replicador en un juego de coordinación en el marco de la teoría de los juegos evolutiva. La tercera simulación usa una aproximación de ascenso por gradiente, en el marco de la *gradient symbolic computation*. Se comparan resultados de patrones agregados de proporciones de error / acierto globales y por sesiones, para cada alumno, entre la simulación y los datos del corpus.

Se considerarán las siguientes variables, que resultaron en una *disminución* de la chance de error para el modelo multinomial y los basados en tiempo hasta un evento. A saber:

Tipo de error	Variables
Género	MOD = "3", Fabs.SC.f = "1", FAM.LEX.f = "1", EST1 = "1"
-e- epentética	MORF.f = "1", Fabs.SC.f = "1", FAM.LEX.f = "1", EST5 = "1", EST2 = "1"
Plural	MOD = "2/3", ES = "2", MORF.f = "1/2", Fabs.SC.f = "1", FAM.LEX.f = "1"
Mixto	Fabs.SC.f = "1", FAM.LEX.f = "1", EST1 = "1", EST5 = "1"

Cuadro 9.1: Variables que disminuyen la chance de error, según resultados de los capítulos 7 y 8.

También se tuvo en cuenta, para cada instancia, la proporción de errores cometidos hasta la instancia anterior, contando desde el inicio de cada sesión. Por ejemplo en SONIA, la primera sesión consta de las siguientes instancias: ["muchos personas" (error), "muchos pueblos" (correcto), "mucho pensadores" (error), "otros autores" (correcto)] y tiene las siguientes proporciones de errores anteriores a una instancia en cuestión: [0; 1; 0,5; 0,66]. Es decir que en "otros autores" hubo dos errores sobre tres instancias producidas antes, contando desde el inicio de la sesión.

Las simulaciones que siguen tienen siempre dos atractores *estables*: "correcto" y "error". Para cada instancia, se intenta recrear un sendero de "activación" en el "paisaje de eventos mentales" hasta que el sistema se acerque a uno de los atractores. Las variables especificadas en cada tipo de error se toman como un "contexto" que sesga el flujo de activación hacia uno de los atractores. Para lograr esto, se plantean diferentes mecanismos de similitud entre las variables de la instancia que se procesa y cada uno de los contextos que disminuyen el error. O sea, la idea fue combinar la lejanía a los contextos que favorecen el buen desempeño junto con los errores cometidos anteriormente, para sesgar el flujo de activación hacia uno de los dos atractores.

A continuación se detallan los modelos y los resultados obtenidos.

9.2. Modelo competitivo de Lokta-Volterra.

Los modelos de competición de *Lokta-Volterra* intentan captar la interacción entre (dos) especies. Dichas especies se interfieren entre sí, disminuyendo (las unas a las otras) las tasas de crecimiento de abundancias poblacionales (Kot, 2001; Murray, 2002). Se trata de una extensión a varias especies del modelo logístico de una especie. Sean entonces dos especies, cada una con abundancias N_i , $i = 1, 2$. En ausencia de la otra, la abundancia de cada especie sigue un crecimiento logístico cuya tasa decrece linealmente con el tamaño poblacional. Cada especie tiene su propia tasa constante de crecimiento r_i y su propia capacidad de carga ambiental K_i (“carring capacity”, la máxima capacidad de “carga” del ambiente, respecto de cada especie). El modelo logístico reza:

$$\begin{aligned}\frac{dN_1}{dt} &= r_1 N_1 \left(1 - \frac{N_1}{K_1}\right) \\ \frac{dN_2}{dt} &= r_2 N_2 \left(1 - \frac{N_2}{K_2}\right)\end{aligned}$$

El término cuadrático representa la competencia intra-específica, o sea, entre miembros de la misma especie. A medida que $t \rightarrow \infty$, $N_i \rightarrow K_i$, para toda abundancia inicial $N_{0i} > 0$. Ahora introdúzcase un efecto de competición inter-específica, en el que cada especie causa un *decrecimiento* en la abundancia de la otra especie. Ya que las especies son diferentes, el efecto de disminución puede ser más fuerte o más débil. Dicho efecto se representa en dos parámetros de competencia: a_{ij} : la “fuerza” del efecto de competencia de la especie j respecto de la especie i ; a_{12}, a_{21} . El modelo queda:

$$\begin{aligned}\frac{dN_1}{dt} &= r_1 N_1 \left(1 - \frac{N_1}{K_1} - \alpha_{12} \frac{N_2}{K_1}\right) = \frac{r_1}{K_1} N_1 (K_1 - N_1 - \alpha_{12} N_2) \\ \frac{dN_2}{dt} &= r_2 N_2 \left(1 - \frac{N_2}{K_2} - \alpha_{21} \frac{N_1}{K_2}\right) = \frac{r_2}{K_2} N_2 (K_2 - N_2 - \alpha_{21} N_1)\end{aligned}$$

Aquí las abundancias de las dos especies serán: (i) N_1 : la abundancia de especie “correcto”; (ii) N_2 : la abundancia de la especie “error”. Las nulclinas del sistema son los puntos (N_1, N_2) donde $\dot{N}_i = \frac{dN_i}{dt} = 0$. Estas son:

$$\begin{aligned}N_1 : N_2 &= 0; N_2 = K_2 - \alpha_{21} N_1 \\ N_2 : N_1 &= 0; N_1 = K_1 - \alpha_{12} N_2\end{aligned}$$

Los puntos fijos del sistema son los puntos en donde las nulclinas se intersectan entre sí o intersectan los ejes del plano:

$$pf = \left\{ P_1 = (0, 0), P_2 = (K_1, 0), P_3 = (0, K_2), P_4 = \left(\frac{K_1 - \alpha_{12} K_2}{1 - \alpha_{12} \alpha_{21}}, \frac{K_2 - \alpha_{21} K_1}{1 - \alpha_{12} \alpha_{21}} \right) \right\}$$

El sistema exhibe cuatro dinámicas posibles. A saber:

- **Caso I** [$a_{12} < \frac{K_1}{K_2}; a_{21} < \frac{K_2}{K_1}$]: el punto fijo interior P_4 es un nodo estable que atrae todas las órbitas del primer cuadrante ($\mathbb{R}_{>0}^2$). El resto de los puntos fijos son inestables; los puntos P_2 y P_3 son puntos sillas (*saddles*) y P_1 es un nodo repulsor. La fuerza de la competencia entre las especies no es muy fuerte. Por lo tanto, hay coexistencia pero con abundancias menores a la capacidad de carga del ambiente de cada especie.
- **Caso II** [$a_{12} > \frac{K_1}{K_2}; a_{21} > \frac{K_2}{K_1}$]: el punto fijo interior P_4 es un nodo *saddle* (inestable). P_1 es un nodo repulsor (inestable) y los puntos P_2 y P_3 son nodos atractores (estables). Hay una *separatrix* que delimita el comportamiento de la dinámica: sobre ésta las órbitas son atraídas hacia P_3 y debajo de ésta son atraídas hacia P_2 . Las órbitas que comienzan en la *separatrix* (excepto el origen) convergen hacia P_4 . La competencia inter-específica es fuerte y eventualmente una de las especies gana y la otra se extingue, dependiendo de las condiciones iniciales.
- **Caso III** [$a_{12} < \frac{K_1}{K_2}; a_{21} > \frac{K_2}{K_1}$]: No hay punto fijo interior P_4 . los puntos P_1 y P_3 son inestables (repulsor y *saddle*, respectivamente) y P_2 es un nodo estable que atrae todas las órbitas. La primera especie hace una competencia más fuerte que la segunda y lleva a ésta última a la extinción.

- **Caso IV** [$a_{12} > \frac{K_1}{K_2}; a_{21} < \frac{K_2}{K_1}$]: No hay punto fijo interior P_4 . los puntos P_1 y P_2 son inestables (repulsor y *saddle*, respectivamente) y P_3 es un nodo estable que atrae todas las órbitas. La segunda especie hace una competencia más fuerte que la primera y lleva a ésta última a la extinción.

No son posibles ciclos. Obsérvese que el parámetro r_i no influye. El siguiente Cuadro resume las dinámicas, suponiendo, como se hará aquí, que $K_i = 1$. El análisis de estabilidad detallado se muestra en el apéndice del capítulo.

Clase	Condiciones	Puntos Fijos			Resultado
		(0, 0)	(1, 0)	(0, 1)	
I	$a_{12} < 1; a_{21} < 1$	repulsor	saddle		atractor (N_1, N_2)
II	$a_{12} > 1; a_{21} > 1$	repulsor	atractor		saddle $(1, 0), (0, 1)$
III	$a_{12} < 1; a_{21} > 1$	repulsor	atractor	saddle	no existe $(1, 0)$
IV	$a_{12} > 1; a_{21} < 1$	repulsor	saddle	atractor	no existe $(0, 1)$

Cuadro 9.2: Dinámicas del modelo competitivo de *Lokta-Volterra*.

La Figura que sigue ilustra los planos de fase de los cuatro tipos de dinámicas. Las rectas son las nulclinas. Las “flechas” muestran hacia dónde va el flujo.

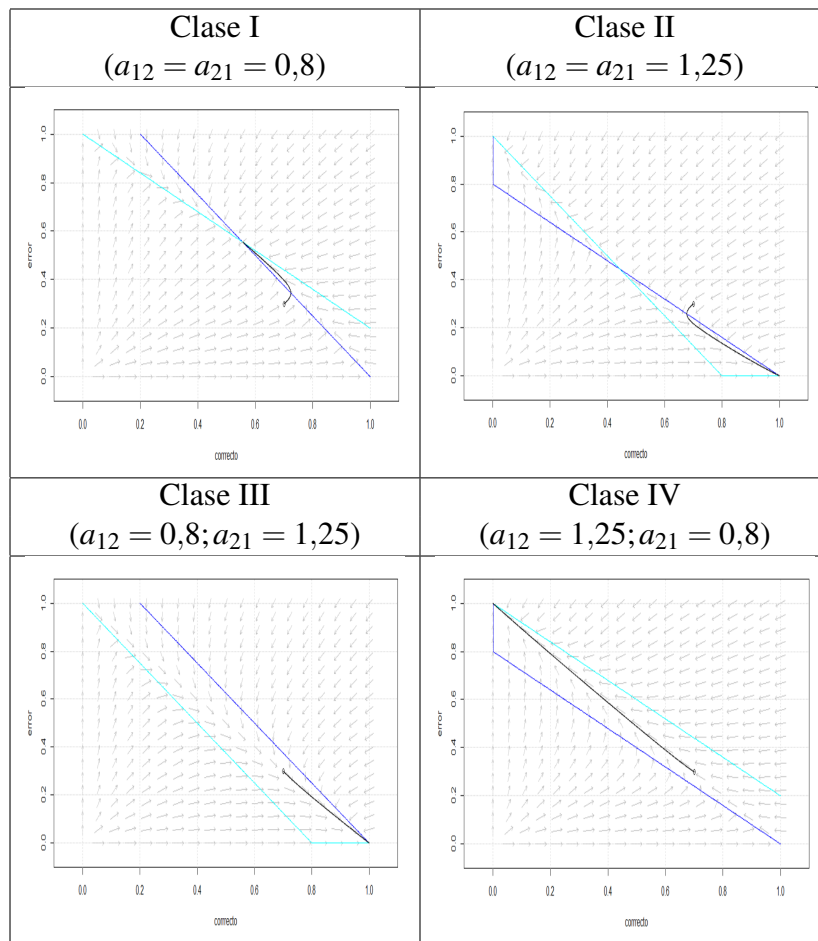


Figura 9.1: Planos de Fase de los cuatro tipos de dinámica. Para todos los casos: $r_1 = r_2 = 0,5; a_{11} = a_{22} = 1$. La condición inicial es siempre $N_0 = (0,7,0,3)$.

No hay aplicaciones a la adquisición de lenguas extranjeras. En cuanto a la concordancia en *L1*, Smith, Frank & Tabor (2018) propusieron un enfoque basado en el modelo de *Lokta-Volterra* generalizado a n especies, para simular datos experimentales del fenómeno de atracción de concordancia, en construcciones partitivas (“un tarro de galletitas ESTÁ/ESTÁN”). El modelo tenía un atractor para NP (“un tarro”) y otro

para *PP* (“de galletitas”), lo cual producía concordancia plural o singular en el verbo. El modelo lograba aproximar las probabilidades de los patrones de concordancia experimentales de los sujetos (ver cap. 3).

9.2.1. Aplicación y resultados.

Como es necesario que haya dos atractores, la simulación se basó en la clase III. Se tomaron los siguientes parámetros: $r_1 = r_2 = 0,5$; $\alpha_{11} = \alpha_{22} = 1$; $\alpha_{12} = \alpha_{21} = 1,5$. Con lo cual, el punto interior *saddle* fue $P_4 = \left(\frac{1-1,5}{1-(1,5)^2}; \frac{1-1,5}{1-(1,5)^2} \right) = (0,4; 0,4)$, y los atractores son: $(1, 0) = \text{“correcto”}$ y $(0, 1) = \text{“error”}$. Para la integración numérica se realizaron 100 iteraciones. Se tomó la siguiente estrategia para sesgar el flujo hacia alguno de los dos atractores:

1. Calcular la distancia entre las variables de cada contexto de la tabla 1 y las mismas variables en la instancia en cuestión. Se cuentan cuántas de ellas *no* coinciden en sus valores. También se computa la igualdad consigo misma, la cual es *cero*. O sea: $m_i = \# \text{variables diferentes}; i = 1, \dots, 5$. Luego se calcula la probabilidad para cada m_i bajo una distribución logística¹ *truncada* entre -1 y 6 , con locación igual a 2 y escala de 1 . Ya que los posibles valores de m_i van entre 0 y 5 . La función de densidad para estos valores es: $P(0) \approx 0,07, P(1) \approx 0,23, P(2) \approx 0,48, P(3) \approx 0,73, P(4) \approx 0,89, P(5) \approx 0,96$.
2. Se calcula la siguiente suma pesada: $E = w \times \text{media}(P(m_1), \dots, P(m_5)) + (1 - w) \times e_{\text{antes}}$; donde e_{antes} es la proporción de errores hasta la instancia anterior y $w \in (0, 1)$ se parametrizó utilizando el error cuadrático de predicción: $RMSE = \sqrt{\frac{1}{4} \sum_{i=1}^4 (p_{\text{sim},i} - p_{\text{error},i})^2}$, donde p_{sim} es la proporción de instancias marcadas como “error” de la simulación (o sea, aquellas en donde gana el atractor “error”) y p_{real} la proporción de errores real, para cada aprendiente. Se elige el valor de w con $RMSE$ mínimo. Se obtuvo $w = 0,71$.
3. Se toma como condición inicial para las ecuaciones diferenciales $N_0 = (1 - E, E)$. Si $E < \frac{1}{2}$, gana el atractor “correcto”; y si $E > \frac{1}{2}$, gana el atractor “error”. Si $E = \frac{1}{2}$, se elige al azar entre los valores $\{0,25, 0,75\}$. Esto último es así dado que la condición inicial $(\frac{1}{2}, \frac{1}{2})$ lleva el flujo hacia el nodo *saddle* P_4 (porque el punto inicial se encuentra sobre la *separatrix*).

Se ejemplifica el procedimiento con la instancia “mucho pensadores” [SONIA, sesión 1, línea 173]. La instancia posee las siguientes variables: MOD = “2”; Fabs.SC.f = “0”; FAM.LEX.f = “0”; MORF.f = “2”; ES = “1”; EST1 = “0”; “EST2” = “0”; EST5 = “1”. Además, el error hasta la instancia anterior es $e_{\text{antes}} = 0,5$. Por supuesto, consigo misma la distancia es *cero*. Con el contexto que baja la chance de error de género, la distancia es *cuatro* (de las cuatro variables que conforman el contexto, la instancia no coincide en sus valores con ninguna [comparar con el Cuadro 1]). Respecto de los contextos de “e-epentética”, “plural” y “mixto” las distancias son, respectivamente: *cuatro, tres, tres*. Por lo tanto: $m_1 = 0, m_2 = 4, m_3 = 4, m_4 = 3, m_5 = 3$. Las probabilidades según la distribución logística truncada son: $P(0) \approx 0,07; P(4) \approx 0,73; P(4) \approx 0,89$. Con lo cual el promedio de las probabilidades de las distancias resulta: $\frac{1}{5} \sum_{i=1}^5 P(m_i) = \frac{1}{5} (0,07 + 0,89 + 0,89 + 0,73 + 0,73) = 0,66$. Con lo cual: $E = w \times \text{media}(P(m_1), \dots, P(m_5)) + (1 - w) e_{\text{antes}} = 0,71 \times 0,66 + 0,29 \times 0,5 = 0,61$. Por lo tanto, las condiciones iniciales son $N_0 = (1 - E, E) = (0,39, 0,61)$. Dado que $0,61 > 0,5$, el flujo se dirige hacia el atractor $(0, 1) = \text{“error”}$.

A continuación se muestra la abundancia de $N_1 = \text{“correcto”}$ y $N_2 = \text{“error”}$ para dos instancias: *muchos personas* [SONIA, sesión 1, línea 56] y *otros autores* [SONIA, sesión 1, línea 187]. Además se ilustra el plano de fase con la recorrida del flujo.

¹La función de distribución acumulada de la distribución logística es: $F(X < x; \mu, \sigma) = \frac{1}{1 + e^{-(x-\mu)/\sigma}}$; con: $x \in (-\infty, +\infty)$, locación $\mu \in \mathbb{R}$ y escala $\sigma > 0$.

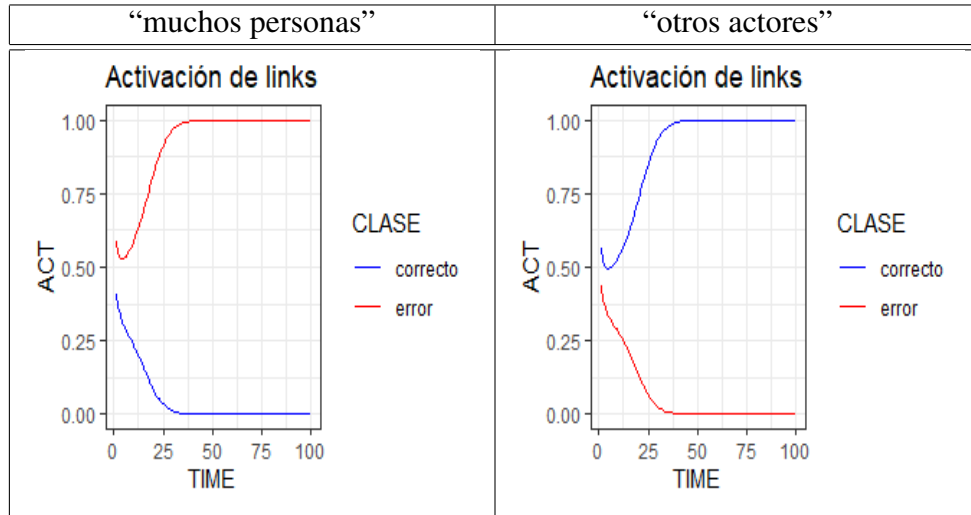


Figura 9.2: Abundancias de las especies “correcto” y “error” para las instancias “muchos personas” y “otros autores”.

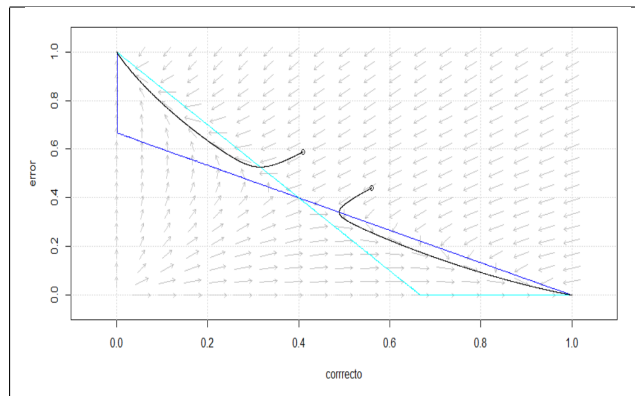


Figura 9.3: Plano de fase para los flujos de las instancias: “muchos personas” y “otros autores”.

El Cuadro que sigue muestra: (i) las proporciones globales de “error” entre la simulación y los datos; (ii) los p-valores de un test asintótico de diferencia entre proporciones entre los errores simulados y los errores de los datos; (iii) los valores de la distancia de Minkowski $D = \sqrt[p]{(x_i - y_i)^p}$; ($p = 3$), entre los errores simulados y reales por sesión (a menor distancia, mejor). Por otro lado, las Figuras muestran un gráfico de barras de proporción de errores y de correctos; así como una gráfica que compara los datos reales y los de la simulación respecto de la proporción de errores en la serie temporal de sesiones.

Se observa que SONIA subestima el error por aproximadamente cinco puntos y JAKO lo sobre-estima por la misma cantidad. Sin embargo, las diferencias entre las proporciones de error no resultan significativas. En cuanto al patrón por sesiones SONIA y NATI se aproximan mejor a las series de tiempo de los datos y MIRKA resulta la de peor desempeño.

	error.sim	error.real	p.value	dist
SONIA	0.160	0.215	0.101	8.392
NATI	0.318	0.336	0.647	8.316
JAKO	0.209	0.164	0.092	11.363
MIRKA	0.258	0.291	0.185	20.279

Cuadro 9.3: Resultados de la simulación: (i) proporciones entre errores reales y simulados; (ii) p-valores de un test para diferencia entre dichas proporciones; (iii) distancia de Minkowski entre los errores simulados y reales por sesiones.

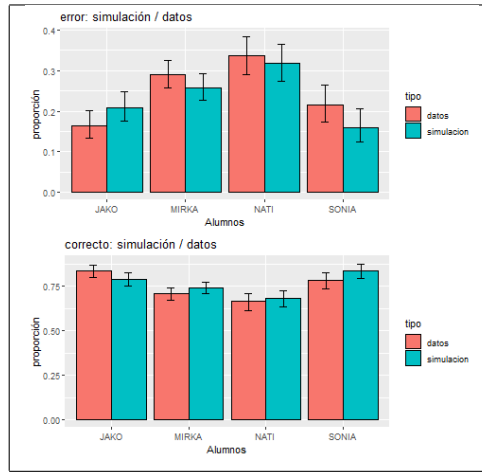


Figura 9.4: Barras de proporción de errores simulados y reales; y de correctos simulados y reales.

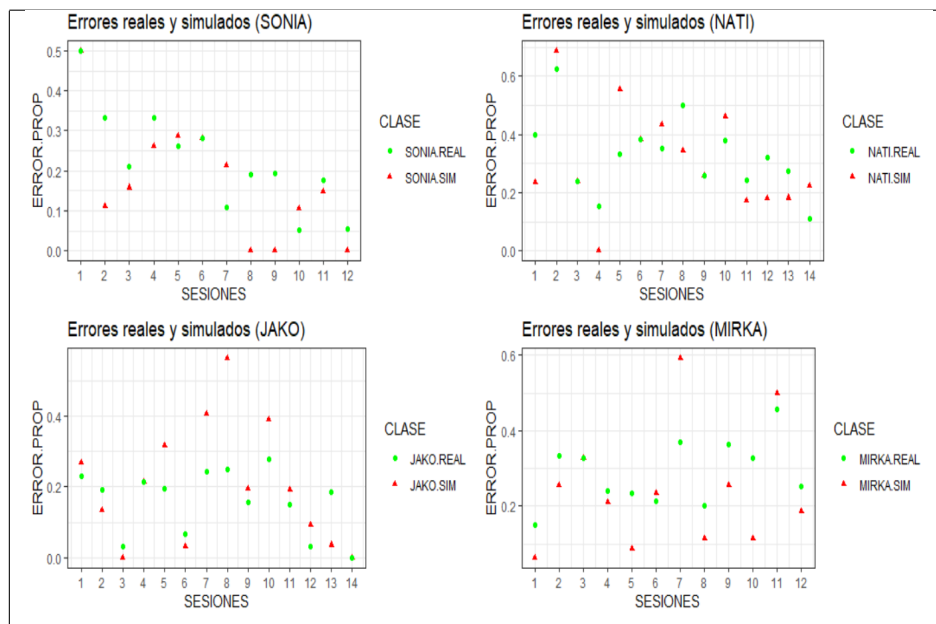


Figura 9.5: Series de tiempo de proporción de errores simulados y reales, por sesiones.

9.3. Teoría de los juegos evolutiva.

Se presentan conceptos clave de teoría de los juegos evolucionista (se sigue a: Weibull, 1997). Se restringe el tratamiento a la forma normal de juegos finitos. Sea $I = \{1, 2, \dots, n\}$ el conjunto de jugadores ($n \in \mathbb{N}$). Para cada jugador $i \in I$, sea $S_i = \{1, 2, \dots, m_i\}$ su conjunto de estrategias puras ($m_i > 2$). Además, sea $s = (s_1, s_2, \dots, s_n)$ el vector de estrategias para el jugador i , su perfil de estrategias puras. La función de utilidad $u_i : s_1 \times s_2 \times \dots \times s_n \rightarrow \mathbb{R}$ asocia un pago o ganancia a las estrategias. Con dos jugadores las funciones de pago u_1 y u_2 se representan como una matriz de dimensión $m_1 \times m_2$. La matriz del primer jugador se denota $A = (a_{hk})$ donde $a_{hk} = u_1(h, k)$ para $h \in S_1$ y $k \in S_2$, es decir, el pago cuando el primer jugador juega la estrategia h dado que el segundo juega la estrategia k . Análogamente, para el segundo jugador $B = (b_{hk})$ donde $b_{hk} = u_2(h, k)$. Cada fila de la matriz es una estrategia pura del primer jugador 1 y cada columna, una estrategia pura del jugador 2. Las matrices se pueden combinar en una bi-matriz (A, B) . En el ejemplo siguiente (“dilema del prisionero”) hay dos jugadores $I = \{1, 2\}$, cada uno con las mismas dos estrategias $S_1 = S_2 = (1, 2)$ y $s_i = (C = \text{“cooperate”}, D = \text{“defect”})$; y donde $a_{11} = u_1(1, 1) = 3$ significa que si el primer jugador juega la estrategia (fila) 1 cuando el segundo también juega la estrategia (columna) 1, entonces recibirá un pago de 3. Análogamente para $b_{11} = u_2(1, 1)$.

$$(A,B) = \begin{matrix} & C & D \\ C & 3,3 & 0,4 \\ D & 4,0 & 2,2 \end{matrix} \Rightarrow A = \begin{pmatrix} 3 & 0 \\ 4 & 2 \end{pmatrix}; B = \begin{pmatrix} 3 & 4 \\ 0 & 2 \end{pmatrix}$$

juego simétrico de dos jugadores. Si hay dos jugadores $I = \{1,2\}$ con el mismo número de estrategias $S_1 = S_2$ y $u_2(h,k) = u_1(k,h)$, o sea si $A^T = B$. Es decir que el pago cuando el jugador 2 usa la estrategia h y el jugador 1 la estrategia k es el mismo que si el jugador 1 usa la estrategia k y el jugador 2 la estrategia h . El ejemplo anterior es simétrico. Por otra parte, un juego es *doblemente* simétrico si $A^T = A$ con $B = A$.

estrategia mixta. Una estrategia mixta para el jugador i es una distribución de probabilidad sobre el conjunto S_i de estrategias puras. El perfil de estrategias mixtas para i es el vector $x_{ih} = (x_1, \dots, x_{m_i})^T \in \mathbb{R}_+^{m_i}$, donde cada componente h indica la probabilidad asociada a la estrategia pura h (y m es el número de estrategias puras). Ya que todas las probabilidades son no-negativas y suman *uno*, dicho vector pertenece al *simplex* unitario Δ_i de dimensión $m_i - 1$ definido como: $\Delta_i = \left\{ x_i \in \mathbb{R}_+^{m_i} : \sum_{h=1}^{m_i} x_{ih} = 1 \right\}$. Los vértices del simplex son $e_i^1 = (1,0, \dots, 0)$, $e_i^2 = (0,1, \dots, 0)$, ..., $e_i^{m_i} = (0,0, \dots, 1)$, los cuales se corresponden con las estrategias puras. Las funciones de pago asociadas con los perfiles de estrategias mixtas $x_1 \in \Delta_1$ y $x_2 \in \Delta_2$ para dos jugadores serán la esperanzas:

$$u_1(x) = \sum_{h=1}^{m_1} \sum_{k=1}^{m_2} x_{1h} a_{hk} x_{2k} = x_1^T A x_2$$

$$u_2(x) = \sum_{h=1}^{m_1} \sum_{k=1}^{m_2} x_{1h} b_{hk} x_{2k} = x_1^T B x_2$$

Por ejemplo, sea el siguiente juego (“batalla de los sexos”):

$$(A,B) = \begin{matrix} & F & B \\ F & 2/3 & 1/3 \\ B & 1/3 & 0,0 \end{matrix} \Rightarrow A = \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix}; B = \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix}$$

las estrategias son $s_1 = s_2 = (B = \text{”Ballet”}, F = \text{”Football”})$ y las probabilidades asociadas (estrategias mixtas) son $x_1 = (\frac{2}{3}, \frac{1}{3})$ y $x_2 = (\frac{1}{3}, \frac{2}{3})$. La ganancia esperada de los jugadores² son $u_1(x) = u_2(x) = \frac{2}{3}$.

dominancia estricta. Una estrategia domina estrictamente a otra si obtiene siempre un pago más alto, cualquiera sea la estrategia que juegue el otro jugador. Es decir: $y_i \in \Delta_i$ domina estrictamente a $x_i \in \Delta_i$ si $u_i(y_i, z_{-i}) > u_i(x_i, z_{-i})$. Por ejemplo, en el juego del “dilema del prisionero” la estrategia pura $D = \text{”defect”}$ domina estrictamente a $C = \text{”cooperate”}$. En efecto, para el jugador 1, $u_1(D,C) = 4$, $u_1(D,D) = 2$ es mayor que $u_1(C,C) = 3$, $u_1(C,D) = 0$. Sin importar si 2 juega D o C , jugar la estrategia D paga siempre más para el jugador 1.

mejor respuesta (BR). Una mejor respuesta para el jugador i frente al perfil de estrategias y del otro jugador es una estrategia *pura* $s_i \in S_i$ tal que ninguna otra estrategia que el jugador tenga disponible le dé un mejor pago frente a y . En el “dilema del prisionero” si cualquier jugador juega la estrategia C lo mejor que puede hacer el otro es jugar D ; y, si cualquier jugador juega D , lo mejor que puede hacer el otro también es jugar D . Por lo tanto, D , es mejor respuesta.

²Efectivamente,

$$u_1(x) = x_1^T A x_2 = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \end{pmatrix} \begin{pmatrix} 2 & 0 \\ 0 & 1 \end{pmatrix} \begin{pmatrix} \frac{1}{3} \\ \frac{2}{3} \end{pmatrix} = \begin{pmatrix} \frac{4}{3} & \frac{1}{3} \end{pmatrix} \begin{pmatrix} \frac{1}{3} \\ \frac{2}{3} \end{pmatrix} = \frac{4}{3} \cdot \frac{1}{3} + \frac{1}{3} \cdot \frac{2}{3} = \frac{4}{9} + \frac{2}{9} = \frac{6}{9} = \frac{2}{3}$$

$$u_2(x) = x_1^T B x_2 = \begin{pmatrix} \frac{2}{3} & \frac{1}{3} \end{pmatrix} \begin{pmatrix} 1 & 0 \\ 0 & 2 \end{pmatrix} \begin{pmatrix} \frac{1}{3} \\ \frac{2}{3} \end{pmatrix} = \begin{pmatrix} \frac{2}{3} & \frac{2}{3} \end{pmatrix} \begin{pmatrix} \frac{1}{3} \\ \frac{2}{3} \end{pmatrix} = \frac{2}{3} \cdot \frac{1}{3} + \frac{2}{3} \cdot \frac{2}{3} = \frac{2}{9} + \frac{4}{9} = \frac{6}{9} = \frac{2}{3}$$

Una mejor respuesta BR para el jugador i frente al perfil de estrategias $y \in \Delta_{-i}$ del otro jugador es una estrategia mixta $x_i \in \Delta_i$ tal que ninguna otra estrategia mixta que el jugador tenga disponible le de un mejor pago frente a y . Formalmente:

$$BR_i(y) = \{x \in \Delta_i : u(x, y) > u(w, y), \forall w \in \Delta_i\}$$

equilibrio de Nash (NE). Para un juego bimatricial (h^*, k^*) hay equilibrio de Nash si un jugador supone que el otro se queda con una determinada estrategia de una fila / columna, al otro no le conviene pasarse de fila / columna. O sea si el jugador 2 elige la columna k^* y el 1 elige la fila h^* , al jugador 1 no le conviene desviarse y elegir otra estrategia (elegir otra fila $-h$) porque pierde ganancia. Lo mismo sucede desde la perspectiva del jugador 2. En el “dilema del prisionero” $(e^2, e^2) = (2, 2)$ es el equilibrio de Nash.

$$Nash = \begin{cases} a_{h^*, k^*} \geq a_{-h, j^*} \\ b_{h^*, k^*} \geq b_{h^*, -j} \end{cases}$$

En el caso de las estrategias mixtas, el perfil de estrategias mixtas $x_{ih}^* = (x_1^*, \dots, x_{m_i}^*) \in \Delta_i$ es un equilibrio de Nash si para todo jugador $i = 1, \dots, N$ y para todo $x_i \in \Delta_i$, se tiene que: $u_i(x_i^*, x_{-i}^*) \geq u_i(x_i, x_{-i}^*)$. O sea que para el jugador i , la estrategia mixta x_i^* es una mejor respuesta a x_{-i}^* . Para estrategias mixtas, existe al menos un equilibrio de Nash. En el caso de la “batalla de los sexos” hay dos equilibrios de Nash puros: $(1, 2)$ y $(2, 1)$. El equilibrio mixto es³: $(x_{1h}^* = (p_1^*, 1 - p_1^*) = (\frac{2}{3}, \frac{1}{3}), x_{2h}^* = (p_2^*, 1 - p_2^*) = (\frac{1}{3}, \frac{2}{3}))$.

En ambos casos es un *Nash estricto* si la desigualdad es estricta ($>$). En otras palabras, un NE es un par de respuestas $(\bar{x}, \bar{y}) \in \Delta_1 \times \Delta_2 = \Delta$ tal que \bar{x} es una mejor respuesta a \bar{y} e \bar{y} es una mejor respuesta a \bar{x} . En un Nash estricto, \bar{x} e \bar{y} son las *únicas* mejores respuestas, la una contra la otra.

Estrategia evolutivamente estable (ESS). Supóngase una población grande de individuos que son seleccionados aleatoriamente para jugar un juego simétrico de a dos. Cada individuo está programado para jugar las mismas estrategias puras. Sea $S = \{1, 2, \dots, k\}$ el conjunto de dichas estrategias. Sea $\Delta_i = \left\{ x \in \mathbb{R}_+^k : \sum_{i=1}^k x_i = 1 \right\}$ el conjunto de estados poblacionales, o sea que cada x_i en el vector $x = (x_1, \dots, x_k)$ es la fracción de individuos que juega la estrategia $i \in S$ (notar que Δ es idéntico formalmente al conjunto de estrategias mixtas). Supóngase, además, que inicialmente el estado poblacional es $x \in \Delta$, entonces la ganancia poblacional promedio (“fitness”) es $u(x, x) = x^T A x$. A continuación, entra a la población un *pequeño grupo* de mutantes con fracción poblacional ε que juega el perfil $y \in \Delta$. Entonces, luego de la introducción de los mutantes, el estado poblacional será $w = (\varepsilon y + (1 - \varepsilon)x)$. Luego de la introducción de los mutantes, la ganancia poblacional promedio (“fitness”) de los no-mutantes que son apareados aleatoriamente con mutantes será $u(x, y) = x^T A y$. La ganancia promedio de los no-mutantes será $u(x, w) = u(x, \varepsilon y + (1 - \varepsilon)x) = \varepsilon u(x, y) + (1 - \varepsilon) u(x, x)$. Análogamente, la ganancia promedio de los mutantes será: $u(y, w) = u(y, \varepsilon y + (1 - \varepsilon)x) = \varepsilon u(y, y) + (1 - \varepsilon) u(y, x)$. Se supone que las “fuerzas evolutivas” no van a seleccionar a la estrategia del mutante si su “fitness” resulta ser menor que el del no-mutante. Es decir si:

$$u(x, w) = \varepsilon u(x, y) + (1 - \varepsilon) u(x, x) > u(y, w) = \varepsilon u(y, y) + (1 - \varepsilon) u(y, x)$$

Se dice que x es una estrategia evolutivamente estable (ESS) si para todo $x \neq y$ y para todo $\varepsilon \ll 1$ vale la desigualdad anterior. Es decir que x es resistente a las perturbaciones. Los mutantes no pueden invadir la población y eventualmente se extinguen. Ahora bien, la desigualdad anterior puede escribirse como:

³Para encontrar p_1^* se resuelve:

$$(p_1^*, 1 - p_1^*) B \begin{pmatrix} 1 \\ 0 \end{pmatrix} = (p_1^*, 1 - p_1^*) B \begin{pmatrix} 0 \\ 1 \end{pmatrix} \Leftrightarrow p_1^* = 2(1 - p_1^*) \Leftrightarrow p_1^* = \frac{2}{3} \Rightarrow 1 - p_1^* = \frac{1}{3}$$

Para encontrar p_2^* se resuelve:

$$(1, 0) A \begin{pmatrix} p_2^* \\ 1 - p_2^* \end{pmatrix} = (0, 1) A \begin{pmatrix} p_2^* \\ 1 - p_2^* \end{pmatrix} \Leftrightarrow 2p_2^* = (1 - p_2^*) \Leftrightarrow p_2^* = \frac{1}{3} \Rightarrow 1 - p_2^* = \frac{2}{3}$$

$$u(x,x) - u(y,x) + \varepsilon \{u(x,y) - u(x,x) + u(y,x) - u(y,y)\} > 0; \varepsilon \ll 1.$$

Es fácil ver que para que esta última desigualdad sea cierta deben cumplirse las condiciones: (1) $u(x,x) \geq u(y,x)$; o sea es un equilibrio de Nash; (2) si $u(x,x) = u(y,x)$, es necesario que sea cierto que: $u(x,y) > u(y,y)$.

En resumen: x es ESS si y solo si:

- (x,x) es un equilibrio de Nash, o sea: $u(x,x) = x^T Ax \geq u(y,x) = y^T Ax$.
- Si sucede que $x^T Ap = y^T Ax$, entonces: $u(x,y) = x^T Ay > u(y,y) = y^T Ay$.

Obsérvese que si x es un Nash estricto entonces es ESS (porque no se da la segunda condición). A modo de ejemplo, considérese el siguiente “juego de coordinación”:

$$A = \begin{pmatrix} 2 & 3 \\ 0 & 4 \end{pmatrix}$$

Los equilibrios de Nash son $NE = \{e^1, e^2, \hat{x}\}$ donde $\hat{x} = (\frac{1}{3}, \frac{2}{3})$ es la estrategia mixta⁴. Las dos primeras puras son Nash estrictos, por lo tanto son ESS. Para examinar si la estrategia mixta es ESS, supóngase la estrategia mixta del mutante $y = (p, 1-p)$. Primero se debe ver si hay un Nash estricto: $x^T Ax \geq y^T Ax$; no lo hay porque $x^T Ax = y^T Ax = \frac{8}{3}$. O sea que tanto los mutantes como los no-mutantes cuando son seleccionados para jugar contra los mutantes reciben un “fitness”⁵ de $\frac{8}{3}$. Como se cumple la igualdad se debe verificar la segunda condición: $x^T Ay > y^T Ay$; pero $x^T Ay = \frac{11}{3} - \frac{9}{3}p$ y $y^T Ay = 4p^2 - 9p + 7$; con lo cual: $\frac{11}{3} - \frac{9}{3}p > 4p^2 - 9p + 7 \iff -4p^2 + 6p - \frac{10}{3} > 0$ ($p \in (0, 1)$) no es cierta, entonces⁶: $x^T Ay < y^T Ay$. Por lo tanto, los no-mutantes que se aparean aleatoriamente con mutantes tienen un “fitness” menor que los mutantes que se aparean aleatoriamente con mutantes. Como no se cumple ninguna de las dos condiciones, $\hat{x} = (\frac{1}{3}, \frac{2}{3})$ no es ESS.

Sea ahora el siguiente juego simétrico generalizado para dos jugadores cuya matriz de pago es A . Restando a_{21} de la primera columna y a_{12} de la segunda columna, se obtiene la matriz normalizada equivalente A^* que resulta ser doblemente simétrica, donde $a_1 = a_{11} - a_{21}$ y $a_2 = a_{22} - a_{12}$.

$$A = \begin{pmatrix} a_{11} & a_{12} \\ a_{21} & a_{22} \end{pmatrix} \implies A^* = \begin{pmatrix} a_{11} - a_{21} & 0 \\ 0 & a_{22} - a_{12} \end{pmatrix} = \begin{pmatrix} a_1 & 0 \\ 0 & a_2 \end{pmatrix}$$

Estableciendo condiciones sobre a_1 y a_2 , es posible clasificar los juegos simétricos de 2×2 en cuatro clases, y caracterizarlos según sus NE y ESS. Esto se muestra en el siguiente Cuadro. Notar que la clase IV es la imagen especular de la clase I.

⁴La cual se encuentra resolviendo:

$$\begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} 2 & 3 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} p^* \\ 1-p^* \end{pmatrix} = \begin{pmatrix} 0 & 1 \end{pmatrix} \begin{pmatrix} 2 & 3 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} p^* \\ 1-p^* \end{pmatrix} \iff 2p+3(1-p^*) = 4(1-p^*) \Rightarrow p^* = \frac{1}{3}, (1-p^*) = \frac{2}{3}$$

⁵Esto se ve calculando:

$$x^T Ax = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \end{pmatrix} \begin{pmatrix} 2 & 3 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} \frac{1}{3} \\ \frac{2}{3} \end{pmatrix} = \begin{pmatrix} \frac{2}{3} & \frac{11}{3} \end{pmatrix} \begin{pmatrix} \frac{1}{3} \\ \frac{2}{3} \end{pmatrix} = \frac{8}{3} \quad \text{e} \quad y^T Ax = \begin{pmatrix} p & 1-p \end{pmatrix} \begin{pmatrix} 2 & 3 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} \frac{1}{3} \\ \frac{2}{3} \end{pmatrix} = \begin{pmatrix} 2p & 4-p \end{pmatrix} \begin{pmatrix} \frac{1}{3} \\ \frac{2}{3} \end{pmatrix} = \frac{8}{3}$$

⁶Efectivamente:

$$x^T Ax = \begin{pmatrix} \frac{1}{3} & \frac{2}{3} \end{pmatrix} \begin{pmatrix} 2 & 3 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} p \\ 1-p \end{pmatrix} = \begin{pmatrix} \frac{2}{3} & \frac{11}{3} \end{pmatrix} \begin{pmatrix} p \\ 1-p \end{pmatrix} = \frac{2}{3}p + \frac{11}{3}(1-p) = \frac{11}{3} - \frac{9}{3}p$$

$$y^T Ay = \begin{pmatrix} p & 1-p \end{pmatrix} \begin{pmatrix} 2 & 3 \\ 0 & 4 \end{pmatrix} \begin{pmatrix} p \\ 1-p \end{pmatrix} = \begin{pmatrix} 2 & 7-4p \end{pmatrix} \begin{pmatrix} p \\ 1-p \end{pmatrix} = 2p + (7-4p)(1-p) = 4p^2 - 9p + 7$$

Clase	Ejemplo representativo	Condición	NE	ESS
I	“Prisoner’s dilemma”	$a_1 < 0, a_2 > 0$	(e^2, e^2)	e^2
II	“Coordination game”	$a_1 > 0, a_2 > 0$	$(e^1, e^1), (e^2, e^2), (\hat{x}, \hat{x}) ; \hat{x} = \left(\frac{a_2}{a_1+a_2}, \frac{a_1}{a_1+a_2} \right)$	e^1, e^2
III	“Hawk-Dove” ⁷	$a_1 < 0, a_2 < 0$	$(e^1, e^2), (e^2, e^1), (\hat{x}, \hat{x}) ; \hat{x} = \left(\frac{a_2}{a_1+a_2}, \frac{a_1}{a_1+a_2} \right)$	$(e^1, e^2), (e^2, e^1), \hat{x}$
IV	-	$a_1 > 0, a_2 < 0$	(e^1, e^1)	e^1

Cuadro 9.4: Clasificación de juegos simétricos 2×2 según sus conjuntos de NE y ESS.

9.3.1. Dinámica del replicador.

En la sección anterior se ha visto que $ESS \Rightarrow NE$ (pero la recíproca no es cierta, excepto si se trata de un NE estricto). Sin embargo, ambos conceptos representan equilibrios estáticos. En este apartado se supone ahora que cada individuo está programado para usar una estrategia pura, que *replica* a su descendencia. La dinámica del replicador muestra cómo la distribución de estrategias puras de la población cambia con el tiempo siguiendo un proceso de selección, y que dicha distribución converge a un ESS . Se presenta el modelo para el caso de dos jugadores. Sea $t \in [0, \infty)$ el tiempo continuo; y denótese $x_1(t)$ a la fracción de individuos que juega la estrategia pura 1 en el tiempo t . En consecuencia, $x_2(t) = 1 - x_1(t)$ será la fracción de individuos que juega la estrategia pura 2 en el tiempo t . Por tanto, el estado de la población es $x(t) = [x_1(t), 1 - x_1(t)] \in \Delta$. Ahora bien, sea el juego simétrico 2×2 : $A = \begin{pmatrix} a & b \\ c & d \end{pmatrix}$. El pago promedio de un individuo que usa la estrategia 1 es:

$$u(e^1, x) = e_1^T A \begin{pmatrix} x(t) \\ 1 - x(t) \end{pmatrix} = \begin{pmatrix} 1 & 0 \end{pmatrix} \begin{pmatrix} a & b \\ c & d \end{pmatrix} \begin{pmatrix} x(t) \\ 1 - x(t) \end{pmatrix} = ax(t) + b(1 - x(t))$$

Por otro lado, el pago promedio de un individuo contra cualquier otro (promedio poblacional) será:

$$u(x, x) = \begin{pmatrix} x(t) & 1 - x(t) \end{pmatrix} A \begin{pmatrix} x(t) \\ 1 - x(t) \end{pmatrix} = ax^2 + bx(t)(1 - x(t)) + c(1 - x(t))x(t) + d(1 - x(t))^2$$

La ecuaciones diferenciales para la dinámica del replicador son pues:

$$\frac{dx_1}{dt} = x_1 [u(e^1, x) - u(x, x)] = x_1(1 - x_1) [x_1(a + d - b - c) + b - d]$$

$$\frac{dx_2}{dt} = -\frac{dx_1}{dt}$$

Por lo tanto, la tasa de crecimiento de la fracción poblacional de la estrategia 1 es proporcional a la diferencia entre su “fitness” y el “fitness” poblacional. Si esta diferencia es positiva, la tasa de crecimiento aumentará; y habrá más individuos jugando dicha estrategia en la próxima generación, en $t + dt$. Ya que $\dot{x}_2 = -\dot{x}_1$, la dimensión del sistema es 1. Por lo tanto, se concentrará el análisis en la primera ecuación. Sus puntos fijos serán los que hagan que $\dot{x}_1 = 0$, lo cual sucede cuando $P_1 = (x_1^* = 1)$, $P_2 = (x_1^* = 0)$, $P_3 = (x_1^* = \frac{d-b}{a+d-b-c})$. En el caso de los juegos simétricos de 2×2 con matriz normalizada: $A = \begin{pmatrix} a_1 = a - b & 0 \\ 0 & a_2 = d - b \end{pmatrix}$, y cuando $a_1 a_2 \neq 0$, la dinámica se vuelve: $\frac{dx_1}{dt} = x_1(1 - x_1) [a_1 x_1 - a_2(1 - x_1)]$. En cuyo caso, el punto fijo P_3 se volverá $x_1^* = \frac{a_2}{a_1 + a_2}$. El Cuadro que sigue muestra el análisis de estabilidad para cada punto fijo, respecto de cada clase de juego simétrico. En el apéndice se muestra el análisis detallado. En lo que respecta a la condición $a_1 a_2 < 0$ (clases I y IV); si $a_1 < 0, a_2 > 0$ (clase I) la proporción de

⁷ Este juego es: $A = \begin{pmatrix} \frac{v-c}{2} & v \\ 0 & \frac{v}{2} \end{pmatrix}$; y normalizado es: $A^* = \begin{pmatrix} a_1 = \frac{v-c}{2} & 0 \\ 0 & a_2 = -\frac{v}{2} \end{pmatrix}$;

Por lo tanto: $\hat{x} = \left(\frac{a_2}{a_1+a_2}, \frac{a_1}{a_1+a_2} \right) = \left(\frac{v}{c}, 1 - \frac{v}{c} \right)$.

la estrategia 1 siempre descende, con lo cual el sistema se aproxima a $P_2 = 0$, extinguiéndose, con lo cual gana siempre la estrategia 2 (e^2). Por otro lado, si $a_1 > 0, a_2 < 0$ (clase IV), la proporción de la estrategia 1 siempre crece, y el flujo es atraído hacia $P_1 = 1$, con lo cual gana la estrategia 1 (e^1). En lo que atañe a la condición $a_1 a_2 > 0$ (clases II y III), si $a_1 > 0, a_2 > 0$, existe un nodo interior repulsor, que coincide con la primera estrategia del equilibrio de Nash mixto $\hat{x} = (\hat{x}_1, \hat{x}_2)$, $\hat{x}_1 = \frac{a_2}{a_1 + a_2}$. Recuérdese que en el juego de coordinación \hat{x} no es ESS, con lo cual es un equilibrio inestable: las condiciones iniciales $x_1^0 > \hat{x}$ son atraídas hacia $P_1 = 1$ (gana la estrategia e^1) y aquellas $x_1^0 < \hat{x}$ son atraídas hacia $P_2 = 0$ (gana la estrategia e^2). Por lo tanto los “basin of attraction” [BA] son: $BA(P_1) = (\hat{x}, 1)$ y $BA(P_2) = (0, \hat{x})$. Por otro lado, si $a_1 < 0, a_2 < 0$ (clase III), \hat{x} es un ESS, por lo tanto es un equilibrio estable (atractor), con lo cual atrae las órbitas que comiencen fuera de dicho punto (excepto, por supuesto, en los puntos fijos), y los puntos P_1 y P_2 son consecuentemente repulsivos. La Figura muestra el estado de fase para las cuatro clases.

Clase	Ejemplo representativo	Condición	P_1	P_2	P_3	Resultado (ESS)
I	“Prisoner’s dilemma”	$a_1 < 0, a_2 > 0, a_1 a_2 < 0$	repulsor	atractor	no existe	e^2
II	“Coordination game”	$a_1 > 0, a_2 > 0, a_1 a_2 > 0$	atractor	atractor	repulsor	e^1, e^2
III	“Hawk-Dove”	$a_1 < 0, a_2 < 0, a_1 a_2 > 0$	repulsor	repulsor	atractor	$\hat{x}_1 = \frac{a_2}{a_1 + a_2}$
IV	-	$a_1 > 0, a_2 < 0, a_1 a_2 < 0$	atractor	repulsor	no existe	e^1

Cuadro 9.5: Clasificación de juegos simétricos 2×2 según la dinámica del replicador.

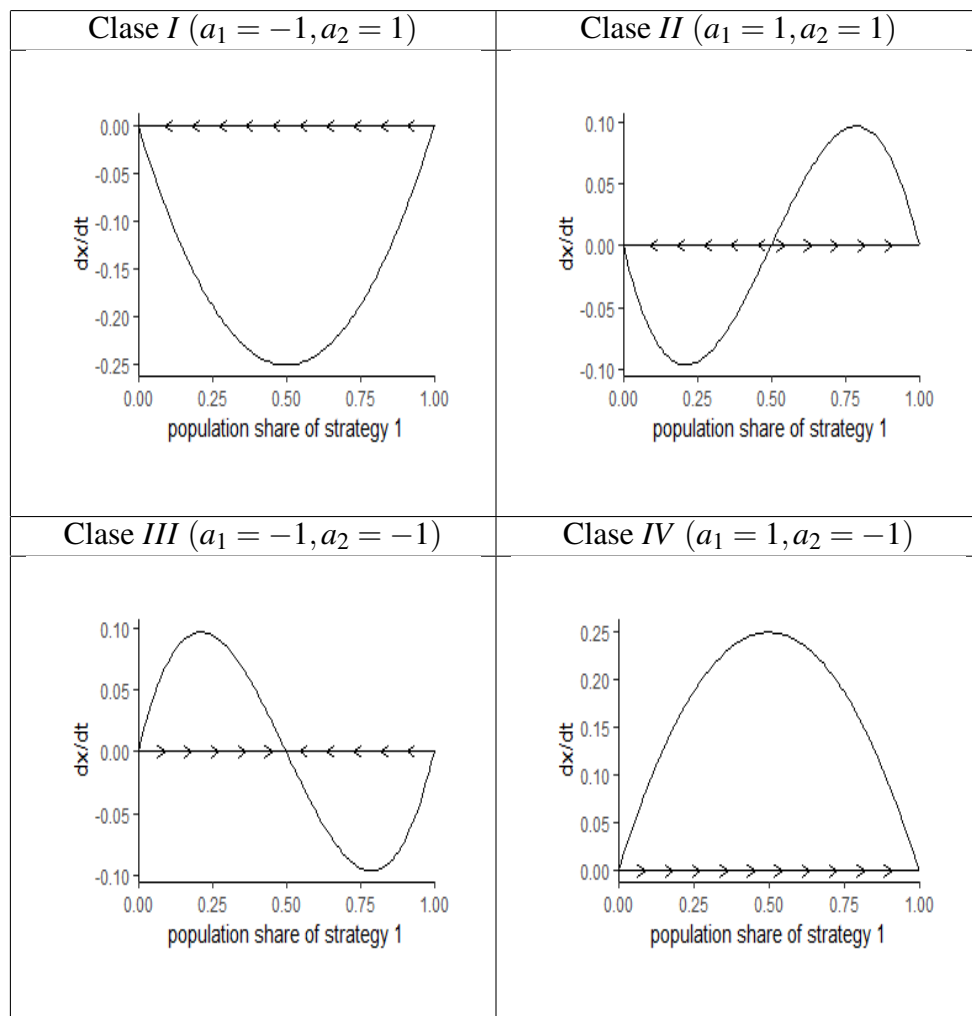


Figura 9.6: Estados de fase para las clases I a IV.

No se han hecho aplicaciones de teoría de juegos evolutiva al fenómeno de la concordancia. Sin embargo Deo (2015) aplicó la teoría para modelar el ciclo progresivo - imperfecto en el cambio semántico diacrónico

de los verbos en pasado. El ciclo tiene cuatro estadios consecutivos (estrategias), a saber: (a) se usa solo una forma lingüística X para denotar significado imperfecto (“comía”) y progresivo (“estaba comiendo”); (b) se introduce la forma Y para denotar progresivo (“mutante”) pero no es de uso obligatorio en contextos de progresivo (en inglés medio); (c) la forma Y se vuelve obligatoria para el significado progresivo y la X queda para el significado de imperfecto (en inglés actual); (d) X desaparece y la forma Y se usa para ambos significados, cerrando el ciclo. El modelo logra captar cuantitativamente el ciclo $a \rightarrow b \rightarrow c \rightarrow d$, pero falla en la predicción de que en la etapa (c) la gran mayoría de los hablantes utiliza la estrategia (c) y muy pocos la (a) o la (b), como en los datos de inglés actual; o que la mayoría de los hablantes utilizaba la estrategia (b), como muestran los datos en inglés medio. Yanovich (2017) mejoró dichas predicciones modificando el modelo para una población finita.

9.3.2. Aplicación y resultados.

Del apartado precedente ha quedado claro que la única clase en la que hay dos atractores es aquella a la que pertenece el juego de coordinación. Por lo tanto, se eligió una matriz simétrica 2×2 normalizada con $a_1 = 3$ y $a_2 = 1$ (ambos positivos, como requieren las condiciones de dicho juego). Con lo cual: $A = \begin{pmatrix} 3 & 0 \\ 0 & 1 \end{pmatrix}$; y $\hat{x} = \left(\frac{a_2}{a_1+a_2}, \frac{a_1}{a_1+a_2} \right) = \left(\frac{1}{4}, \frac{3}{4} \right) \in NE$. Se eligieron dichos valores de los coeficientes para que $\hat{x}_1 = \frac{1}{4}$ fuera similar al porcentaje real de error de los datos de los cuatro alumnos ($\approx 25\%$). La estrategia 1 fue “error” y la otra, “correcto”. Por la dinámica del replicador, las condiciones iniciales $x_1^0 > \frac{1}{4}$ serán atraídas hacia la estrategia “error” y con $x_1^0 < \frac{1}{4}$ habrá atracción hacia la estrategia “correcto”. En consecuencia, el sesgo del modelo depende enteramente de las condiciones iniciales, como en el modelo de *Lokta-Volterra*. Se aplica el siguiente mecanismo para determinarlo.

1. Calcular la distancia entre las variables de cada contexto de la tabla I y las mismas variables en la instancia en cuestión. Se cuentan cuántas de ellas *no* coinciden en sus valores. No se incluye la similitud consigo misma. O sea: $m_i = \#variables\ diferentes; i = 1, \dots, 4$.
2. Establecer $\theta = (\theta_1 < \frac{1}{4}, \theta_2 > \frac{1}{4})$ y asignar $p(\theta_j) = (e_{antes}, 1 - e_{antes})$, donde e_{antes} es la proporción de errores hasta la instancia anterior. Luego, para cada contexto i establecer la distribución de probabilidades a posteriori de θ_j utilizando la regla de Bayes: $p(\theta_j | x = m_i) = \frac{p(\theta_j)p(x=m_i|\theta_j)}{\sum_{\theta} p(\theta_j)p(x=m_i|\theta_j)}$; $j = 1, 2$ donde $x \sim Binom(n = c, p = \theta_j)$, con $c =$ número de variables en el contexto; luego elegir de acuerdo al máximo a posteriori (MAP): $\max(p(\hat{\theta}_1), p(\hat{\theta}_2))$ (se asigna θ_j de acuerdo a la probabilidad a posteriori más alta).
3. A continuación se cuentan las veces que se asignó θ_1 y θ_2 , en los contextos y se asigna a la variable E el θ_j con conteo máximo (si hay empate se elige al azar). Por último, el vector de condiciones iniciales resulta $x^0 = (x_1^0 = 1 - E, x_2^0 = E)$. Por lo tanto si $E < \frac{1}{4}$, gana el atractor “error”, y si $E > \frac{1}{4}$, gana el atractor “correcto”. El vector θ se parametrizó utilizando el error cuadrático de predicción: $RMSE = \sqrt{\frac{1}{4} \sum_{i=1}^4 (p_{sim,i} - p_{error,i})^2}$, donde p_{sim} es la proporción de instancias marcadas como “error” de la simulación (o sea, aquellas en donde gana el atractor “error”) y p_{real} la proporción de errores real, para cada aprendiente. Se elige el valor de θ con $RMSE$ mínimo. Se obtuvo $\theta = (\theta_1 = 0,24, \theta_2 = 0,85)$.

Se ejemplifica el procedimiento con la instancia “los estudio” [SONIA, sesión 4, línea 80]. La instancia posee las siguientes variables: MOD = “0”; Fabs.SC.f = “1”; FAM.LEX.f = “1”; MORF.f = “1”; ES = “0”; EST1 = “1”; “EST2” = “0”; EST5 = “0”. Además, el error hasta la instancia anterior es $e_{antes} = 0,5$. Con el contexto que baja la chance de error de género, la distancia es *uno* (de las cuatro variables que conforman el contexto, la instancia no coincide en sus valores con una [comparar con el Cuadro I]). Respecto de los contextos de “e-epentética”, “plural” y “mixto” las distancias son, respectivamente: *dos*, *uno*, *uno*. Por lo tanto: $m_1 = 1, m_2 = 2, m_3 = 1, m_4 = 1$. Además, $e_{antes} \approx 0,28$. Se considera m_1 . Los valores a elegir para \hat{x}_1^0 son $\theta = (\theta_1 = 0,24, \theta_2 = 0,85)$. Para ello se aplica la regla bayesiana con un prior

$p(\theta_j) = (e_{antes} = 0,29, 1 - e_{antes} = 0,71)$, de la siguiente forma. Como hay cuatro variables en el contexto fijemos $n = 4$ e indiquemos los valores de las verosimilitudes $p(x = i | \theta_j)$, con $i = \{0, \dots, 4\}$ y $j = \{1, 2\}$, como se observa en la siguiente tabla.

θ_j	$p(\theta_j)$	$p(x = i \theta_j)$	$x = 0$	$x = 1$	$x = 2$	$x = 3$	$x = 4$
$\theta_1 = 0,24$	0,29	$p(x = i \theta_1)$	0,3336	0,4214	0,1996	0,0420	0,0033
$\theta_2 = 0,85$	0,71	$p(x = i \theta_2)$	0,0005	0,0114	0,0975	0,3684	0,5220

La siguiente tabla muestra la probabilidad conjunta y la marginal:

θ_j	$p(\theta_j)p(x = i \theta_j)$	$x = 0$	$x = 1$	$x = 2$	$x = 3$	$x = 4$
$\theta_1 = 0,24$	$0,29p(x = i \theta_1)$	$0,29 \times 0,3336 = 0,0967$	0,1222	0,0578	0,0121	0,0009
$\theta_2 = 0,85$	$0,71p(x = i \theta_2)$	$0,71 \times 0,0005 = 0,0003$	0,0081	0,0692	0,2616	0,3706
Marginal	$\sum_{\theta} p(\theta_j)p(x = i \theta_j)$	0,0971	0,1303	0,1271	0,2738	0,3715

Por último, la posterior normalizada es:

$\frac{p(\theta_j)p(x=i \theta_j)}{\sum_{\theta} p(\theta_j)p(x=i \theta_j)}$	$x = 0$	$x = 1$	$x = 2$	$x = 3$	$x = 4$
$\frac{\frac{1}{3}p(x=i \theta_1)}{\sum_{\theta} p(\theta_j)p(x=i \theta_j)}$	$\frac{0,0967}{0,0971} = 0,9962$	0,9375	0,4553	0,0445	0,0025
$\frac{\frac{2}{3}p(x=i \theta_2)}{\sum_{\theta} p(\theta_j)p(x=i \theta_j)}$	$\frac{0,0003}{0,0971} = 0,0037$	0,0624	0,5446	0,9554	0,9974

Como para $m_1, x = 1$ y $n = 4$, la posterior es $p(\theta_j | x = 1) = (0,93, 0,06)$. Con lo cual el máximo a posteriori es 0,93. Por lo tanto se elige $\theta_1 = 0,24$. Si esto se repite con el resto de los valores de m_i , se obtiene que se elige siempre θ_1 , con lo cual el conteo es: (4,0), cuyo máximo corresponde a $\theta_1 = 0,24$. Si el conteo hubiera dado empatado (2,2) se hubiera elegido θ al azar. En consecuencia $E = 0,24$ y $x^0 = (0,76, 0,24)$, con lo cual $x_1^0 = 1 - E = 0,76 > 0,25$ y el flujo es atraído hacia el atractor “error”. La Figura que sigue muestra el diagrama de fase y el flujo de activación.

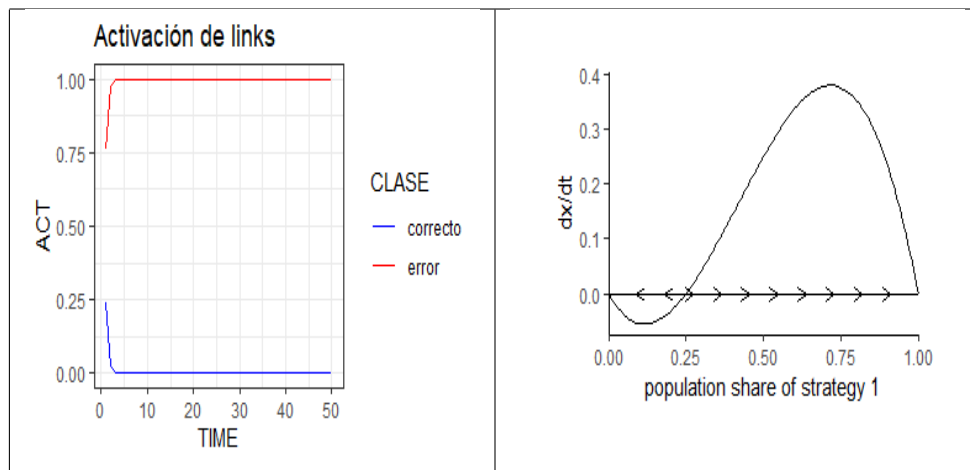


Figura 9.7: Flujo de activación de la instancia “los estudio” [SONIA, sesión 4, línea 80] (izquierda) y diagrama de fase del juego de coordinación ($a_1 = 3, a_2 = 1$).

El Cuadro que sigue muestra los resultados: (i) las proporciones globales de “error” entre la simulación y los datos; (ii) los p-valores de un test asintótico de diferencia entre proporciones entre los errores simulados y los errores de los datos; (iii) los valores de la distancia de Minkowski $D = \sqrt[p]{(x_i - y_i)^p}$; ($p = 3$), entre los errores simulados y reales por sesión (a menor distancia, mejor). Por otro lado, las Figuras muestran un

gráfico de barras de proporción de errores y de correctos; así como una gráfica que compara los datos reales y los de la simulación respecto de la proporción de errores en la serie temporal de sesiones.

Se observa que las diferencias entre las proporciones de error no resultan significativas. En cuanto al patrón por sesiones SONIA y NATI se aproximan de modo similar a las series de tiempo de los datos. MIRKA resulta la de peor desempeño y JAKO es el que mejor se acerca al patrón de los datos.

	error.sim	error.real	p.value	dist
SONIA	0.192	0.215	0.551	10.746
NATI	0.377	0.336	0.262	10.477
JAKO	0.155	0.164	0.788	9.899
MIRKA	0.275	0.291	0.550	13.024

Cuadro 9.6: Resultados de la simulación: (i) proporciones entre errores reales y simulados; (ii) p-valores de un test para diferencia entre dichas proporciones; (iii) distancia de Minkowski entre los errores simulados y reales por sesiones.

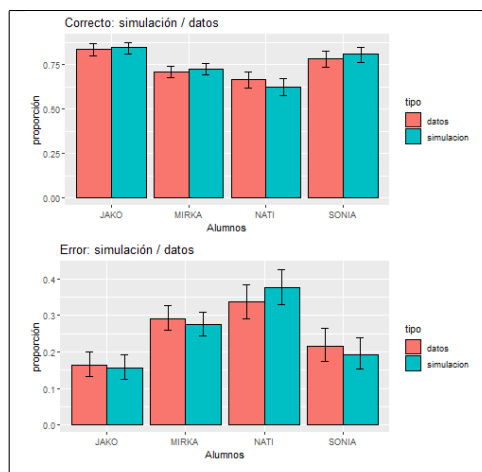


Figura 9.8: Barras de proporción de errores simulados y reales; y de correctos simulados y reales.

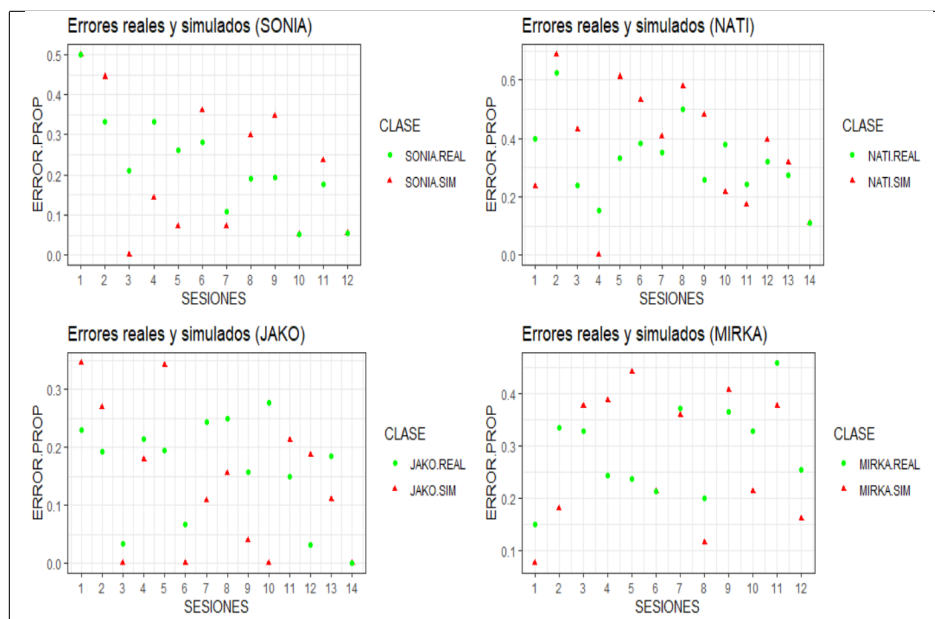


Figura 9.9: Series de tiempo de proporción de errores simulados y reales, por sesiones.

9.4. Ascenso de gradiente de armonía.

Gradient Symbolic Computation (GSC) [Smolensky et al., 2014; Cho et al., 2017, 2018] postula un sistema dinámico estocástico, para tiempo y estados continuos. El modelo construye gradualmente estructuras simbólicas discretas (atractores) a partir de un espacio de estados continuo. Para ver cómo funciona el modelo, considérese a modo de ejemplo el lenguaje formal L . Dicho lenguaje está caracterizado por las siguientes reglas de estructura de frase: $S \rightarrow S[1] \mid S[2] \mid S[3] \mid S[4]$, con: $S[1] \rightarrow AB$, $S[2] \rightarrow AC$, $S[3] \rightarrow DB$, $S[4] \rightarrow DC$. El parseado de cada secuencia es: $T_1 = [{}_s [{}_{s[1]} [AB]]]$, $T_2 = [{}_s [{}_{s[2]} [AC]]]$, $T_3 = [{}_s [{}_{s[3]} [DB]]]$, $T_4 = [{}_s [{}_{s[4]} [DC]]]$. Cada parseado se representa en términos de “roles” [*span roles*] y de constituyentes que llenan esos roles [*fillers*]. Los *span roles* son etiquetas que se asignan a cada una de las cuatro posiciones: S , $S[i]$, A , B . Aquí las etiquetas son $(0, 1)$, $(1, 2)$, $(0, 1, 2)$, $(0, 2)$. Los *span roles* de la forma (i, j) son las etiquetas de los constituyentes que abarcan de la posición i a la posición j en la secuencia que se parsea. Por otro lado, un *span role* de la forma (i, k, j) abarca de la posición i a la j y tiene dos “hijos”, uno que abarca de la posición i a la k y otro de k a j . Por ejemplo, en T_1 el “filler” A tiene el rol $(0, 1)$ formando la unión [*binding*] $A/(0, 1)$. Por tanto, para T_1 se tendrá el siguiente conjunto de uniones: $T_1 = [{}_s [{}_{s[1]} [AB]]] \equiv \{B/(1, 2), S/(0, 2), A/(0, 1), S[1]/(0, 1, 2)\}$, para $i = 0$, $j = 2$, $k = 1$. Esquemáticamente:

$S/(0, 2)$ $w = 2; b = -1$	
$S[1]/(0, 1, 2)$ $w = 2; b = -3$	
$A/(0, 1)$ $w = 2; b = -1$	$B/(1, 2)$ $w = 2; b = -1$

Cuadro 9.7: Representación de T_1 en términos de *fillers* y *span roles*.

Cada unión de *filler* f_n con *span roll* r_n del constituyente n –ésimo se denotará f_n/r_n ($n = 1, \dots, k$; para un total de k constituyentes). Se representa en un espacio continuo mediante el *producto exterior* entre un vector de activación de n componentes $f = (f_1, \dots, f_n)$ (que representa el *filler*) y otro vector de activación de n componentes $r = (r_1, \dots, r_n)$ (representando el *span rol*). El resultado es una matriz $n \times n$.

$$f/r \equiv f \otimes r \equiv fr^T = \begin{bmatrix} f_1 r_1 & f_1 r_2 & \dots & f_1 r_n \\ f_2 r_1 & f_2 r_2 & \dots & f_2 r_n \\ \vdots & \vdots & \ddots & \vdots \\ f_n r_1 & f_n r_2 & \dots & f_n r_n \end{bmatrix}$$

La codificación de un *conjunto de uniones* para un determinado parsing es la suma de los productos externos de cada unión:

$$\{f_1/r_1, f_2/r_2, \dots, f_k/r_k\} \equiv \sum_{n=1}^k f_n \otimes r_n = f_1 \otimes r_1 + f_2 \otimes r_2 + \dots + f_k \otimes r_k$$

El modelo de *GSC* usa una *Harmonic Grammar* [*HG*] (Smolensky, 2006) para codificar restricciones que imponen “premios” (pesos w de valor positivo) o “castigos” (b de valor negativo) sobre la buena formación (o “armonía”) de las estructuras simbólicas. *HG* asigna un peso positivo w a cada par de uniones “padre-hijo” ($f_i/r_i, r_j/r_j$) en el árbol de parsing, por ejemplo ($S[1]/(0, 1, 2), A/(0, 1)$). En el Cuadro, $w = 2$ para cada par de uniones. Además, se asigna una penalidad $-b$ a cada *filler*, donde b es el numero de conexiones que el *filler* debe tener para que la estructura sea gramatical (por ejemplo, A no podría tener dos conexiones en el árbol). Si todas las conexiones resultan ser legales, entonces, las suma de los valores de b cancelan a los de w y la armonía total del parsing es cero, el máximo valor posible. La armonía de un conjunto de uniones f/r es la suma de sus restricciones, cuya expresión es:

$$H_G(x, e_n) = \frac{1}{2} x^T W_n x + b_n^T x + e_n^T x$$

donde: x es el vector de activación, W es la matriz de pesos positivos w , b es el vector de penalidades negativas y e es el vector de activación que corresponde al *input* externo. Ahora bien, se necesitan otros términos para penalizar representaciones que no son aproximadamente discretas (que es a dónde debe llegar el sistema). A saber:

- Baseline: H_B requiere que el vector de activación x no se aleje del valor de *baseline* z_n :

$$H_B = -\frac{1}{2}\beta \|x - z_n\|^2 = -\frac{1}{2}\beta (x - z_n)^T (x - z_n)$$

- Cuantización: H_Q asegura que cada *span rol* se una a un solo *filler*, y la activación sea igual a 1. Esto se logra mediante dos restricciones, a saber: (i) H_{Q_0} : la activación de cada unión [*binding*] sea *cero* o *uno*; (ii) H_{Q_1} asegura que la suma de las activaciones de las uniones al cuadrado, en cada grupo de uniones que compiten entre sí; deba ser igual a 1. H_{Q_1} introduce competición entre los *fillers* para unirse a cada rol (o sea que se penalizan las representaciones con múltiples *fillers* para un determinado *span rol*). Ambas restricciones se expresan como:

$$H_{Q_0}(x) = -\sum_r \sum_f (x_{f,r})^2 (1 - x_{f,r})^2$$

$$H_{Q_1}(x) = -\sum_r \left(1 - \sum_f x_{f,r}^2\right)^2$$

y uniendo ambos: $H_Q(x) = \frac{1}{2}[H_{Q_0}(x) + H_{Q_1}(x)]$. La armonía total del estado de activación x , dado el *input* externo e_n y la “presión hacia una representación discreta” q resulta ser:

$$H(x, e_n, q) = H_G(x, e_n) + H_B(x) + qH_Q(x)$$

donde $q \geq 0$ es un parámetro que pesa la restricción de «cuantización». Aquí se asume que q aumenta con el tiempo de manera lineal. Valores grandes de q fuerzan el modelo a elegir entre *fillers* competidores.

El modelo *GSC* es un proceso estocástico dinámico que optimiza las restricciones siguiendo el gradiente de armonía; o sea aumentando dicha armonía con el tiempo. Se expresa como:

$$\frac{dx}{dt} = \nabla_x H(x, e_n, q) dt + \sqrt{2T} dW$$

donde W es el proceso de Wiener, T es el nivel de ruido (que puede ser constante o decrecer con el tiempo de modo exponencial) y ∇_x es el gradiente de de la armonía total evaluada en el vector de activación x . El primer término es el determinístico (*drift*) y el segundo es la parte estocástica.

9.4.1. Aplicación y resultados.

Considérese el modelo para una dimensión donde la variable x constituye un continuo de activación entre $[-1, 1]$. En los extremos x representa dos estados “puros” o atractores discretos: $x = -1$ es el símbolo “correcto” y $x = 1$ representa el símbolo “error”. En una dimensión el vector de activación x es un número. Los valores de $x \in (-1, 1)$ son estados “mezcla” entre ambos símbolos. Se desea que el sistema comience en un estado arbitrario y la activación vaya hacia alguno de los estados “puros”. Por otro lado, el parámetro $e \in [-1, 1]$ es el valor que sesga dicha atracción hacia algún valor del continuo. En este caso solo hay un *filler* por *span roll*, con lo cual el término de competición H_{Q_1} se ignora. Los componentes de armonía quedan:

$$H_G = \frac{1}{2}Wx^2 + bx + ex$$

$$H_B = -\frac{1}{2}\beta (x - z)^2$$

$$H_{Q_0} = -[x(1-x)]^2 = -[(x-1)(x+1)]^2 = -(x-1)^2(x+1)^2$$

Como no hay estructura jerárquica, $W = b = 0$, y se establece: $z = 0, \beta = 4$. En una sola dimensión, el gradiente para cada componente es simplemente su derivada respecto de x :

$$\begin{aligned} \nabla H_G &= Wx + b + e \\ \nabla H_B &= -\beta(x - z) \\ \nabla H_{Q_0} &= -4x(x - 1)(x + 1) \end{aligned}$$

Con lo cual⁸: $\nabla_x H(x, e_n, q) = \nabla H_G + \nabla H_B + q\nabla H_{Q_0}$ en: $\frac{dx}{dt} = \nabla_x H(x, e_n, q) dt + \sqrt{2T}dW$.

Ahora bien supóngase $e = 1$, es decir que el sistema está sesgado hacia el atractor “error” (si $e = 0$ no hay sesgo). La Figura que sigue ilustra el comportamiento del sistema a medida que cambia el parámetro dinámico $q \in [0, 5]$. Si $q = 0$ la “superficie” de armonía tiene la forma de una parábola con máximo global en $x = 0,25$. Por lo tanto, dicho máximo es el único atractor y representa un estado “mezcla”, intermedio entre “error” y “correcto”, con un poco de sesgo hacia el “error”. Si se aumenta el valor de q , se observará que aparecen dos máximos locales. Ha sucedido una bifurcación y se ha pasado de un atractor a dos atractores. A medida que q sigue aumentando, el “valle” entre ambos máximos locales se hace más “profundo”. El sistema “escalará” el gradiente de armonía y eventualmente alcanzará $x = 1$, el atractor del “error”. Si la cantidad de ruido T que se agrega a la parte determinística de la ecuación es pequeño, será difícil que haya valores que estén atraídos hacia el atractor “correcto” y superar “hacia atrás” el valle entre los máximos locales. En cambio, si T es grande, el sistema podría terminar en $x = -1$, “correcto”. En el apéndice se ofrece un análisis detallado.

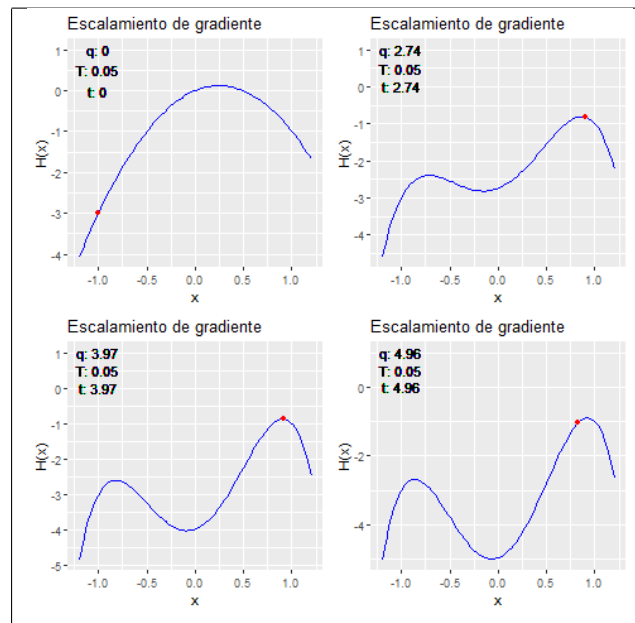


Figura 9.10: Ascenso de gradiente de armonía según $q = 0$ (arriba izquierda), $q = 2,74$ (arriba derecha), $q = 3,97$ (abajo izquierda), $q = 4,96$ (abajo derecha).

El parámetro de sesgo e se calcula de la siguiente manera.

1. Se usa la medida de Jaccard para calcular la distancia entre las variables de cada contexto de la tabla I y las mismas variables en la instancia en cuestión. No se incluye la similaridad consigo misma. El índice de Jaccard mide la similitud entre dos conjuntos A y B . Es la cardinalidad de la intersección entre ambos conjuntos, dividido por la cardinalidad de su unión⁹: $J(A, B) = \frac{|A \cap B|}{|A \cup B|} = \frac{|A \cap B|}{|A| + |B| - |A \cap B|}$. La distancia de Jaccard es $d_{jaccard} = 1 - J(A, B) \in [0, 1]$, siendo cero si ambas variables coinciden en

⁸Se implementó la ecuación para R a partir de un script de *Phyton* en el sitio: <https://cocalc.com/projects/18c77389-a5c3-49de-946a-7593b53d3fb2/>

⁹Dados dos vectores A y B con la misma cantidad de elementos; el índice se calcula como: $J(A, B) = \frac{\text{sum}(A \times B)}{(\text{sum}(A^2) + \text{sum}(B^2) - \text{sum}(A \times B))}$.

todos los valores de las variables que comparten. Por lo tanto: $m_i = 1 - J(\text{instancia}, \text{contexto})$; $i = 1, \dots, 4$.

2. Se calcula un *kernel* gaussiano como: $K = \exp \left\{ -\frac{1}{2\text{error}_{\text{antes}}} [\text{mediana}(m_1 + \dots + m_4)]^2 \right\}$; donde $\text{error}_{\text{antes}}$ es la proporción de errores hasta la instancia anterior.
3. Se calcula el percentil correspondiente a K , bajo una distribución logística truncada entre -1 y 1 , centrada en 0 y con escala 1 (o sea, con $P(X < -1) = 0$, $P(X < 0) = 0,5$ y $P(X < 1) = 1$). Este es el valor de x al cual se sesga el modelo.
4. Dados $t, q \in [0, 5]$, se corre el modelo hasta $t = q = 5$ (partiendo de $x(0) = -1$) y se declara “error” si el valor final de $x(t) > s$. En caso contrario se declara “correcto”. s es una barrera que declara cuán cerca es necesario estar del atractor “error” $x = 1$. Se lo parametrizó mediante el error cuadrático de predicción: $RMSE = \sqrt{\frac{1}{4} \sum_{i=1}^4 (p_{\text{sim},i} - p_{\text{error},i})^2}$, donde p_{sim} es la proporción de instancias marcadas como “error” de la simulación (o sea, aquellas en donde gana el atractor “error”) y p_{real} la proporción de errores real, para cada aprendiente. Se elige el valor de θ con $RMSE$ mínimo. Se obtuvo $s = 0,9$.

Se ilustra el procedimiento con la instancia errónea “muchas lugares” [SONIA, sesión 4, línea 56]. La instancia posee las siguientes variables: MOD = “2”; Fabs.SC.f = “0”; FAM.LEX.f = “1”; MORF.f = “2”; ES = “1”; EST1 = “0”; “EST2” = “0”; EST5 = “0”. Además el error hasta la instancia anterior es $\text{error}_{\text{antes}} = 0,5$. Entonces el conjunto de valores de la instancia es $A = \{0, 1, 0, 0\}$ y el del último contexto es $B = \{1, 1, 1, 1\}$; y $|A \cap B| = 1$, $|A \cup B| = 4$ El índice de Jaccard para el último contexto es $J(A, B) = \frac{1}{4} = 0,25$; con lo cual $m_4 = 1 - J(A, B) = 0,75$. Para el resto de los contextos son: , $m_1 = 0,30$, $m_2 = 0,57$, $m_3 = 0,15$. Luego se calcula el kernel gaussiano como: $K = \frac{1}{2(0,214)} e^{-[\text{mediana}(0,3,0,57,0,15,0,75)]^2} \approx 0,64$. Y su percentil bajo la distribución logística truncada es $q_{\text{logis}}(0,64) = 0,26$; y por tanto: $e = 0,26$. Como se observa en la Figura que sigue, en $q = 0$, hay un solo atractor con máximo aproximadamente en 0 . Luego se comienzan a formar dos atractores. Eventualmente el sistema termina en $x(t = 5) = 0,917$. Como $x(t = 5) > 0,9$, se declara “error”.

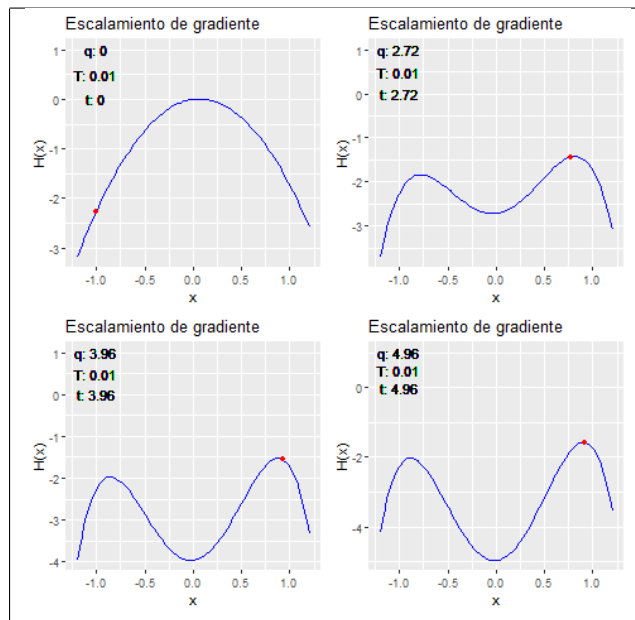


Figura 9.11: Ascenso de gradiente de armonía para la instancia “muchas lugares” [SONIA, sesión 4, línea 56]; según $q = 0$ (arriba izquierda), $q = 2,72$ (arriba derecha), $q = 3,96$ (abajo izquierda), $q = 4,96$ (abajo derecha).

Ya que se trata de un modelo estocástico, se corrieron 50 simulaciones completas. El Cuadro que sigue muestra: (i) las proporciones medias globales de la simulación y las proporciones de los datos reales; (ii) el

error típico [se] de las medias de errores simulados; (iii) a partir de 50 tests asintóticos de diferencia entre proporciones entre los errores simulados y los errores de los datos, la frecuencia relativa de p-valores resultantes menores de 0.05; (iv) los valores medios de la distancia de Minkowski $D = \sqrt[p]{(x_i - y_i)^p}; (p = 3)$, entre los errores simulados y reales por sesión (a menor distancia, mejor). Por otro lado, las Figuras muestran un gráfico de barras de proporción de errores y correctos; así como una gráfica que compara los datos reales y los de la simulación respecto de la proporción de errores en la serie temporal de sesiones (con barras de error: *media (errores simulados) ± se*).

Se observa que las diferencias entre las proporciones de error no resultan significativas. En el patrón global, SONIA y MIRKA son las que más se acercan a los datos reales. En cuanto al patrón por sesiones, SONIA es la de mejor desempeño en la aproximación las series de tiempo de los datos, el resto registra un desempeño similar.

	error.sim.mean	error.sim.se	error.real	p.value.freq	dist.mean
1	0.1999	0.0004	0.2147	0.0000	6.4544
2	0.3750	0.0004	0.3359	0.0600	11.5035
3	0.1937	0.0003	0.1638	0.0600	12.1196
4	0.2839	0.0003	0.2909	0.0000	12.5854

Cuadro 9.8: Resultados de la simulación: (i) proporciones de errores reales y proporciones medias de simulados; (ii) test para diferencia entre dichas proporciones: frecuencias de p.valores < 0.05; (iii) distancia media de Minkowski entre los errores simulados y reales por sesiones.

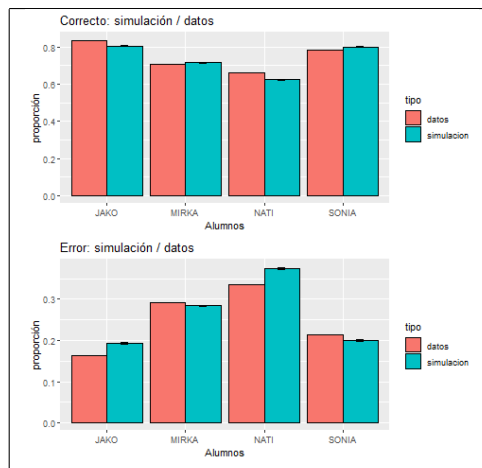


Figura 9.12: Barras de proporción de errores simulados y reales; y de correctos simulados y reales.

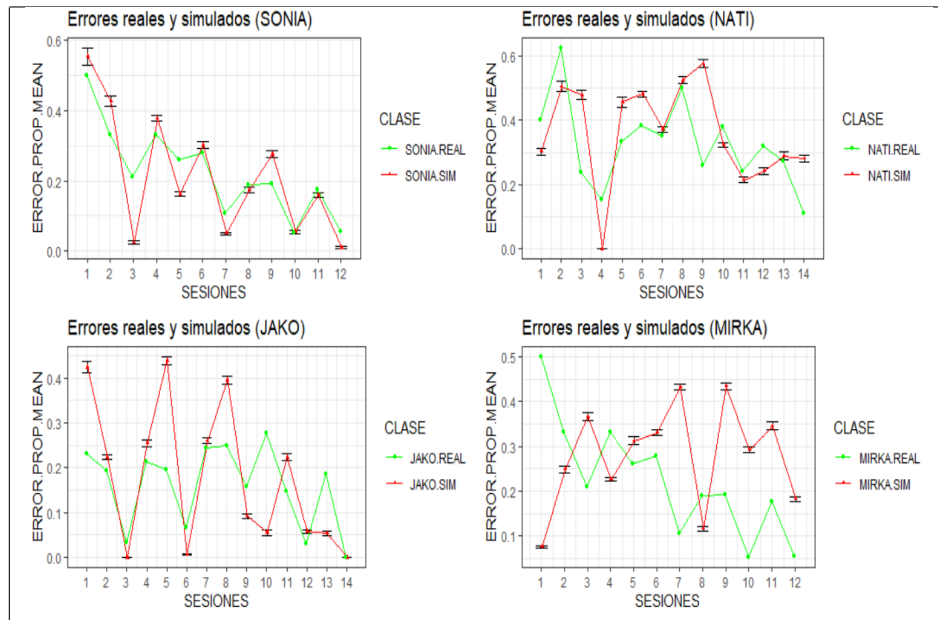


Figura 9.13: Series de tiempo de proporción de errores simulados y reales, por sesiones.

9.5. Conclusiones.

Si cada concordancia fuera un individuo que toma decisiones de acuerdo a la dinámica de los modelos especificados; entonces, a partir del conjunto de dichas decisiones *emerge* un patrón global de error y otro por serie de tiempo. Es decir que los patrones de aprendizaje se auto-organizan y emergen de la dinámica ejercida por los individuos. En general los modelos aproximan bien los datos globales de error reales. En cuanto a la aproximación a las series de tiempo reales, SONIA alcanza el mejor desempeño con el modelo *GSC*, NATI con el de *Lokta-Volterra* y JAKO con el del juego de coordinación. MIRKA resultó la de peor desempeño en todos los modelos. Todos los modelos dinámicos analizados consideran al error como atractor que se alcanza (o no) en el último paso temporal. Es decir que el error es una representación continua y solamente al final, en el atractor, se vuelve una representación discreta. “Error” o “Correcto” son propiedades emergentes de los sistemas dinámicos descriptos.

9.6. Apéndice de capítulo.

9.6.1. Modelo de Lokta-Volterra: análisis de estabilidad.

Se sigue en su mayoría a Murray (2002). Las ecuaciones diferenciales del modelo de competencia entre especies de Lokta-Volterra rezan:

$$\begin{aligned}\frac{dN_1}{dt} &= r_1 N_1 \left(1 - \frac{N_1}{K_1} - \alpha_{12} \frac{N_2}{K_1} \right) \\ \frac{dN_2}{dt} &= r_2 N_2 \left(1 - \frac{N_2}{K_2} - \alpha_{21} \frac{N_1}{K_2} \right)\end{aligned}$$

En primer lugar se lleva el modelo al formato adimensional, lo cual produce un modelo equivalente con menos parámetros. Sea: $u_1 = \frac{N_1}{K_1} \rightarrow N_1 = K_1 u_1$; $u_2 = \frac{N_2}{K_2} \rightarrow N_2 = K_2 u_2$; $\tau = r_1 t \rightarrow t = \frac{\tau}{r_1}$; $a_{12} = \alpha_{12} \frac{K_2}{K_1}$; $a_{21} = \alpha_{21} \frac{K_1}{K_2}$; $\rho = \frac{r_2}{r_1}$. Con lo cual se obtiene:

$$\begin{aligned}\frac{d(K_1 u_1)}{d(\tau/r_1)} &= K_1 r_1 \frac{d(u_1)}{d(\tau)} = r_1 N_1 \left(1 - \frac{N_1}{K_1} - \alpha_{12} \frac{K_2}{K_1} u_2 \right) = \frac{r_1 N_1}{r_1 K_1} \left(1 - \frac{N_1}{K_1} - \alpha_{12} u_2 \right) \\ \frac{d(K_2 u_2)}{d(\tau/r_1)} &= K_2 r_1 \frac{d(u_2)}{d(\tau)} = r_2 N_2 \left(1 - \frac{N_2}{K_2} - \alpha_{21} \frac{K_1}{K_2} u_1 \right) = \frac{r_2 N_2}{r_1 K_2} \left(1 - \frac{N_2}{K_2} - \alpha_{21} u_1 \right)\end{aligned}$$

Y se llega al modelo reducido, reemplazando por u_1, u_2, ρ en la parte derecha.

$$\begin{aligned}f_1(u_1, u_2) &= \frac{du_1}{d\tau} = u_1 (1 - u_1 - a_{12} u_2) \\ f_2(u_1, u_2) &= \frac{du_2}{d\tau} = \rho u_2 (1 - u_2 - a_{21} u_1)\end{aligned}$$

Los puntos fijos son las soluciones simultáneas de: $f_1(u_1, u_2) = u_1 = 0$ y $f_2(u_1, u_2) = u_2 = 0$. Esto sucede si: $P_1 = (u_1^* = 0, u_2^* = 0)$, $P_2 = (u_1^* = 1, u_2^* = 0)$, $P_3 = (u_1^* = 0, u_2^* = 1)$. Para P_4 debe resolverse el sistema: $u_1 (1 - u_1 - a_{12} u_2) = 0 \Leftrightarrow u_1 = 1 - a_{12} u_2$; $u_2 (1 - u_2 - a_{21} u_1) = 0 \Leftrightarrow u_2 = 1 - a_{21} u_1$. Y sustituyendo para calcular u_1^* : $u_1 = 1 - a_{12} u_2 = 1 - a_{12} (1 - a_{21} u_1) = 1 - a_{12} + a_{21} a_{12} u_1 \Rightarrow u_1 - a_{21} a_{12} u_1 = 1 - a_{12} \Rightarrow (1 - a_{12} a_{21}) u_1 = 1 - a_{12} = \frac{1 - a_{12}}{1 - a_{12} a_{21}} = u_1^*$.

Y análogamente para u_2^* ; obteniéndose:

$$P_4 = \left(u_1^* = \frac{1 - a_{12}}{1 - a_{12} a_{21}}, u_2^* = \frac{1 - a_{21}}{1 - a_{12} a_{21}} \right); a_{12} a_{21} \neq 1 \text{ (con lo cual se está en el primer cuadrante: } u_1^* \geq 0; u_2^* \geq 0).$$

Con el objetivo de descubrir el tipo de punto fijo, se hace necesario calcular el Jacobiano y evaluarlo en el punto fijo en cuestión $J(u_1^*, u_2^*)$:

$$J = \begin{pmatrix} J_{11} = \frac{\partial f_1}{\partial u_1} & J_{12} = \frac{\partial f_1}{\partial u_2} \\ J_{21} = \frac{\partial f_2}{\partial u_1} & J_{22} = \frac{\partial f_2}{\partial u_2} \end{pmatrix}_{u_1^*, u_2^*} = \begin{pmatrix} 1 - 2u_1 - a_{12} u_2 & -a_{12} u_1 \\ -\rho a_{21} u_2 & \rho (1 - 2u_2 - a_{21} u_1) \end{pmatrix}_{u_1^*, u_2^*}$$

Luego hay que calcular la traza y el determinante de J : $Tr(J) = J_{11} + J_{22}$, $det(J) = J_{11} J_{22} - J_{21} J_{12}$. Los autovalores de J serán entonces: $\lambda_{1,2} = \frac{Tr(J) \pm \sqrt{[Tr(J)]^2 - 4det(J)}}{2}$. Por último los puntos fijos se clasifican siguiendo el Cuadro (se incluyen solamente los tipos pertinentes para el análisis):

	$Tr(J) < 0$	$Tr(J) > 0$
$[Tr(J)]^2 > 4det(J)$	$\lambda_1 < 0, \lambda_2 < 0$: Atractor (estable)	$\lambda_1 > 0, \lambda_2 > 0$: Repulsor (inestable)
$det(J) < 0$	$\lambda_1 < 0, \lambda_2 > 0$: Saddle (inestable)	$\lambda_1 < 0, \lambda_2 > 0$: Saddle (inestable)

Cuadro 9.9: Clasificación de Puntos Fijos.

Siguiendo lo expuesto, se aplica la fórmula para los autovalores y se clasifica los puntos fijos según el Cuadro anterior.

$$P_1 = (u_1^* = 0, u_2^* = 0): J = \begin{pmatrix} 1 & 0 \\ 0 & \rho \end{pmatrix}; det(J) = \rho; Tr(J) = (1 + \rho); \lambda_{1,2} = \frac{1}{2} \left\{ (1 + \rho) \pm \sqrt{(1 + \rho)^2 - 4\rho} \right\} \Rightarrow \lambda_1 = \rho; \lambda_2 = 1. \text{ Clasificación: como } \lambda_1 > 0, \lambda_2 > 0, \text{ es Repulsor.}$$

$$P_2 = (u_1^* = 1, u_2^* = 0): \quad J = \begin{pmatrix} -1 & -a_{12} \\ 0 & \rho(1-a_{21}) \end{pmatrix}; \det(J) = -\rho(1-a_{21}); \operatorname{Tr}(J) = \rho(1-a_{21}) - 1;$$

$\lambda_{1,2} = \frac{1}{2} \left\{ \rho(1-a_{21}) - 1 \pm \sqrt{(\rho(1-a_{21}) - 1)^2 - [4(-\rho)(1-a_{21})]} \right\} \Rightarrow \lambda_1 = -1; \lambda_2 = \rho(1-a_{21})$. Clasificación: si $a_{21} > 1 \Rightarrow \lambda_1 < 0, \lambda_2 < 0$, por lo tanto es Atractor. En cambio, si $a_{21} < 1 \Rightarrow \lambda_1 < 0, \lambda_2 > 0$, por lo tanto es Saddle (notar que $\det(J) < 0$).

$$P_3 = (u_1^* = 0, u_2^* = 1): \quad J = \begin{pmatrix} 1-a_{12} & 0 \\ -\rho a_{21} & -\rho \end{pmatrix}; \det(J) = -\rho(1-a_{12}); \operatorname{Tr}(J) = (1-a_{12}) - \rho;$$

$\lambda_{1,2} = \frac{1}{2} \left\{ (1-a_{12}) - \rho \pm \sqrt{((1-a_{12}) - \rho)^2 - [4(-\rho)(1-a_{12})]} \right\} \Rightarrow \lambda_2 = 1-a_{12}; \lambda_1 = -\rho$. Clasificación: Si $a_{12} > 1 \Rightarrow \lambda_1 < 0, \lambda_2 < 0$, por lo tanto es Atractor. En cambio, si $a_{21} < 1 \Rightarrow \lambda_1 < 0, \lambda_2 > 0$, por lo tanto es Saddle (notar que $\det(J) < 0$).

$$P_4 = \left(u_1^* = \frac{1-a_{12}}{1-a_{12}a_{21}}, u_2^* = \frac{1-a_{21}}{1-a_{12}a_{21}} \right): \quad J = (1-a_{12}a_{21})^{-1} \begin{pmatrix} a_{12}-1 & a_{12}(a_{12}-1) \\ \rho a_{21}(a_{21}-1) & \rho(a_{21}-1) \end{pmatrix};$$

Para el determinante se aplica la regla: $\det(cA) = c^n \det(A); A_{n \times n}$. Con lo cual:

$$\begin{aligned} \det(J) &= (1-a_{12}a_{21})^{-2} [(a_{12}-1)\rho(a_{21}-1) - \rho a_{21}(a_{21}-1)a_{12}(a_{12}-1)] \\ &= (1-a_{12}a_{21})^{-2} \rho(a_{12}-1)(a_{21}-1)(1-a_{12}a_{21}) \end{aligned}$$

Para la traza se usó la regla: $\operatorname{Tr}(cA) = c \operatorname{Tr}(A); A_{n \times n}$. Por lo tanto:

$$\operatorname{Tr}(J) = (1-a_{12}a_{21})^{-1} [(a_{12}-1) + \rho(a_{21}-1)].$$

Y los autovalores son:

$$\begin{aligned} \lambda_{1,2} &= \frac{1}{2} \left[(1-a_{12}a_{21})^{-1} [(a_{12}-1) + \rho(a_{21}-1)] \pm \left\{ [(1-a_{12}a_{21})^{-1} [(a_{12}-1) + \rho(a_{21}-1)]]^2 - 4(1-a_{12}a_{21})^{-2} \rho(a_{12}-1)(a_{21}-1)(1-a_{12}a_{21}) \right\}^{\frac{1}{2}} \right] \\ &= \frac{1}{2} \left[(1-a_{12}a_{21})^{-1} [(a_{12}-1) + \rho(a_{21}-1)] \pm (1-a_{12}a_{21})^{-1} \left\{ [(a_{12}-1) + \rho(a_{21}-1)]^2 - 4\rho(a_{12}-1)(a_{21}-1)(1-a_{12}a_{21}) \right\}^{\frac{1}{2}} \right] \end{aligned}$$

con lo cual se tiene:

$$\lambda_{1,2} = [2(1-a_{12}a_{21})]^{-1} \left[(a_{12}-1) + \rho(a_{21}-1) \pm \left\{ [(a_{12}-1) + \rho(a_{21}-1)]^2 - 4\rho(a_{12}-1)(a_{21}-1)(1-a_{12}a_{21}) \right\}^{\frac{1}{2}} \right]$$

Clasificación: si $a_{12} < 1; a_{21} < 1 \Rightarrow \lambda_1 < 0, \lambda_2 < 0$, por lo tanto es un Atractor. Por otra parte si $a_{12} > 1; a_{21} > 1 \Rightarrow \lambda_1 < 0, \lambda_2 > 0$ y el punto fijo es un Saddle.

9.6.2. Dinámica del replicador: análisis de estabilidad.

En el caso de los juegos simétricos de 2×2 con matriz normalizada: $A = \begin{pmatrix} a_1 = a-b & 0 \\ 0 & a_2 = d-b \end{pmatrix}$, y cuando $a_1 a_2 \neq 0$, la dinámica del replicador está gobernada por las ecuaciones:

$$\dot{x}_1 = x_1 [u(e^1, x) - u(x, x)] = x_1 (1-x_1) [a_1 x_1 - a_2 (1-x_1)]$$

$$\dot{x}_2 = -\dot{x}_1$$

Como en realidad hay una sola dimensión, los puntos fijos son las soluciones de: $f_1(x_1) = \dot{x}_1 = 0$. Estos son: $P_1 = (x_1^* = 1)$, $P_2 = (x_1^* = 0)$, $P_3 = \left(x_1^* = \frac{a_2}{a_1+a_2} \right)$. Con el objetivo de descubrir el tipo de punto fijo, se hace necesario calcular la derivada con respecto a x_1 y evaluarla en el punto fijo en cuestión: $\lambda = \left[\frac{df_1(x)}{dx} \right]_{x^*}$. Si $\lambda < 0$, el punto fijo es atractor, si $\lambda > 0$, es repulsor. Si $\lambda = 0$ puede ser atractor, repulsor o saddle; y es necesario examinar el signo la segunda derivada. Si es un saddle, la segunda derivada tiene el mismo signo en un punto $x = x^* \pm \varepsilon$, a ambos lados de x^* . En un saddle, el punto fijo atrae en una dirección y repele en la otra (o atrae a la izquierda y repele a la derecha o viceversa). Para el caso de un atractor con $\lambda = 0$, la segunda derivada cambia de signo cuando pasa por x^* : es positiva en un punto a la izquierda de x^* y negativa a la derecha de x^* ; y al revés con un repulsor: es negativa en un punto a la izquierda de x^* y positiva a su derecha. Aplicando lo dicho, se examina cada punto fijo.

$$f'_1(x_1) = -3(a_1 + a_2)x_1^2 + 2(a_1 + 2a_2)x_1 - a_2$$

- $P_1 = (x_1^* = 1) \Rightarrow f'_1(1) = -a_1$. Si $a_1 > 0 \Rightarrow f'_1(1) < 0$, entonces es atractor. Si $a_1 < 0 \Rightarrow f'_1(1) > 0$, entonces es repulsor.
- $P_2 = (x_1^* = 0) \Rightarrow f'_1(0) = -a_2$. Si $a_2 > 0 \Rightarrow f'_1(0) < 0$, entonces es atractor. Si $a_2 < 0 \Rightarrow f'_1(0) > 0$, entonces es repulsor.
- $P_3 = \left(x_1^* = \frac{a_1}{a_1 + a_2}\right) \Rightarrow f'_1\left(\frac{a_1}{a_1 + a_2}\right) = \frac{a_1 a_2}{a_1 + a_2}$. Sea $a_1 a_2 > 0$. Si $a_1, a_2 > 0 \Rightarrow f'_1(0) > 0$, entonces es repulsor. Si $a_1, a_2 < 0 \Rightarrow f'_1(0) < 0$, entonces es atractor.

9.6.3. Ascenso de gradiente de armonía: análisis de bifurcación.

Como no hay estructura jerárquica, $W = b = 0$, y se establece: $z = 0, \beta = 4$. En una sola dimensión, el gradiente para cada componente es simplemente su derivada respecto de x :

$$\begin{aligned}\nabla H_G &= Wx + b + e \\ \nabla H_B &= -\beta(x - z) \\ \nabla H_{Q_0} &= -4x(x - 1)(x + 1) = -4x(x^2 - 1)\end{aligned}$$

Con lo cual: $\nabla_x H(x, e_n, q) = \nabla H_G + \nabla H_B + q \nabla H_{Q_0}$ y entonces:

$$\frac{dx}{dt} = \nabla_x H(x, e_n, q) dt + \sqrt{2T} dW$$

Se analizará la parte determinística de la ecuación diferencial estocástica.

$$\frac{dx}{dt} = -4x - 4xq(x^2 - 1) + e$$

Si el término de sesgo $e = 0$, el sistema es equivalente al de una bifurcación tridente super-crítica con simetría. Para ver ésto, primero se practica una expansión de *Taylor*, alrededor de $c = 0$:

$$P(x) = f(c) + \frac{f^{(1)}(c)}{1!}(x - c) + \frac{f^{(2)}(c)}{2!}(x - c)^2 + \frac{f^{(3)}(c)}{3!}(x - c)^3 + \dots$$

con lo cual: $f(c) = 0$; $f^{(1)}(x) = -4 - 4q(3x^2 - 1) \Rightarrow f^{(1)}(0) = -4 + 4q$; $f^{(2)}(x) = -4q(6x) = -24qx \Rightarrow f^{(2)}(0) = 0$; $f^{(3)}(x) = -24q \Rightarrow f^{(3)}(0) = -24q$.

Entonces:

$$P(x) = 0 + (-4 + 4q)x + 0 + \frac{-24}{6}qx^3$$

La ecuación diferencial queda:

$$\frac{dx}{dt} = (-4 + 4q)x - 4qx^3$$

Para que esto se parezca a la forma normal de una bifurcación tridente super-crítica, es necesario deshacerse de los coeficientes del término cúbico. Para ello, sea la sustitución $x = \alpha u$, con α a determinarse. Con lo cual se tiene:

$$\begin{aligned}\frac{d(\alpha u)}{dt} &= (-4 + 4q)\alpha u - 4q(\alpha u)^3 \\ \frac{\alpha du}{dt} &= (-4 + 4q)\alpha u - 4q\alpha^3 u^3 \\ \frac{du}{dt} &= (-4 + 4q)\frac{\alpha}{\alpha}u - 4q\frac{\alpha^3}{\alpha}u^3 \\ \frac{du}{dt} &= (-4 + 4q)u - 4q\alpha^2 u^3\end{aligned}$$

tomando: $\alpha = \frac{1}{2q^{\frac{1}{2}}}$; y reemplazando:

$$\frac{du}{dt} = (-4 + 4q)u - 4q \left(\frac{1}{2q^{\frac{1}{2}}} \right)^2 u^3 = (-4 + 4q)u - u^3$$

y considerando $r = -4 + 4q$, se obtiene la forma normal de la bifurcación tridente super-crítica:

$$\frac{du}{dt} = ru - u^3$$

Para hallar los puntos críticos se resuelve¹⁰: $\frac{du}{dt} = ru - u^3 = 0$. con $r > 0 \Rightarrow q > 1$. Se encuentran tres puntos críticos: $u_1^* = 0$; $u_2^* = \sqrt{r}$; $u_3^* = -\sqrt{r}$ (obsérvese la simetría de los puntos dos y tres respecto del primero). Con $r = 0 \Rightarrow q = 1$, $u(r - u^2) = 0 \Rightarrow -u^2 = 0 \Rightarrow u^* = 0$. Y con: $r < 0 \Rightarrow q < 1$, $u(r - u^2) = 0 \Rightarrow u^* = 0$. Es decir que con $r \leq 0$ solo está el punto fijo $u^* = 0$. Para analizar la estabilidad de los tres puntos fijos, es preciso examinar el signo de la primera derivada: $f^{(1)}(u) = r - 3u^2$, evaluada en $u = u^*$.

- $u_1^* = 0$: $f^{(1)}(0) = r$. Si $r > 0 \Rightarrow q > 1$ es inestable (repulsor) y si $r < 0 \Rightarrow q < 1$ es estable (atractor). Ahora bien si: $r = 0 \Rightarrow q = 1$, la derivada es cero pero el punto no es un *saddle* ya que la segunda derivada $f^{(2)}(u) = 1 - 6u$ evaluada en $u_1^* - \varepsilon$, tiene signo positivo a la izquierda del punto fijo y a su derecha, $u_1^* + \varepsilon$, tiene signo negativo. Por lo tanto sigue siendo un punto estable pero débil (el flujo se acerca a $u_1^* = 0$ más lentamente).
- Los otros dos puntos fijos solo existen con $r > 0 \Rightarrow q > 1$. En primer lugar, $u_2^* = \sqrt{r}$: $f^{(1)}(\sqrt{r}) = r - 3r = -2r$, por lo tanto es atractor. Para el último: $u_3^* = -\sqrt{r}$: $f^{(1)}(-\sqrt{r}) = r - 3 \left((-1)^2 (\sqrt{r})^2 \right) = r - 3r = -2r$, por tanto también es atractor.

En suma, en $r \leq 0 \Rightarrow q \leq 1$ hay un solo punto fijo $u_1^* = 0$ y es estable. En $r = 0 \Rightarrow q = 1$ dicho punto sufre la bifurcación tridente (y es débilmente estable). Luego de la bifurcación, para $r > 0 \Rightarrow q > 1$, aparecen dos puntos fijos a distancia simétrica del anterior: $u_2^* = \sqrt{r}$ y $u_3^* = -\sqrt{r}$, ambos estables y el punto $u_1^* = 0$ se convierte en inestable.

En el caso de $e \neq 0$, aparece el fenómeno de una bifurcación tridente super-crítica imperfecta. En dicho caso, la constante e es un parámetro de perturbación del sistema que rompe la simetría. El sistema queda:

$$\frac{du}{dt} = ru - u^3 + \frac{1}{\alpha}e$$

Ahora el comportamiento del sistema pasa a depender de dos parámetros: $h = \frac{1}{\alpha}e$ y $r = -4 + 4q$, con $\alpha = \left(2q^{\frac{1}{2}}\right)^{-1}$. Los puntos fijos se encuentran donde se intersectan $f(u) = ru - u^3$ y $g(u) = h$. O sea:

$$\underbrace{ru - u^3}_{f(u)} + \underbrace{h}_{g(u)} = 0 \implies ru - u^3 = -h \implies f(u) = -g(u)$$

En el siguiente gráfico se muestra la dinámica del sistema para $r \leq 0 \Rightarrow q \leq 1$ (izquierda) y $r > 0 \Rightarrow q > 1$ (derecha)¹¹.

¹⁰En efecto: $r > 0 \Leftrightarrow -4 + 4q > 0 \Leftrightarrow 4q > 4 \Leftrightarrow q > \frac{4}{4} \Leftrightarrow q > 1$.

¹¹Imágenes tomadas de: https://it.wikipedia.org/wiki/Biforcazione_imperfetta.

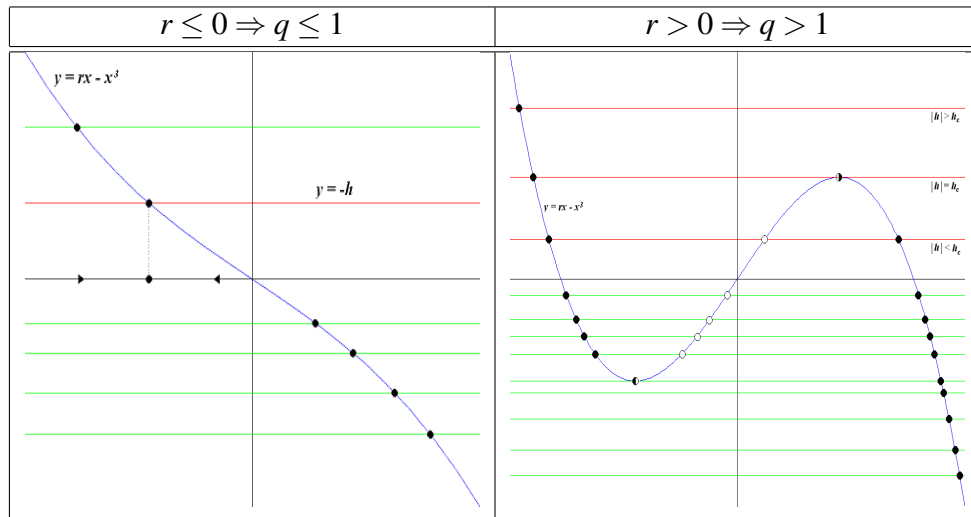


Figura 9.14: Bifurcación tridente super-crítica imperfecta.

En el caso $r \leq 0 \Rightarrow q \leq 1$ ambas funciones se intersectan siempre en un punto, no importa el valor de $-h$. Dicho punto fijo es siempre estable. Por ejemplo, para $h = 1$ y $q = 0 [\Rightarrow r = -4]$ el punto fijo es $u^* \approx \frac{1}{4}$ (Esto se puede ver en la Figura 10, arriba a la izquierda, donde la parábola tiene su máximo aproximadamente en 0.25, que coincide con el atractor).

En lo que atañe al caso $r > 0 \Rightarrow q > 1$, puede haber un punto fijo o tres, dependiendo del punto crítico $h_c(r)$. En el punto crítico, la función $g(u)$ es tangente a $f(u)$ y, dado que allí hay un punto máximo o mínimo, $f^{(1)}(u) = 0$. La condición de tangencia establece que $f^{(1)}(u) = g^{(1)}(u)$. Con lo cual se tiene:

$$f^{(1)}(u) = r - 3u^2; g^{(1)}(u) = 0; \Rightarrow f^{(1)}(u) = g^{(1)}(u) \Leftrightarrow r - 3u^2 = 0 \Leftrightarrow u_c^* = \pm\sqrt{\frac{r}{3}}.$$

Luego el punto crítico resulta:

$$h_c(r) = ru - u^3 = r\sqrt{\frac{r}{3}} - \frac{r}{3}\sqrt{\frac{r}{3}} = \left(1 - \frac{1}{3}\right)r\sqrt{\frac{r}{3}} = \frac{2}{3}r\sqrt{\frac{r}{3}}$$

Según se trate del máximo o del mínimo de $f(u)$, se tendrá: $h = \pm h_c(r) = \frac{2}{3}r\sqrt{\frac{r}{3}}$.

En dicho punto se lleva a cabo una bifurcación *saddle node*. Se tienen los casos:

condición	puntos fijos	estabilidad
$ h > h_c$	un punto fijo	estable
$ h < h_c$	tres puntos fijos	dos estables y uno inestable
$ h = h_c$	punto fijo crítico u_c^*	<i>saddle node</i>

Cuadro 9.10: Puntos fijos para: $r > 0 \Rightarrow q > 1$

Mírese la Figura 14 a la derecha, comenzando desde arriba. Cuando $|h| > h_c = +\frac{2}{3}r\sqrt{\frac{r}{3}}$, hay un único punto fijo estable. Luego en $|h| = h_c$ se produce una bifurcación *saddle node*. Y a partir de allí, en $|h| < h_c$ hay dos puntos fijos más, uno inestable y otro estable. Los puntos estables se encuentran a ambos lados del inestable pero no a distancia simétrica. En consecuencia se ha obtenido el resultado de una bifurcación tridente super-crítica pero a través de una bifurcación *saddle-node*.

En el ejemplo de la Figura 10, debajo a la derecha, $q = 4,96 > 1 \Rightarrow r = 15,94 > 0$; las raíces (puntos fijos) escaladas son $x^* = \alpha u^* = \left(2q^{\frac{1}{2}}\right)^{-1} u^* = (-0,86; -0,06; 0,92)$ [$u^* = (-0,28; -3,83; 4,11)$], que coinciden aproximadamente con los dos máximos y el mínimo global del gráfico en cuestión.

Capítulo 10

Dinámica de series de error.

10.1. Introducción.

El presente capítulo intentará cuantificar la pregunta general “¿el alumno mejora en su aprendizaje?”. Desde el punto de vista de los sistemas complejos, esto implica detectar posibles transiciones entre regímenes de errores, sesiones de dinámica de error similar y aquellas con dinámica diferente. Se responderán las siguientes preguntas: (i) ¿cuáles sesiones tienen dinámica de error similar?; (ii) ¿Qué regímenes de dinámica diferente es posible identificar?; (iii) En particular, ¿cuáles sesiones destacan frente al resto por tener una mejor dinámica del error?; (iv) si hay un cambio en el aprendizaje, ¿Dónde se encuentra la transición?.

10.2. “Motivos”.

10.2.1. Motif y LDA.

Los “motivos” [*motif*] son series de símbolos discretos $S = \{S_1, \dots, S_{\mathcal{L}}\}$ pertenecientes a un alfabeto $A = \{A_1, \dots, A_n\}$, cuyo tamaño es $|A| = \mathcal{N}$. Para descubrir “motivos” se hace correr en la serie una ventana de largo \mathcal{L} . En cada ventana se normalizan los datos. Se establece un alfabeto A de tamaño $a = |A|$. Luego, se establecen B “puntos de corte” en los cuantiles de la normal sobre los datos estandarizados (restándoles la media y dividiendo por el desvío) en la ventana, tal que la serie de cortes $B = \beta_1, \dots, \alpha_{a-1}$, tenga un área bajo la curva de la normal entre β_i y β_{i+1} igual a $\frac{1}{a}$. Es decir que los “puntos de corte” producen áreas de igual tamaño bajo la curva de la distribución. Los datos que caen bajo cada área definida por los cortes reciben un símbolo común del alfabeto. El procedimiento puede ser precedido por una reducción de dimensión de los n datos a w datos ($w < n$) en el cual se dividen los datos en w segmentos y se toma en cada segmento la media (Lonardi et al., 2002). Esto último no se llevó a cabo aquí. Se partió de la respuesta binaria “1” = *error*; “0” = *no error* y se la consideró como números enteros. Se estableció $A = \{a, b\}$, $a = |A| = 2$, $\mathcal{L} = 3$. Por ejemplo si una ventana es [011] su estandarización es $[-1,15; 0,57; 0,57]$. Hay un solo punto de corte en $p = \frac{1}{a} = 0,5$, por lo tanto el “motivo” será: “abb”. El Cuadro que sigue muestra los motivos presentes en las series de los cuatro aprendientes.

	motivo	secuencia
1	aaa	000
2	aab	001
3	aba	010
4	abb	011
5	baa	100
6	bab	101
7	bba	110
8	bbb	111

Cuadro 10.1: Motivos y secuencias que representan.

Para cada sesión de cada aprendiz se encontraron los “motivos”. Se consideró que cada *motivo* representaba un “estado” diferente de dinámica, con dos atractores “a = correcto” y “b = error” según: (i) atractor “a” fuerte: *aaa*; (ii) atractor “a” débil: *aab, aba, baa*; (iii) atractor “b” fuerte: *bbb*; (iv) atractor “b” débil: *bba, bab, abb*. Se interpretó “fuerte” como permanencia en el atractor y “débil” como un solo apartamiento del atractor hacia el otro. El objetivo consistió en clasificar las sesiones en grupos de dinámica similar, según “estados” similares. Para ello, se consideró a cada sesión como un “texto” cuyas palabras eran los “motivos”. Se organizaron los datos en términos de una matriz $d \times t$, con d sesiones / documentos en las filas y t “motivos” / términos en las columnas. En cada celda se contó el número de ocurrencias de un “motivo” determinado en una sesión determinada. Se asignó a cada término / “motivo” t en el “documento” / sesión d un peso según el esquema:

$$w_{t,d} = tf \times idf = tf_{t,d} \times \log_2 \left(\frac{N}{df_t} + 1 \right)$$

Donde: N es el número de documentos, df_t es el número de documentos que contienen el término t , $tf_{t,d}$ es el número de ocurrencias del término t en un documento d . No se consideró la primera sesión de SONIA por tener solamente cuatro instancias, con lo cual hubo 51 sesiones / documentos. El peso aumenta con la cantidad de ocurrencias de un “motivo” en una sesión y también con su rareza en la colección de documentos / “sesiones”. A continuación se hizo un análisis para descubrir “tópicos”, como si cada tópico fuera un grupo de dinámica diferente. Para ello se aplicó *latent Dirichlet allocation (LDA)* [Blei et al., 2003]. Esta técnica permite descubrir los tópicos de los que hablan los textos, asignando una probabilidad a cada tópico. En concreto, en LDA cada documento se representa como una mezcla de distribuciones sobre tópicos latentes (no observados) y cada tópico se caracteriza por una distribución asimétrica sobre palabras, en donde solo algunas palabras de V tendrán alta probabilidad. Defínase: (i) una palabra w de un vocabulario $V = \{1, \dots, V\}$ como un vector de V componentes donde $w_i = 1$ si $V = i$ y $w_i = 0$ si $V = j$ ($i \neq j$); (ii) un documento w como una secuencia de N palabras denotada por $w = \{w_1, \dots, w_N\}$; (iii) un corpus D como una secuencia de M documentos denotada por $D = \{w_1, \dots, w_M\}$; (iv) un tópico φ de la colección $\Phi = \{\varphi_1, \dots, \varphi_k\}$. LDA es un modelo bayesiano generativo (en donde las observaciones están generadas por variables latentes), en el cual se asume que cada documento w de largo N_i en el corpus D se genera de la siguiente manera.

1. Se eligen N palabras del documento muestreando de una distribución de *Poisson*: $N \sim \text{Pois}(\lambda)$.
2. Se elige el parámetro θ_i de la distribución del *prior* sobre los tópicos muestreando de una distribución de *Dirichlet*: $\theta_i \sim \text{Dir}(\alpha)$; $i = 1, \dots, M$.
3. Se elige el parámetro φ_k de la distribución del *prior* sobre las palabras muestreando de una distribución de *Dirichlet*: $\varphi_k \sim \text{Dir}(\beta)$; $k = 1, \dots, K$.
4. Para cada palabra j ($j \in \{1, \dots, N_i\}$) en cada documento i ($i \in \{1, \dots, M\}$):
 - elegir un tópico z_{ij} muestreando de una distribución multinomial $z_{ij} \sim M(\theta_i)$.
 - elegir una palabra w_{ij} muestreando de una distribución multinomial $w_{ij} \sim M(\varphi_{z_{ij}})$. Cada palabra w_{ij} tendrá una probabilidad condicional $p(w_{ij} | z_{ij}, \beta)$.

La probabilidad conjunta es la siguiente:

$$p(W, Z, \theta, \varphi; \alpha, \beta) = \prod_{i=1}^M p(\theta_i; \alpha) \prod_{j=1}^N p(\varphi_{z_{ij}}; \beta) \prod_{t=1}^N p(Z_{j,t} | \theta_j) p(W_{j,t} | Z_{j,t}, \varphi)$$

donde W, Z, θ y φ son vectores de parámetros.

Por ejemplo, si se tiene el texto: [el **lunes** desayuné con **cereales**; el **martes**, con **fruta**] y se extraen los términos en negrita (lematizando) para formar el documento: $w = \{\text{lunes}, \text{cereal}, \text{martes}, \text{fruta}\}$. Se eligen entonces $N = 4$ palabras. Luego se eligen las probabilidades sobre los tópicos, por ejemplo que el

documento versará un 50 % sobre *tiempo* [tópico A] y 50 % sobre *comida* [tópico B]. Entonces las palabras (observadas) generadas por los tópicos (variables latentes) serán: (1) “lunes” [tópico A], (2) “cereal” [tópico B], (3) “martes” [tópico A], (3) “fruta” [tópico B].

10.2.2. Análisis y evaluación de resultados.

Se aplico LDA^1 a la matriz $d \times t$ con pesos $tf \times idf$. Se eligieron $K = 3$ tópicos, $\alpha = 0,1$; $\beta = 0,001$. Para evaluar la agrupación encontrada, se formaron los grupos g_i [$i = 1, 2, 3$] de referencia utilizando las frecuencias relativas de error (f_{error}) de las sesiones / documentos d : $g_1 = \{d : f_{error} \in [0,21, 0,35]\}$; $g_2 = \{d : f_{error} \leq 0,2\}$; $g_3 = \{d : f_{error} > 0,35\}$. Los datos con la asignación de tópicos a cada sesión y aquellos de referencia se muestran en el apéndice del capítulo. Se calculó la tabla de confusión con los grupos predichos en las filas y aquellos de referencia en las columnas. De ella se derivaron las medidas siguientes. Si se tiene la tabla de confusión, donde: TP = “true positives”; TN = “true negatives”; FP = “false positives”; FN = “false negatives”; se pueden derivar las medidas descriptas en el Cuadro que le sigue².

	+	-
+	TP	FP
-	FN	TN

Cuadro 10.2: Tabla de confusión.

Medidas	Fórmula	Definición
<i>accuracy (acc)</i>	$\frac{TP+TN}{FP+TN+TP+FN}$	proporción de clasificados correctamente del total
<i>sensitivity (sens)</i> [recall]	$\frac{TP}{TP+FN}$	Proporción de positivos clasificados correctamente y que lo son en el patrón de referencia
<i>specificity (spec)</i>	$\frac{TN}{TN+FP}$	Proporción de positivos clasificados correctamente y que lo son en el patrón de referencia
<i>positive predictive value (ppv)</i> [precisión]	$\frac{TP}{TP+FP}$	Proporción de positivos clasificados correctamente (TP) de todos los clasificados como positivos (TP+FP)
<i>negative predictive value (npv)</i>	$\frac{TN}{TN+FN}$	Proporción de negativos clasificados correctamente (TN) de todos los clasificados como negativos (TN+FN)
<i>Kappa</i>	$(totalAcc - randomAcc) / (1 - randomAcc)$	medida de “acuerdo” entre predichos y referencia
<i>balance accuracy</i>	$\frac{sens+spec}{2}$	accuracy para clases desbalanceadas
<i>F1</i>	$2 \times \left[\frac{precision \times recall}{precision+recall} \right]$	media armónica entre precisión y recall

Cuadro 10.3: Medidas derivadas de la tabla de confusión.

Los Cuadros siguientes muestran la tabla de confusión y las medidas derivadas. La precisión global fue de 0,803 (C.I. 95 % : [0,668;0,901]) y el coeficiente *Kappa* de 0,693. Con respecto a las medidas por clase, casi todas se hallan por encima de 80 %. Por otra parte, el área bajo la curva (AUC) de una gráfica de la *sensibilidad* versus las falsas alarmas (*1-especificidad*) arroja: $AUC(1,2) = 0,753$; $AUC(1,3) = 0,869$; $AUC(2,3) = 0,866$. Ya que $AUC = 1$ indica clasificación perfecta, se nota que el modelo distingue menos entre los grupos 1 y 2.

Por último se muestran gráficamente los tópicos, indicando los “motivos” que los componen. Como el tópico 2 coincide con proporciones de error menores al 21 %, se observa que la mayor cantidad de “motivos” tienen que ver con el atractor “a” [*aaa, baa, aab, aba*]. En este grupo, el estado fuerte del atractor “correcto” es el que tiene más instancias, respecto de los otros dos tópicos. En el otro extremo, el atractor 3 se asocia a probabilidades de error mayores al 35 %, los estados “bbb” [atractor “error” fuerte], “bba”, “abb”, “bab”

¹Utilizando la librería *tex2vec* de R.

²En el caso de la clase ternaria se hace una tabla de confusión de cada clase, como clase positiva, contra la suma de las otras como clase negativa.

	1	2	3
1	18	3	1
2	2	15	1
3	3	0	8

Cuadro 10.4: Matriz de confusión. Filas: predicción, Columnas: referencia

	Class: 1	Class: 2	Class: 3
Sensitivity	0.78	0.83	0.80
Specificity	0.86	0.91	0.93
Pos Pred Value	0.82	0.83	0.73
Neg Pred Value	0.83	0.91	0.95
F1	0.80	0.83	0.76
Balanced Accuracy	0.82	0.87	0.86

Cuadro 10.5: Matriz de confusión. Medidas de evaluación de la clasificación.

[atractor “error” débil] registran más frecuencia comparados con los otros dos tópicos y el atractor “aaa” resulta el de menos frecuencia en comparación. Por último el tópico 1, asociado a proporción de errores de 21 a 35%, registra más frecuencia en los estados débiles de los atractores de “error” [*baa*, *aab*, *aba*] comparado con el tópico 1; y menos en los estados débiles del atractor “no error” [*bba*, *abb*, *bab*] comparado con el tópico 3. Es decir que hay una dinámica intermedia.

En resumen, los tópicos logran captar tres tipos de dinámica asociadas a franjas de probabilidad de error de las sesiones. La tabla tres muestra a las sesiones de cada alumno agrupadas por tópico de dinámica similar. Como es esperable, SONIA y JAKO tienen más sesiones en el tópico 2 [error < 0.21] y ninguna en el 3 [error > 0.35]. En contraste, NATI y MIRKA tienen casi la mitad de las sesiones en el tópico 3 y como máximo dos en el tópico 2.

ALUMNO	TÓPICO 1	TÓPICO 2	TÓPICO 3
SONIA	2,4,6,9,11	3,5,7,8,10,12	-
NATI	3,7,9,11,12,13	4,14	1,2,5,6,8,10
JAKO	1,4,7,8,10,13	2,3,5,6,9,11,12,14	-
MIRKA	3,4,5,6,12	1,8	2,7,9,10,11

Cuadro 10.6: Agrupación de sesiones según tópicos de dinámica similar. Las celdas contienen el número de sesión.



Figura 10.1: Tópicos y motivos que los componen.

10.3. El plano entropía - complejidad.

En el campo de la teoría de la información (Cover & Thomas, 1991), la información de un evento en particular (“errores”, “ítems correctos”, etc.) está definida como el logaritmo (en base 2, si la unidad es el *bit*) de la probabilidad de dicho evento: $\log_2 [p(x)]$. Los eventos improbables nos dan mayor información que los probables (nos “sorprenden” más). La entropía de *Shannon* sobre una función de masa de probabilidad (IMP) de la serie de los eventos³: $\vec{\pi} = \{\pi_1, \dots, \pi_n\}$ reza: $H(S) = -\sum_{i=1}^N \pi_i \log_a \pi_i$. Aquí $a = 2$ y entonces su resultado está expresado en *bits*. Si ahora definimos una ventana de longitud L que se deslice en la serie S (de largo \mathcal{L}) un símbolo a la vez, cada deslizamiento revelará una “palabra” $W(L)$, habiendo en total $\mathcal{L} - L + 1$ palabras presentes de $|\mathcal{A}|^L$ posibles palabras de longitud L (\mathcal{A} es un alfabeto con $|\mathcal{A}|$ cantidad de símbolos). Con el conteo de frecuencias absolutas de palabras $W = \{w_1, \dots, w_{|\mathcal{A}|^L}\}$ se puede construir una IMP $\vec{\pi} = \{\pi_1, \dots, \pi_{|\mathcal{A}|^L}\}$ donde cada probabilidad se normaliza como $\pi_i = w_i / (\mathcal{L} - L + 1)$. De este modo es posible aplicar $H(S)$ a palabras de longitud L , que representarán ahora los estados del sistema de aprendizaje. Es decir, definiéndola como una entropía de *Shannon* de bloque $H(L > 1)$ (Larson et. al., 2011; Crutchfield & Feldman, 2003), que reza ahora : $H(L) = -\sum_{i=1}^N \pi_{L_i} \log_a \pi_{L_i}$, siendo $H(L = 1) = H(S)$. Cuando $H(S) = H(L) = 0$ (es mínima), podemos predecir con exactitud la palabra $W(L)$ que aparecerá a continuación en la serie, o sea tendremos incertidumbre mínima (o lo que es igual, conocimiento máximo), ya que la IMP consta de una única probabilidad: $\vec{\pi} = \{\pi = 1\}$, por ejemplo si $S = \{0, 0, 0, \dots\}$ [ausencia de error]. En cambio nuestra incertidumbre es máxima (y nuestra ignorancia también) cuando las palabras

³Una función de masa de probabilidad (IMP) asocia a cada tipo de símbolo del alfabeto $\mathcal{A} = \{A_1, \dots, A_n\}$ presente en una serie, una probabilidad $\vec{\pi} = \{\pi_1, \dots, \pi_n\}$, de manera que sus valores estén en el rango $0 \leq \pi_i \leq 1$ y sumen 1 ($\sum_{i=1}^N \pi_i = 1$).

$W(L)$ siguen una IMP uniforme: $\pi_i = \frac{1}{|\mathcal{A}|^L}$ y entonces: $H(S) = \log_2 |\mathcal{A}|$ y $H(L > 1) = L \times \log_2 |\mathcal{A}|$. Además, en el caso de una serie periódica con período P , $H(L_P) = \log_2(P)$. Ahora bien, una situación de máximo / mínimo conocimiento-incertidumbre no es compleja; o sea, sencillamente existe completo orden o desorden; por ello muchas medidas de complejidad (C) asignan $C = 0$ cuando H es mínima o máxima. A través de un plano de incertidumbre / complejidad $H \times C$, podemos crear el espacio de fase del sistema de aprendizaje de un aprendiente, en el que cada estado (sesión) es un punto en dicho plano; e investigar el movimiento / trayectoria hacia, por ejemplo, un punto atractor que representa al hablante nativo (ausencia de error) o una región de atracción que represente al nivel superior, una vez cruzado un punto crítico.

10.3.1. Entropía de permutación.

La entropía de permutación [*permutation entropy*] fue originariamente propuesta por Bandt & Pompe (2002). En lugar de tomar como base para el cálculo de la entropía de Shannon palabras W de longitud L , ellos propusieron asignar a cada palabra un patrón ordinal obtenido por comparación de los valores (símbolos) adyacentes que contiene la palabra W . Más formalmente (ver: Zanin et. al., 2012; Ribeiro et. al., 2012), considérese una serie de tiempo $\{x_t : t = 1, \dots, N\}$ y en cada punto s de dicha serie, un vector (palabra W) compuesto por D valores subsiguientes: $s \mapsto (x_s, x_{s+1}, \dots, x_{s+(D-2)}, x_{s+(D-1)})$. D se denomina *dimensión* y define el largo del vector/palabra. A cada palabra se asocia luego un patrón ordinal en términos de la permutación $\pi = (r_0, r_1, \dots, r_{D-1})$ del vector $(0, 1, \dots, D-1)$ de forma que se cumpla el orden creciente: $x_{s+r_0} \leq x_{s+r_1} \leq \dots \leq x_{s+r_{(D-2)}} \leq x_{s+r_{(D-1)}}$. Por ejemplo, para la serie $\{3, 6, 8, 9, 5, 10, 2\}$ y con $D = 2$ tenemos para $s = 1$ la palabra $(3, 6)$ que ya se encuentra en orden creciente, produciendo el patrón $\pi = (01)$. Para $s = 2$, la palabra $(6, 8)$ también produce $\pi = (01)$; en cambio para $s = 4$ obtenemos la palabra $(9, 5)$; que puesta en orden creciente es $(5, 9)$, y se le asigna el patrón $\pi = (10)$. Se habrá notado entonces que para $D = 2$, solo es posible un patrón creciente o uno decreciente, es decir, $\vec{\pi} = \{\pi_1 = (01), \pi_2 = (10)\}$; en general los posibles patrones para cada D siguen al factorial $D!$ si los valores de la serie son diferentes entre sí (no se repiten). El siguiente paso consiste en generar las probabilidades para cada patrón, siguiendo: $p(\pi_i) = \frac{\#\{s:s < N-D+1; s \in \pi_i\}}{N-D+1}$, donde $\#$ representa las frecuencias absolutas de cada patrón π posible. Se forma entonces la IMP $\vec{\pi} = \{p(\pi_1), \dots, p(\pi_{D!})\}$. Se define pues la entropía de permutación $H_s[P]$ en términos de la entropía de Shannon para los patrones π normalizada sobre el logaritmo de los patrones π posibles como sigue:

$$H_s[P] = - \frac{1}{\log_2 D!} \sum_{i=1}^{D!} p(\pi_i) \log_2 p(\pi_i)$$

Como se usa el logaritmo en base dos, está expresada en bits.

La IMP para el ejemplo es: $\vec{\pi} = \{p(\pi_1 = 01) = \frac{4}{6}, p(\pi_2 = 10) = \frac{2}{6}\}$.

Con lo cual se tiene: $H_s[P] = - \frac{1}{\log_2(2)} (\frac{4}{6} \log_2 \frac{4}{6} + \frac{2}{6} \log_2 \frac{2}{6}) = 0,9183$.

La medida de complejidad asociada (Lamberti et al., 2004; Ribeiro et al., 2012) está definida como el producto de la entropía de shannon normalizada $H_s[P]$ y una medida de desequilibrio $Q_j[P, P_e]$, la denominada distancia de Jensen - Shannon, y que mide el grado de separación entre la distribución efectiva $P = \vec{\pi}$ de los patrones ordinales y una hipotética distribución uniforme de éstos: $P_e = \{\frac{1}{D!}\}$. Formalmente:

$$C_{sj}[P] = Q_j[P, P_e] H_s[P]$$

$$Q_j[P, P_e] = \frac{S[\frac{P+P_e}{2}] - \frac{S[P]}{2} - \frac{S[P_e]}{2}}{Q_{max}}$$

$$Q_{max} = -\frac{1}{2} \left[\frac{D!+1}{D!} \log_2(D!+1) - 2 \log_2(2D!) + \log_2(D!) \right]$$

La última expresión es la distancia máxima entre P y P_e cuando uno de los componentes de P tiene probabilidad 1 y los demás probabilidad 0. Con esto es posible investigar la dinámica del sistema de aprendizaje en el plano $\vec{H}_s \times \vec{C}_{js}$.

10.3.2. Significancia de las medidas.

En primer lugar se describe el test de permutación para tendencia central para dos muestras *independientes* [Boninni et al., 2014].

Asumiendo homoscedasticidad en la hipótesis nula, supóngase dos distribuciones cualesquiera continuas y no-degeneradas: F_1 y F_2 de la misma familia F . H_0 implica el intercambio de los datos observados respecto de los dos grupos. O sea, se los puede considerar como si hubieran sido asignados de modo aleatorio a ambos grupos pero en realidad provienen de la misma población. Considerar $X = X_1 \uplus X_2$, la concatenación de dos vectores con los datos observados de las dos muestras; con los primeros n_1 elementos de X_1 y los demás n_2 elementos de X_2 . El estadístico de permutación es $T^* = \bar{X}_1^* - \bar{X}_2^*$ donde $\bar{X}_j^* = \frac{1}{n_j} \sum_{i=1}^{n_j} X_{ij}^*$ ($j = 1, 2$) es la media muestral de los primeros n_1 elementos y los segundos n_2 de X^* . Además X^* representa la base de datos permutada, es decir, asignando aleatoriamente n_1 elementos al primer grupo y n_2 elementos al segundo. Bajo H_0 , la distribución de T^* se obtiene permutando la base de datos B veces de modo independiente. El p-valor del test es $p = \sum_{b=1}^B \frac{1}{B} I(T_{(b)}^* \geq T^0)$, o sea, la proporción de T^* valores del estadístico iguales o mayores a T^0 ; donde $T_{(b)}^*$ es el valor del estadístico de la b-ésima permutación y T^0 es el valor observado del estadístico en la muestra conjunta X ,

Ahora bien si se calculan las medidas H_s^{ruido} y C_{js}^{ruido} para un proceso de ruido estas deberían dar cercanas a uno para la entropía y cercanas a cero para la complejidad. Se genera una muestra de ruido para cada sesión (de n igual al de la sesión). Luego, se comparan los vectores con los valores de entropía H_s con H_s^{ruido} y C_{js} con C_{js}^{ruido} para cada aprendiente mediante el test de permutación anterior. Si se rechaza la hipótesis nula de igualdad de medias (diferencia de medias igual a cero) entre los datos observados y los que provienen de datos sin estructura, entonces se concluye que los datos observados poseen información relevante.

10.3.3. Resultados.

Se utilizó la respuesta categórica con los valores $\{0, 1, 2, 3, 4\}^4$ y se la consideró como cuantitativa discreta a los efectos de aplicar ambas medidas. Se calcularon H_s y C_{js} para cada sesión de cada aprendiente (los valores obtenidos se hallan en el apéndice del capítulo). Se estableció $D = 2$ ya que hay sesiones con pocos datos⁵. Se evaluó la significatividad de las medidas mediante el procedimiento detallado en el apartado anterior. El Cuadro que sigue muestra los p-valores del test de permutación y del test (no paramétrico) de correlación de *Spearman*. Según este último no hay correlaciones significativas entre los vectores de medidas derivados de los datos y aquellos provenientes del ruido. Por lo tanto, el test de permutación para dos muestras independientes resulta una buena elección. Este último rechaza la hipótesis nula de igualdad de medias para todos los aprendientes, con lo cual se considera que las medidas son informativas.

	id	e.cor	e.perm	c.cor	c.perm
1	1	0.875	0.011	0.265	<0.001
2	2	0.298	<0.001	0.344	<0.001
3	3	0.976	<0.001	0.697	<0.001
4	4	0.837	<0.001	0.862	<0.001

Cuadro 10.7: Test de permutación: e.cor= p valor de S de Spearman para PE; e.perm= p valor del test para PE; c.cor= p valor de S de Spearman para C; e.perm= p valor del test para C.

⁴Recuérdese la codificación: 0 = "no error"; 1 = "error de género"; 2 = "error de e-pentética"; 3 = "error de plural"; 4 = "error mixto".

⁵Amigó et al. (2008) proponen la relación $\mathcal{L} \geq 5D!$ (\mathcal{L} es el largo de la serie), con $D = 2$ se tiene $\mathcal{L} \geq 5(2!) = 10$. La única sesión con $n \leq 10$ es la primera de SONIA.

En la Figura 2 se muestra el plano entropía-complejidad para todos los aprendientes. Se han ubicado en cada sesión su proporción de errores y se colorearon los puntos según las franjas de frecuencia relativa de error: (i) $f_{error} \leq 0,20$; (ii) $0,21 \leq f_{error} \leq 0,35$; (iii) $f_{error} > 0,35$.

Se ubicó el atractor “sin error” en el punto $(H_s = 0, C_{js} = 0)$ y el del “error” en el punto $(H_s = 1, C_{js} = 0)$. Se observa que para $H_s < 0,55$ la mayoría de los puntos coincide con la primera franja. En contraste, para $H_s > 0,80$ la mayoría de los puntos está en la tercera franja. Por lo tanto se estableció el punto crítico (arbitrario) $f_{error} = 0,21$ para designar aquel punto más allá del cual se ingresa en el “basin of attraction” del atractor “no error”; y el punto crítico (arbitrario) $f_{error} = 0,35$ para designar aquel punto más allá del cual se ingresa en el “basin of attraction” del atractor “error”. La región intermedia se designó de “transición” entre ambos regímenes. Los gráficos de la Figura 3 ilustran los resultados para cada aprendiente. El Cuadro 8 ubica las sesiones de los aprendientes en cada régimen.

JAKO es el aprendiente con más sesiones en el régimen “NO_ERROR”, además la sesión 14 corresponde al punto del atractor (“no error”), y posee más sesiones cercanas a dicho punto. No hay sesiones en la región “ERROR”. Nótese además que hay tanto sesiones iniciales como finales en la región “NO_ERROR”. Por lo tanto, JAKO solamente consolida su nivel de competencia, que es el más alto de los cuatro. SONIA, a pesar de haber hecho un curso de nivel A evidencia solamente un par de sesiones en la región “ERROR”, además son las dos sesiones iniciales de la serie en dicha región. Por ende, SONIA está dejando el atractor del “ERROR”. MIRKA se encuentra todavía en la región del “ERROR” en tres sesiones, tanto del principio, del medio y del final de la serie. Por lo tanto, la atracción al error es fuerte. NATI parece estar en una situación similar: con sesiones del principio, medio y final de la serie en la región del “ERROR”, pero con una sesión más en la región “NO_ERROR”, incluida la última. En resumen, el plano entropía-complejidad confirma el orden de competencia: JAKO > SONIA > NATI > SONIA.

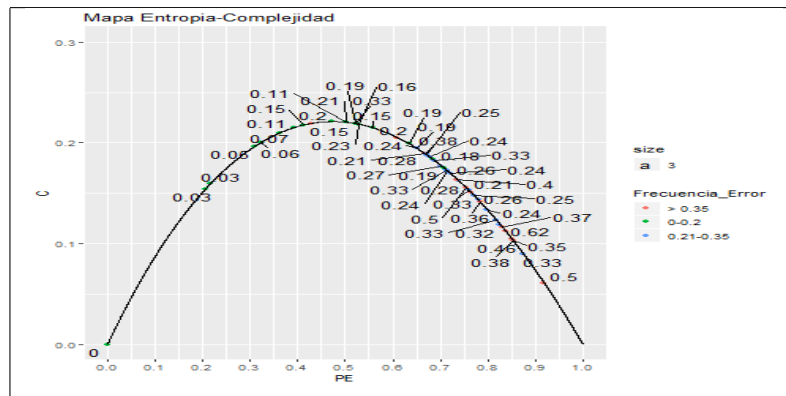


Figura 10.2: El plano entropía-complejidad para todos los aprendientes. Se señaló la proporción de error correspondiente a la sesión. EP: entropía de permutación H_s ; C: complejidad C_{js} .

	NO_ERROR	TRANSICION	ERROR
SONIA	3,7,10,12	4,5,6,8,9,11	1,2
NATI	4,5,14	1,3,6,8,9,11,13	2,7,10,12
JAKO	1,2,3,6,9,12,14	4,5,7,8,10,11,13	
MIRKA	1,8	2,4,5,6,9,10,12	3,7,11

Cuadro 10.8: Sesiones en cada región del plano entropía-complejidad

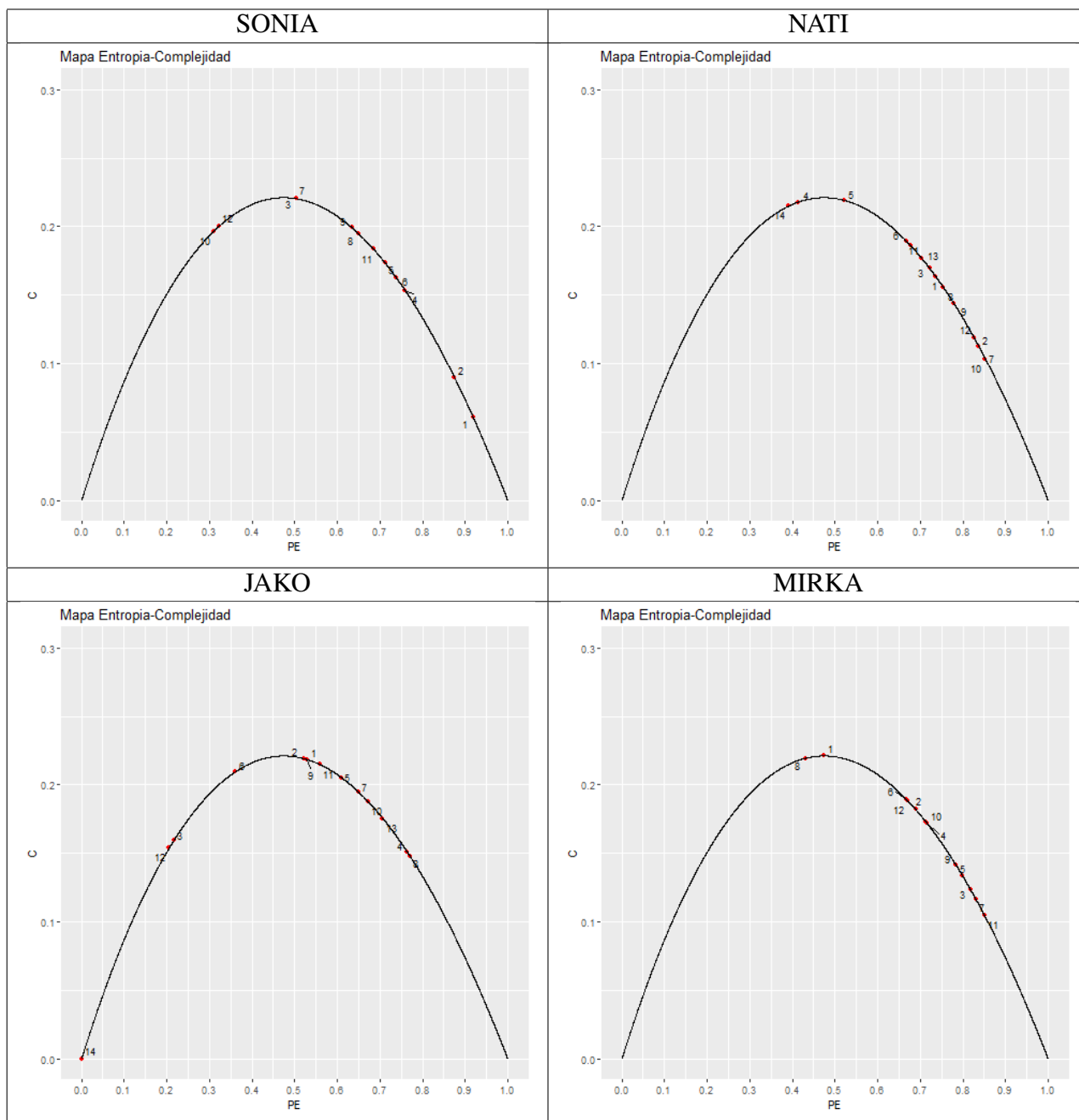


Figura 10.3: El plano entropía-complejidad para cada aprendiente. EP: entropía de permutación H_S ; C: complejidad C_{js} .

10.3.4. Análisis de cambio de punto.

“*Change Point Analysis*” es una metodología que sirve para estimar el punto de una serie en el cual cambian las propiedades estadísticas de dicha serie. Se usará el método no paramétrico propuesto por Dehling et al. (2018), útil para series de datos no independientes y asimétricos. Se supone que las observaciones $(X_n)_{n \geq 1}$ provienen del modelo $X_n = \mu_n + Y_n$, donde $(\mu_n)_{n \geq 1}$ es una secuencia de constantes desconocidas y $(Y_n)_{n \geq 1}$ es un proceso estacionario con media cero. Dada las observaciones X_1, \dots, X_n se desea testear la hipótesis nula de que el proceso es estacionario: $H_0 : \mu_1 = \dots = \mu_n$ contra la alternativa H_1 de que existe un punto k en el tiempo en el cual la locación (mediana) cambia: $\exists k : k \in \{1, \dots, n-1\}$ tal que $\mu_1 = \dots = \mu_k \neq \mu_{k+1} = \dots = \mu_n$. Para cada punto k se calcula el siguiente estadístico basado en el estimador de *Hodges-Lehmann*⁶ entre la serie segmentada entre dos grupos $[1, k]$ y $[k+1, n]$ y se elige el máximo en valor absoluto:

⁶Dados dos grupos independientes de observaciones X_1, \dots, X_{n_1} y Y_1, \dots, Y_{n_2} el estimador de *Hodges-Lehmann* se define como la mediana de sus diferencias: $med \{ (Y_j - X_i) : 1 \leq i \leq n_1, 1 \leq j \leq n_2 \}$.

$$T_n = \sqrt{n} \max_{1 \leq k \leq n} \frac{k}{n} \left(1 - \frac{k}{n} \right) | \text{med} \{ (X_j - X_i) : 1 \leq i \leq k, k+1 \leq j \leq n \} |$$

El estimador rechaza la hipótesis nula para valores (negativos o positivos) grandes del estadístico.

Se calculó H_s ($D = 3$) por medio de una ventana móvil de largo $w = 80$ para cada aprendiente. Ya que los datos de la serie creada por la ventana móvil no son independientes, se aplicó el método explicado para descubrir un punto de cambio de locación⁷. Las gráficas que siguen [Figura 5] muestran la ubicación del cambio de locación para cada aprendiente. Las líneas punteadas marcan la separación entre las sesiones. Además se muestra la serie del estimador en la Figura 4. En esta última, la línea roja marca el máximo (en valor absoluto) y las punteadas son bandas de confianza más allá de las cuales el estadístico es significativo.

Notar que cada punto i representa la entropía de la serie temporal original en los puntos $[i, i + w]$.

Los cambios en locación indican cambios en la dinámica. Para SONIA el cambio se produce entre finales de la sesión 5 y mediados de la sesión 8 (ventana de puntos[112, 192]); para NATI, durante las sesiones 5 y mediados de la 7 (ventana de puntos[82, 162]); en JAKO, durante el periodo entre las sesiones 9 y principios de la 11 (ventana de puntos[280, 360]). Por último, para MIRKA; entre las sesiones 2 y principios de la 3 (ventana de puntos[61, 141]). Sin embargo el p-valor del estadístico en MIRKA es $p = 0,047$, casi no significativo, con lo cual no hay mucha certeza de que el cambio esté allí. Recuérdese también que cuanto más baje H_s menos errores hay en la ventana. Por lo tanto, se observa que SONIA y JAKO cambian a una mejor dinámica; pero NATI y MIRKA a una peor.

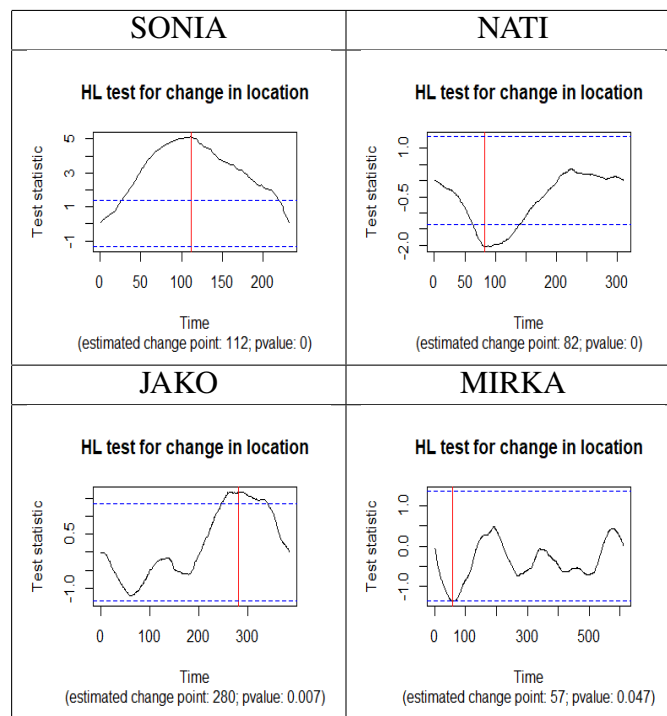


Figura 10.4: Estimador de cambio de locación. La línea roja indica el máximo en valor absoluto. Las líneas punteadas son bandas de confianza.

⁷Se utilizó el paquete *robs* de R

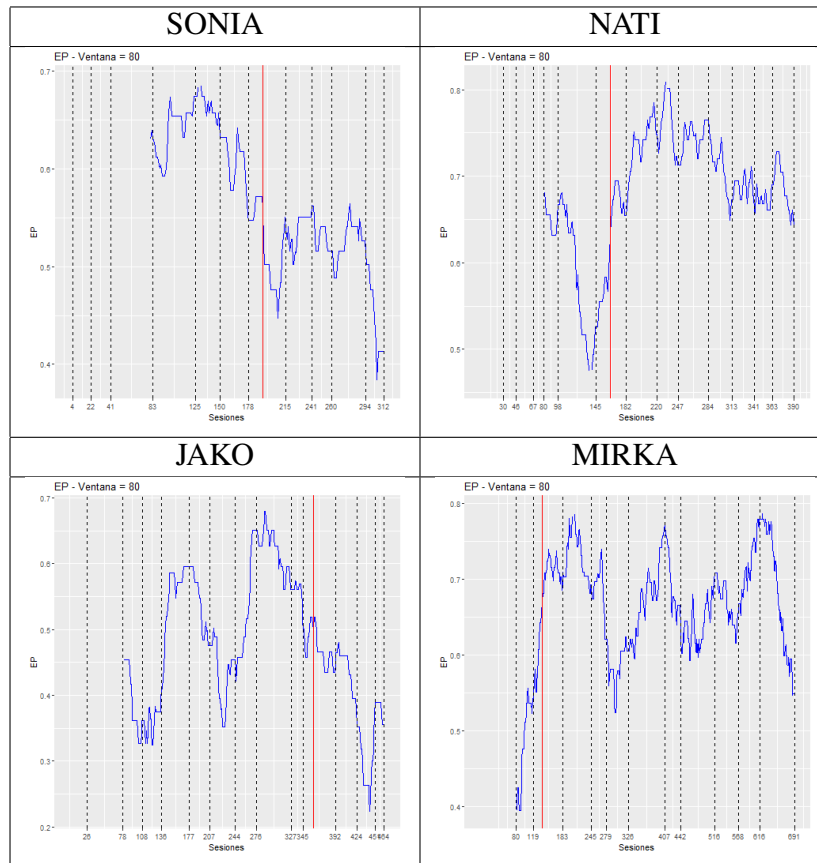


Figura 10.5: Ventana móvil de $H_s(D = 3)$ para cada aprendiente. Las líneas punteadas marcan separación entre sesiones y la roja el lugar de la serie donde se estima el cambio en la locación.

10.4. Análisis cuantitativo de recurrencias (RQA)

10.4.1. Gráficos de recurrencias.

Recurrence Quantification Analysis o *RQA*, por sus siglas en inglés (Eckmann et al., 1987; Webber & Zbilut, 2005; Marwan et al., 2007; Webber & Marwan, 2015; Webber, Ioana & Marwan, 2016; Huffaker et al., 2017) es una técnica que permite medir propiedades de atractores en el espacio de fase [phase space] reconstruido. Los atractores se definen en términos de regiones en el espacio (multi-)dimensional que son visitadas una y otra vez por el sistema dinámico. *RQA* mide el grado de nuevas visitas del sistema en dichas regiones y computa ciertas medidas basadas en dichas recurrencias. El método fue introducido por Eckmann et al. (1987). Supóngase una trayectoria de vectores en el espacio de fase. Los componentes de estos vectores son parámetros. La evolución del sistema esta representada por una serie de estos vectores, que representan su trayectoria. Sin embargo, en los datos reales por lo general no todos los parámetros necesarios para construir el espacio de fase pueden ser medidos; es decir que a menudo se tiene a disposición solamente un observable medido de modo discreto. Pero, si el sistema depende de más parámetros, o sea, si existe en un espacio multi-dimensional, ¿cómo reconstruirlo a partir de una sola variable observada? La solución yace en el hecho de que todas las variables de un sistema dinámico determinístico están relacionadas. Por ende, si en el tiempo τ solamente se conoce el valor de una variable x , entonces otra medida de dicha variable en $t + \tau$ (τ representa el *embedding delay*) portará implícitamente información sobre otras variables del sistema. Entonces, conociendo τ si se deben obtener, por ejemplo, tres variables que actúen en el sistema, es posible reconstruir un vector tridimensional $(x(t), x(t + \tau), x(t + 2\tau))$. Formalmente, esto es el teorema de *Taken*: $y_t = (x(t), x(t + \tau), x(t + 2\tau), \dots, x(t + (m - 1)\tau))$; donde m es la dimensión y τ , el retraso. Cada y_t es un punto en el espacio de fase reconstruido. Luego se calculan las distancias entre dichos puntos (entre los puntos en $t = i$ y $t = j$): $\delta_{ij} = \|y_i - y_j\|$. Se pueden usar diferentes normas para la distancia: *Euclídea*, *Manhattan*, etc. Dos puntos se declaran recurrentes si la distancia calculada es menor a un cierto

radio ε : $|y_i - y_j| < \varepsilon$. Entonces un gráfico de recurrencia [*RP*, *recurrence plot*, por sus siglas en inglés] se basa en las siguientes condiciones en una matriz de recurrencias:

$$R_{i,j} = \Theta(\varepsilon - \|y_i - y_j\|); y_i \in \mathbb{R}^m; i, j = 1, \dots, N.$$

donde: N es el número de estados de la trayectoria y_i , ε es el radio de recurrencia, $\|\cdot\|$ es una norma para calcular la distancia y Θ es la función *Heaviside*:
$$\begin{cases} \Theta(x) = 0 & x > \varepsilon \\ \Theta(x) = 1 & x < \varepsilon \end{cases}$$

Por lo tanto, para todos los estados recurrentes (todos los y_i que se encuentran en el vecindario definido por el radio ε), $R_{ij} = 1$, y se representa mediante un punto oscuro; alternativamente el estado no recurre y $R_{ij} = 0$, representado por un punto blanco. Ya que un estado siempre recurre con sí mismo porque la distancia es cero, $R_{ii} = 1$, la diagonal de la matriz se llama línea de identidad [*line of identity*, *LOI*, por sus siglas en inglés]. La matriz es simétrica respecto a dicha diagonal: $R_{ij} = R_{ji}$. El siguiente ejemplo se tomó de Webber & Zbilut (2005) para ilustrar el procedimiento. Sea el vector de datos:

$$x = \begin{pmatrix} 3,7;9,2;2,1;-5,4;0,0;-10,9;9,2;3,1;1,7;1,8;-0,3;-4,9;2,7;3,5 \\ 7,5;-9,9;-9,9;-4,7;1,3;2,7;7,6;3,9;7,3;8,0;0,3;-1,9;5,1;8,8;8,2 \end{pmatrix}$$

y supóngase $\tau = 8$ y $m = 4$; es decir que se toman datos cada ocho pasos para formar vectores de cuatro dimensiones. Es decir que:

$$y_i = (x(t), x(t+8), x(t+2(8)), x(t+3(8))) = (x(t), x(t+8), x(t+16), x(t+24))$$

por lo tanto los y_i serán: $y_1 = (3,7;1,7;-9,9;0,3)$; $y_2 = (9,2;1,8;-4,7;-1,9)$; $y_3 = (2,1;-0,3;1,3;5,1)$; $y_4 = (-5,4;-4,9;2,7;8,8)$; $y_5 = (0,0;2,7;7,6;8,2)$. Supóngase además la métrica euclídea: $\|y_4 - y_5\|_2 = \sqrt{\sum (y_i - y_j)^2}$. Por ejemplo, tómense los vectores y_4 e y_5 :

$$\|y_4 - y_5\|_2 = \sqrt{((-5,4-0,0)^2 + (-4,9-2,7)^2 + (2,7-7,6)^2 + (8,8-8,2)^2)} = 10,549.$$

La matriz de distancias D_{ij} será:

$$D = \begin{bmatrix} 0 & 7,882893 & 12,451506 & 8,904497 & 19,579326 \\ 7,882893 & 0 & 11,824551 & 20,671236 & 18,405162 \\ 12,451506 & 11,824551 & 0 & 9,646761 & 7,918965 \\ 18,904497 & 20,671236 & 9,646761 & 0 & 10,549408 \\ 19,579326 & 18,405162 & 7,918965 & 10,549408 & 0 \end{bmatrix}$$

Optativamente es posible re-escalar las distancias: (a) dividiendo cada entrada de la matriz por la distancia media de la matriz de distancias: $\frac{d_{ij}}{\frac{1}{N^2} \sum D_{ij}}$; (b) dividiendo cada entrada de la matriz por la distancia máxima de la matriz de distancias: $\frac{d_{ij}}{\max\{D_{ij}\}}$. En este caso, si re-escalamos usando (b): $d_{4,5}/\max\{D_{ij}\} = 10,549408/20,671236 = 0,510342$. La matriz de distancias re-escalada queda:

$$D_{re-escalada} = \begin{bmatrix} 0 & 0,381346 & 0,602359 & 0,914531 & 0,947177 \\ 0,381346 & 0 & 0,572029 & 1 & 0,890375 \\ 0,602359 & 0,572029 & 0 & 0,466675 & 0,383091 \\ 0,914531 & 1 & 0,466675 & 0 & 0,510342 \\ 0,947177 & 0,890375 & 0,383091 & 0,510342 & 0 \end{bmatrix}$$

Por último, si se establece $\varepsilon = 8$ (re-escalado: $\varepsilon = 8/20,671236 = 0,3870112$), y aplicando la función *Heaviside*, la matriz de recurrencias R será:

$$R = \begin{bmatrix} 1 & 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 1 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 1 \end{bmatrix}$$

Por lo tanto existen *dos* recurrencias: $y_{1,2} = y_{3,5} = 1$, sin contar la línea de identidad ni las entradas debajo de ésta, ya que son valores simétricos.

Supongamos ahora que se tienen dos sistemas dinámicos, cada uno representado por las trayectorias x_i e y_i en el mismo espacio m -dimensional. El gráfico de recurrencia cruzada [*CRP*, *cross-recurrence plot*] es una extensión bivariada del *recurrence plot* que permite cuantificar comportamientos similares en la evolución de dichas trayectorias. En este caso las distancias se computan entre ambas trayectorias. Luego el gráfico se construye en modo análogo al anterior:

$$R_{i,j}^{x,y} = \Theta(\varepsilon - \|x_i - y_j\|); x_i, y_i \in \mathbb{R}^m; i = 1, \dots, N; j = 1, \dots, M.$$

El gráfico de recurrencia ilustra ahora cuándo un estado de un sistema “recurrer” con un estado del otro sistema. La diagonal principal se llama “línea de sincronización” [*LOS*, *line of synchronization*], en la cual no necesariamente los sistemas visitan los mismos estados.

10.4.2. Medidas a partir de recurrencias.

Se han propuesto varias medidas para cuantificar la matriz de recurrencias (Webber & Zbilut, 2005; Marwan et al., 2007; Webber & Marwan, 2015; Webber, Ioana & Marwan, 2016).

Recurrencia [*Recurrence Rate*; *RR*]. Porcentaje de puntos que recurren en la matriz R , incluyendo o no a la *LOI* (diagonal principal). En el caso del *CRP* debe estar la diagonal principal si se desea investigar el grado de sincronización entre ambos sistemas.

$$RR = \frac{1}{N^2} \sum_{i,j=1}^N R_{i,j}$$

Determinismo [*Determinism*, *DET*]. Porcentaje de puntos recurrentes que forman líneas diagonales, incluyendo o no a la *LOI*:

$$DET = \frac{\sum_{l=d_{min}}^N lP_D(l)}{\sum_{l=1}^N lP_D(l)}$$

Donde: d_{min} es la longitud mínima para que la secuencia sea considerada como línea diagonal en el conteo, l es la longitud de cada línea y $P(l)$ es el *histograma* de frecuencias absolutas de las líneas diagonales. Para $d_{min} = 1$, $DET = 1$. DET mide la predictibilidad del sistema: cuanto más alto, más se parece a un sistema periódico. En sistemas caóticos las diagonales son cortas; en sistemas periódicos, largas; y en el ruido aleatorio no existirían en principio, salvo algunas de corta longitud que aparecen por azar. Observar que: $\sum_{l=1}^N lP_D(l) = \sum_{i,j=1}^N R_{i,j}$.

RATIO. Es el ratio entre DET y RR . Se ha sugerido (Marwan et al, 2007) que para cierto tipo de transiciones la RR se mantiene constante y la DET baja, haciendo que el $RATIO$ suba.

$$RATIO = \frac{DET}{RR} = N^2 \frac{\sum_{l=d_{min}}^N lP_D(l)}{\left(\sum_{l=1}^N lP_D(l)\right)^2}$$

Línea diagonal máxima: D_{max} [*maximal diagonal line length*]. Es simplemente la longitud máxima de las líneas diagonales: $D_{max} = \arg \max_l P_D(l)$. Una línea diagonal de largo l indica que un segmento de una trayectoria (o sistema) está cerca durante l pasos de tiempo de otro segmento de la misma trayectoria (o de la trayectoria del otro sistema en CRP) que se encuentra en un tiempo diferente. Por ende, cuánta más divergencia, más cortas serán las líneas diagonales. Su inversa ($\frac{1}{L_{max}}$) es la medida llamada *DIV*, la cual cuantifica dicha divergencia.

Línea diagonal promedio: $\langle D \rangle$ [*average diagonal line length*]. Es la longitud media de las líneas diagonales. Es el tiempo promedio en el cual dos segmentos de una trayectoria están cerca uno del otro. Se interpreta como tiempo promedio de predicción.

$$\langle D \rangle = \frac{\sum_{l=d_{min}}^N l P_D(l)}{\sum_{l=d_{min}}^N P_D(l)}$$

ENT [Entropy]. Es la entropía de *Shannon* respecto de la *distribución* de las longitudes de las líneas diagonales. Es una medida de complejidad de las estructuras diagonales (determinismo). Cuanto más baja menos complejidad.

$$ENT = - \sum_{l=l_{min}}^N P(l) \ln [P(l)]; P(l) = \frac{P_D(l)}{\sum_{l=d_{min}}^N P_D(l)}$$

TND [Trend]. Es el coeficiente de correlación respecto de la densidad de recurrencia de las diagonales paralelas a la *LOI*. Constituye una medida de estacionariedad del sistema. Una dinámica cuasi-estacionaria estará indicada por un valor de TND cercano a cero. Valores lejanos al cero indicarían dinámica no estacionaria y posiblemente “drift”, mostrando que el sistema se acerca paulatinamente a estados más estacionarios.

$$TND = \frac{\sum_{i=1}^{\tilde{N}} (i - \tilde{N}/2) (RR_i - \langle RR \rangle)}{\sum_{i=1}^{\tilde{N}} (i - \tilde{N}/2)^2}$$

LAM [Laminarity]. Análogamente a *DET*, es el ratio entre los puntos recurrentes que forman parte de líneas verticales u horizontales y el número total de puntos recurrentes. El análisis separado de líneas horizontales y verticales cobra especial relevancia en el *CRP* ya que se observan comportamientos en dos sistemas diferentes. Se define como:

$$LAM_V = \frac{\sum_{l=v_{min}}^N l P_V(l)}{\sum_{l=1}^N l P_V(l)}; LAM_H = \frac{\sum_{l=h_{min}}^N l P_H(l)}{\sum_{l=1}^N l P_H(l)}$$

Donde: $v_{min}; h_{min}$ es la longitud mínima para que la secuencia sea considerada como línea vertical / horizontal en el conteo, l es la longitud de cada línea y $P_V(l); P_H(l)$ es el *histograma* de frecuencias absolutas de las líneas verticales / horizontales. Está relacionada con periodos de *laminaridad* del sistema, es decir, con segmentos de tiempo en los cuales no hay intermitencia. La laminaridad indica por cuánto tiempo el sistema está “atrapado” expresando recurrencias, o sea en un mismo estado, desde el punto de vista del (sub)sistema vertical (en el eje vertical) o del (sub)sistema horizontal (en el eje horizontal).

Línea horizontal (vertical) máxima: V_{max} y H_{max} [*maximal vertical (horizontal) line length*]. Es la longitud máxima de las líneas verticales u horizontales: $V_{max} = \arg \max_l P_V(l); H_{max} = \arg \max_l P_H(l)$. Mide la duración máxima de recurrencias en el (sub)sistema vertical u horizontal.

Línea horizontal (vertical) promedio: TT_V y TT_H [trapping time]. Es la longitud media de las líneas verticales u horizontales. Indican en tiempo promedio en el cual el (sub)sistema vertical u horizontal se encuentra “atrapado” en un mismo estado de recurrencias. Si TT_V es alto, el sistema del eje vertical tiende a estar atrapado por tiempos prolongados en recurrencias con puntos del sistema en el eje horizontal en tiempos diferentes. Análogamente para TT_H alto el sistema en el eje horizontal está atrapado en el mismo comportamiento recurrente que coincide con el otro sistema en algún punto diferente de tiempo.

$$TT_V = \frac{\sum_{l=v_{min}}^N lP_V(l)}{\sum_{l=v_{min}}^N P_V(l)} ; TT_H = \frac{\sum_{l=h_{min}}^N lP_H(l)}{\sum_{l=h_{min}}^N P_H(l)}$$

El siguiente gráfico ilustra el gráfico de recurrencias para sesiones de SONIA 2 y 12. En el gráfico se traza una línea donde hay un error. El “Array1” (serie del primer término de concordancia) esta en el eje horizontal y el “Array2” (serie del segundo término de concordancia), en el vertical. Las franjas negras son “1” en la matriz de recurrencia; en cambio las blancas son “0”. Dado que en el “Array1” hay tres errores, hay tres franjas blancas de ceros y en el “Array 2” hay cuatro. Obsérvese que en el primer gráfico hacia arriba hay una “franja doble” en el “Array2” ya que se trata de dos errores seguidos. Las líneas horizontales (verticales) de ceros significan que hay un error (falta de recurrencia) en el término i-ésimo del Array en cuestión, y que dicho término i-ésimo tampoco recurre con todos los términos j-ésimos del otro Array. La línea diagonal es la llamada “línea de sincronización”. En la sesión 12 solamente se comete un error en el undécimo término del Array1. Entonces solamente hay una franja vertical. Es fácil ver que un alumno sin errores tendría el gráfico completamente pintado de negro. Para datos categóricos los gráficos tipo “mosaico”, como en el ejemplo, aparecen comúnmente.

datos (sesión 2): Array1 : 0 3 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0
 Array2 : 0 0 0 0 0 0 0 0 4,1 0 0 0 0 0 0 3,1 2,1 0 3,1

datos (sesión 12): Array1 : 0 0 0 0 0 0 0 0 0 0 0 3 0 0 0 0 0 0
 Array2 : 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

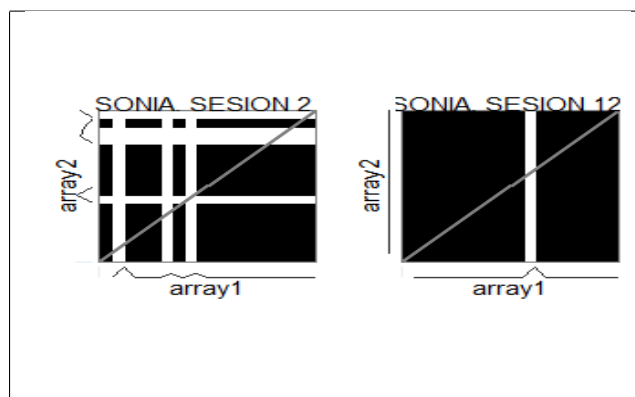


Figura 10.6: Gráficos de recurrencia para SONIA, sesiones 2 y 12.

El Cuadro 4 muestra las frecuencias absolutas de las líneas, necesarias para calcular las medidas que derivan de ellas.

Por último, el Cuadro 5, que le sigue, ilustra el cálculo de las medidas.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
SESIÓN 2	VERTICALES	28	0	14	0	0	0	0	0	0	14	0	0	0	0	0	0	0	0
	HORIZONTALES	15	0	0	0	0	15	15	0	0	0	0	0	0	0	0	0	0	0
	DIAGONALES	28	8	12	6	6	14	8	0	0	0	0	0	0	0	0	0	0	0
SESIÓN 12	VERTICALES	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	17
	HORIZONTALES	0	0	0	0	0	0	18	0	0	18	0	0	0	0	0	0	0	0
	DIAGONALES	4	4	4	4	4	4	26	2	2	2	0	0	0	0	0	0	0	0

Cuadro 10.9: Tabla de frecuencias de líneas diagonales, horizontales, verticales para los ejemplos. Las columnas son el largo de las líneas y las filas, las frecuencias absolutas.

	SESIÓN 2	SESIÓN 12
DET	$\frac{3(12)+4(6)+5(6)+6(14)+7(8)}{1(28)+2(8)+3(12)+4(6)+5(6)+6(14)+7(8)} = 0,83$	$\frac{3(4)+4(4)+5(4)+6(4)+7(26)+8(2)+9(2)+10(2)}{1(4)+2(4)+3(4)+4(4)+5(4)+6(4)+7(26)+8(2)+9(2)+10(2)} = 0,96$
$\langle D \rangle$	$\frac{3(12)+4(6)+5(6)+6(14)+7(8)}{12+6+6+14+8} = 5$	$\frac{3(4)+4(4)+5(4)+6(4)+7(26)+8(2)+9(2)+10(2)}{4+4+4+4+26+2+2+2} = 6,41$
LAM_V	$\frac{6(15)+7(15)}{1(15)+6(15)+7(15)} = 0,92$	$\frac{18(17)}{18(17)} = 1$
LAM_H	$\frac{3(14)+10(14)}{1(28)+3(14)+10(14)} = 0,86$	$\frac{7(18)+10(18)}{7(18)+10(18)} = 1$
TT_V	$\frac{3(14)+10(14)}{15+15} = 6,5$	$\frac{18(17)}{17} = 18$
TT_H	$\frac{6(15)+7(15)}{14+14} = 6,5$	$\frac{7(17)+10(18)}{18+18} = 8,5$

Cuadro 10.10: Ilustración del cálculo de las medidas de recurrencia

10.4.3. Significancia de las medidas.

Una posible pregunta sería si el valor de las medidas obtenidas es significativamente diferente de aquellas obtenidas provenientes de una matriz de recurrencias calculada a partir de una serie de ruido blanco (sin estructura). Es decir, examinar si las medidas son realmente informativas. Para evaluar esto se puede llevar a cabo un test de permutaciones (Webber & Marwan, 2015): (i) mezclar las filas y las columnas de la matriz de recurrencias (esto destruye las estructuras, o sea, los tipos de líneas presentes de la matriz original); (ii) calcular la medida de recurrencia; (iii) repetir el procedimiento B veces; (iv) a partir de la distribución empírica de los B valores de la medida de recurrencia obtenidos calcular los percentiles $\frac{\alpha}{2}$ y $1 - \frac{\alpha}{2}$; (v) definir un intervalo de confianza del $100 - \alpha\%$ como el rango entre dichos percentiles; (vi) declarar a la medida significativa si se encuentra fuera del intervalo.

Otra cuestión a considerar sería si existe una diferencia significativa entre dos valores de la misma medida provenientes de matrices de recurrencias calculadas sobre series sujetas a tratamientos diferentes. Schinkel et. al. (2009) sugieren el siguiente procedimiento, solamente aplicable a medidas derivadas de líneas (diagonales, verticales u horizontales) de la matriz de recurrencias. Es como sigue: (i) dados los histogramas de las líneas diagonales ($P_D(l)$), horizontales ($P_H(l)$) y verticales ($P_V(l)$) [incluyendo las líneas de un solo punto recurrente], re-muestrear con re-emplazo a partir de éstos para obtener otro histograma $P_D(l)^{*,1}$ [o bien: $P_H(l)^{*,1}$, $P_V(l)^{*,1}$]; (ii) A partir de los histogramas de re-muestreo calcular las medidas, por ejemplo $DET^{*,1}$; (iii) repetir la operación anterior B veces; (iv) a partir de la distribución empírica de los B valores de la medida de recurrencia obtenidos, calcular los percentiles $\frac{\alpha}{2}$ y $1 - \frac{\alpha}{2}$; (v) definir un intervalo de confianza del $100 - \alpha\%$ como el rango entre dichos percentiles. Observar que, a diferencia del caso anterior este procedimiento preserva las estructuras de la matriz original de recurrencias. Si se calcula dicho intervalo para las medidas sujetas a dos tratamientos diferentes, se declarará una diferencia significativa si sus intervalos de confianza *no se superponen*. En ambos procedimientos es posible correr una ventana móvil en los datos y calcular el intervalo puntual para cada ventana.

Por último, cabría preguntarse sobre la confianza con la cual una determinada medida logra detectar una transición en la dinámica del sistema o un desvío de la misma. Pero en este caso es necesario hacer correr una ventana móvil (con o sin superposición) y calcular las medidas para cada ventana, precisamente para ser capaz de detectar dichos cambios. Marwan et al. (2013) sugirieron el siguiente método, para rechazar la

hipótesis nula H_0 : “la dinámica del sistema no cambia en el tiempo”. Por ende, la estructura de la matriz de recurrencias se mantiene relativamente constante. Lo cual implica que la medida derivada de recurrencia M tiene una distribución con media $\mu(M)$ y varianza $\sigma(M)$ no nulas aunque desconocidas. Como en el caso anterior, solamente es aplicable a medidas que dependan de líneas (diagonales, horizontales, verticales). Se parte nuevamente de los histogramas de frecuencias de las líneas $P(l)$. Habrá uno para cada ventana. Se los combina a todos para formar un histograma global de líneas $P_{global}(l)$. Luego se lleva a cabo un re-muestreo a partir de dicho histograma global para crear un nuevo histograma $P^*(l)$ y se calcula la medida deseada M^* . Se repite el procedimiento B veces con el fin de obtener una distribución empírica de la medida y derivar los percentiles los percentiles $\frac{\alpha}{2}$ y $1 - \frac{\alpha}{2}$. Como antes, se define un intervalo de confianza del $100 - \alpha\%$ como el rango entre dichos percentiles. La ventana se declara de dinámica diferente si el valor de la medida se ubica *fuera* del intervalo de confianza. Notar que la correspondiente medida se asigna a la ventana $w[i]$; con lo cual, si se detecta allí un cambio de dinámica, este habrá ocurrido presumiblemente en los datos originales entre los tiempos i e $i + w - (m - 1)\tau$ (si $m = 1$, $\tau = 1$ el intervalo es $[i, i + w]$).

10.4.4. Cambios de dinámica en las sesiones.

Para cada alumno, se creó una serie temporal bivariada con categorías por tipo de error en el alfabeto usual $A = \{0, 1, 2, 3, 4\}$, discriminado por término de concordancia. Por ejemplo, la instancia *grande ciudades* [SONIA, sesión 2, línea 12] le corresponde el error $[3,0]$ ya que hay un error de plural en el primer término y ningún error en el segundo término. Es decir que se consideraron dos sub-sistemas dinámicos en proceso de sincronización. Como el interés radicó en evaluar la sincronización entre instancias correctas en ambos términos, se sumó una pequeña cantidad constante de 0,1 a los datos del segundo término que no tuvieran la categoría “0”. De este modo solamente se considerarían recurrentes los pares $[0,0]$ y no los pares de repetencia de error (por ejemplo $[3,3]$). Se llevó a cabo un análisis de recurrencia cruzada por alumno y para cada sesión, con los siguientes parámetros: $m = 1$, $\tau = 1$, $\varepsilon = 0,0001$. El largo mínimo para considerar como línea (diagonal, vertical, horizontal) fue $l = 3$. Para detectar posibles cambios de dinámica se utilizó el método descrito en Marwan et al. (2013), el cual se detalló en el apartado anterior. Se consideraron las medidas: DET [determinismo], $\langle D \rangle$ [línea diagonal promedio], LAM_V [laminaridad vertical], LAM_H [laminaridad horizontal], TT_V [“trapping time” vertical], TT_H [“trapping time” horizontal]⁸.

Los Cuadros que siguen (uno para cada alumno) muestran los resultados de las medidas de recurrencia según las sesiones. La línea punteada marca el límite superior del intervalo puntual de confianza del 95% [el número en negrita especifica el cuantil correspondiente en la distribución de *bootstrap*], más allá del cual se considera un cambio en la dinámica.

⁸Se utilizaron las librerías de R “casnet” y “crqa” para los cálculos.

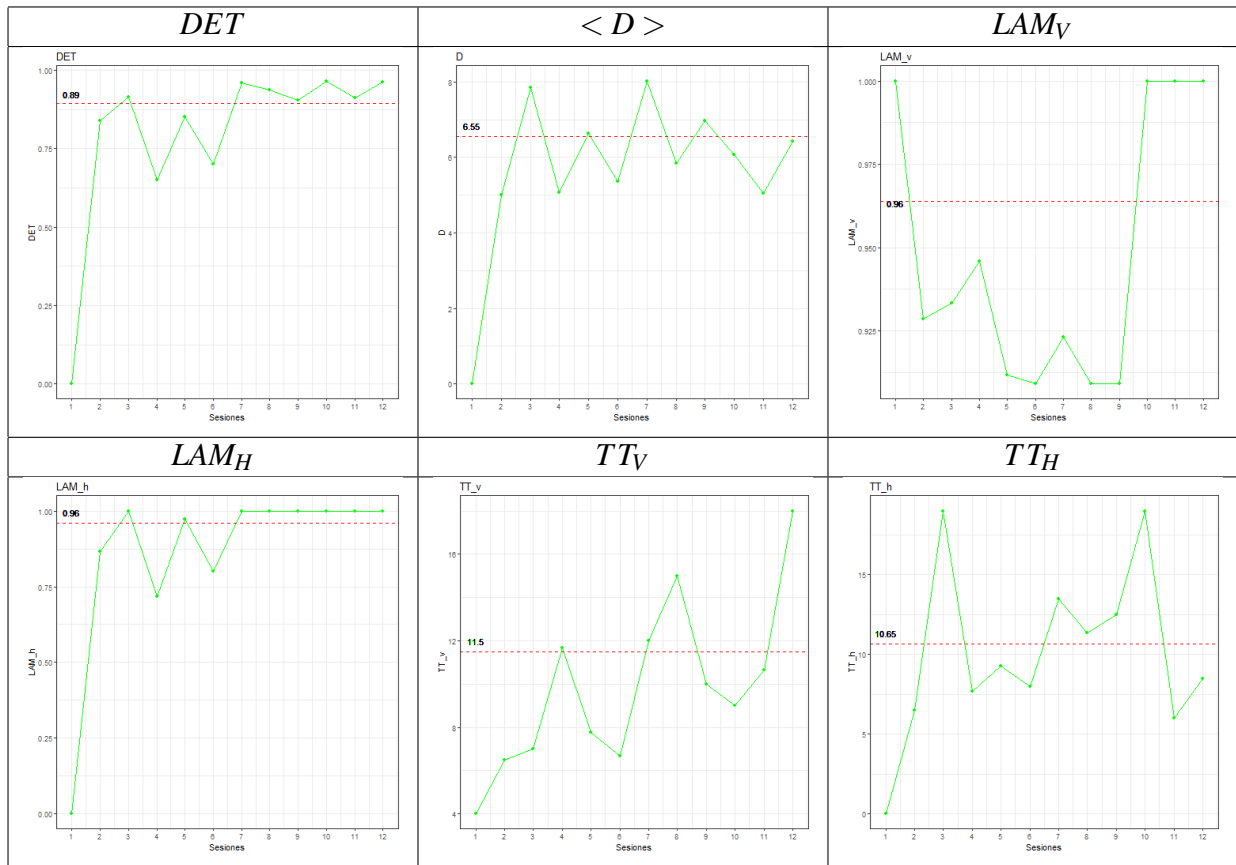


Figura 10.7: SONIA. Medidas de recurrencia.

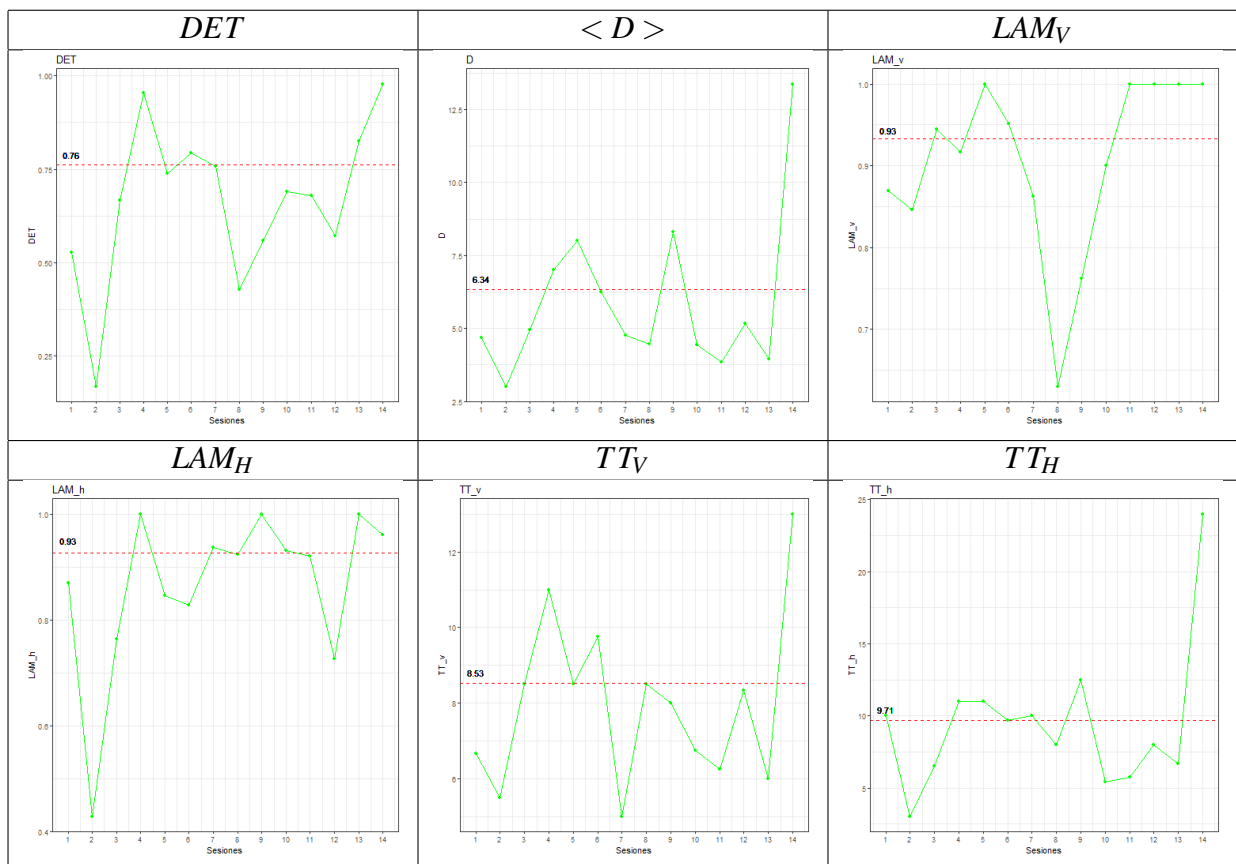


Figura 10.8: NATI. Medidas de recurrencia.

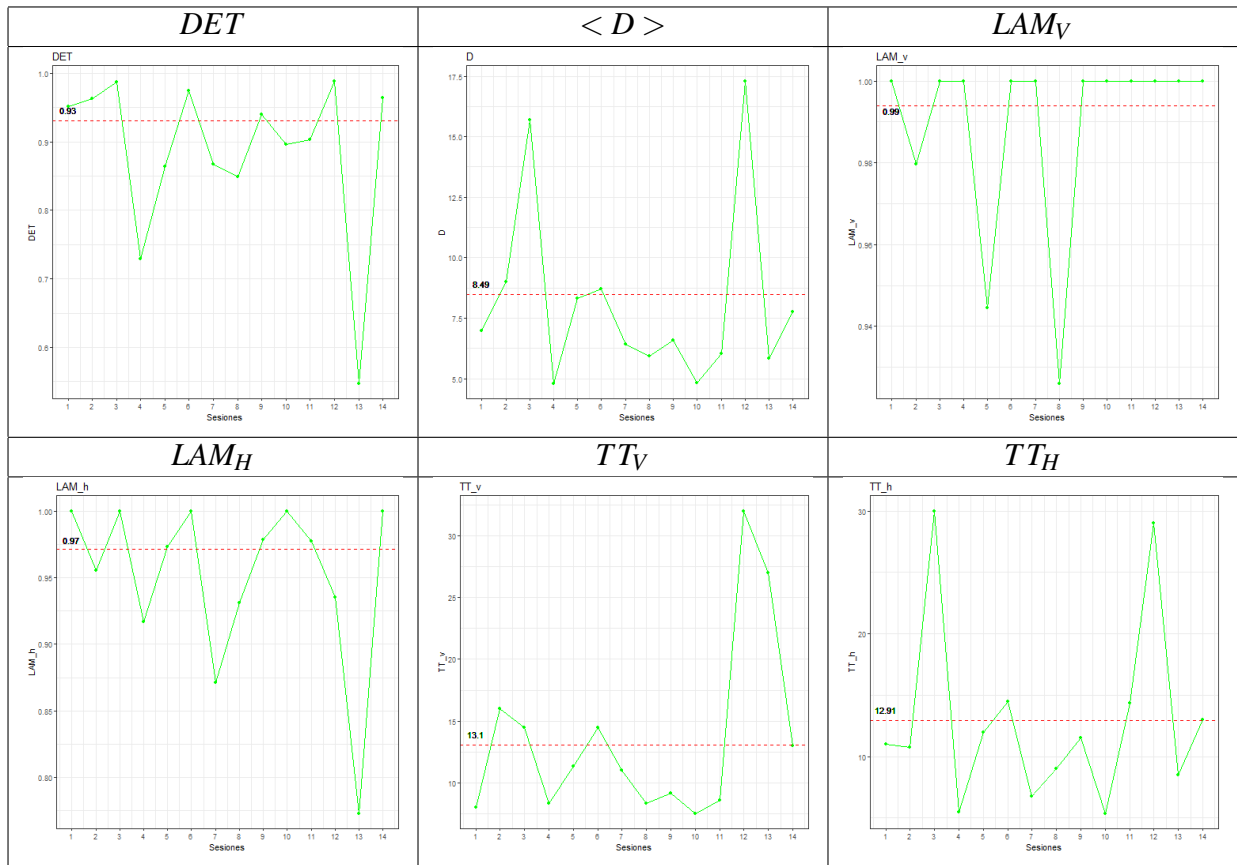


Figura 10.9: JAKO. Medidas de recurrencia.

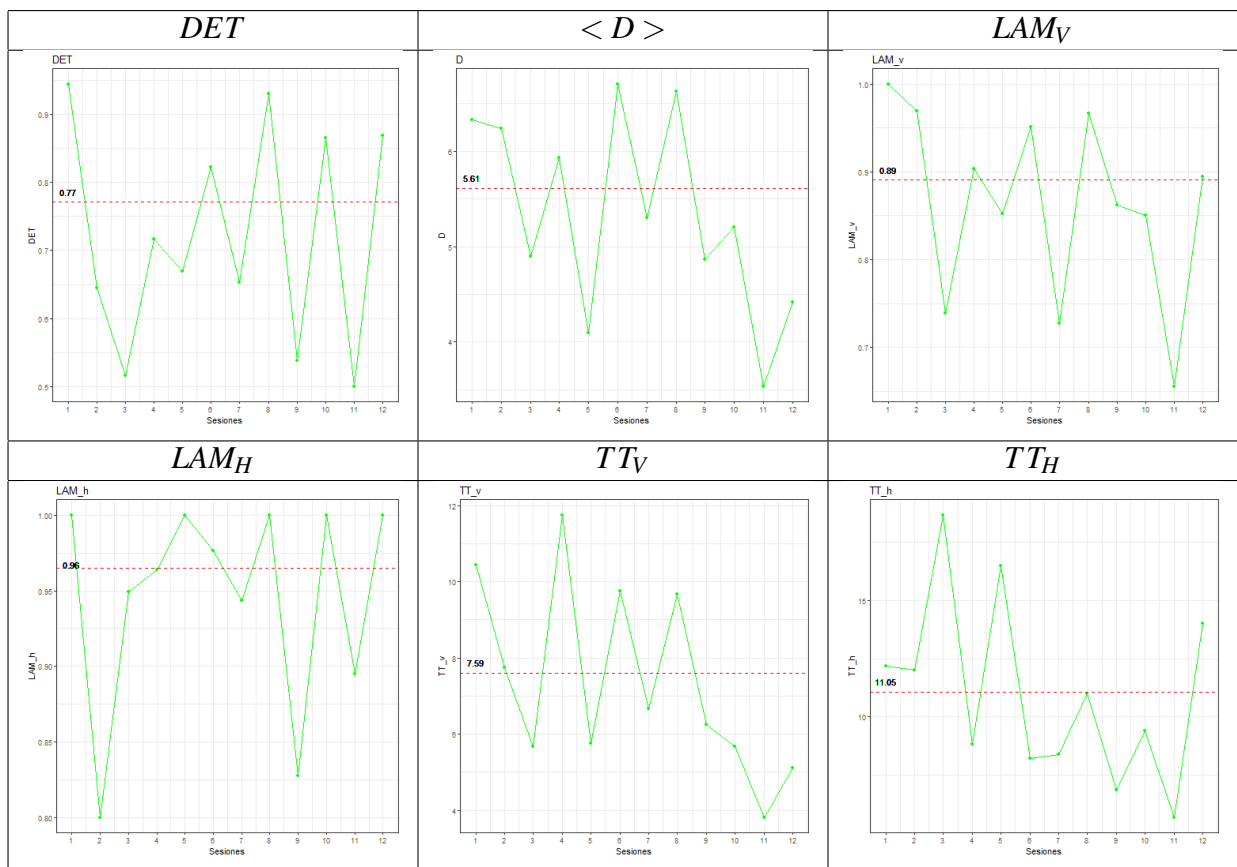


Figura 10.10: MIRKA. Medidas de recurrencia.

Lo primero que salta a la vista es que las medidas no necesariamente coinciden en cuanto al cambio de dinámica en una determinada sesión. Por lo tanto se adoptó una estrategia de “consenso”. Se discretizaron

las medidas en: (i) “1” = “cambia la dinámica” (arriba del límite superior del *I.C.* 95%), (ii) “0” = “no cambia la dinámica”. Luego se sumaron las medidas para cada sesión. Se declaró un cambio de dinámica si dicha suma resultaba ≥ 4 . El Cuadro que sigue muestra lo dicho. Según este criterio, SONIA cambia la dinámica en las sesiones 3,7,8,9,10,12; NATI, en las sesiones 4,14; JAKO en las sesiones 3,6,12,14; y MIRKA en las sesiones 1,2,6,8,12. Las medidas tienen valores más bajos para NATI y MIRKA, como es esperable. Vale la pena aclarar que un cambio en la dinámica del error *no* implica necesariamente que el alumno haga menos errores. Por ejemplo, en las sesiones 2 y 3 de MIRKA se comete un 33% de error, sin embargo la laminaridad y el “trapping time” verticales son mayores en la segunda sesión. Esto significa que el sistema del primer término de la concordancia se encuentra “atrapado” más tiempo en el atractor de lo “correcto”. Además es más alta la línea diagonal promedio, lo cual indica que en media segmentos de ambos términos de la concordancia están más cerca por más tiempo que en la sesión tres. En general la suma de las medidas posee una alta correlación negativa con las proporciones de error. Según el test no paramétrico de *Spearman*: (i) SONIA: $\rho_s = -0,803$ ($p < 0,01$); (ii) NATI: $\rho_s = -0,651$ ($p < 0,05$); (iii) JAKO: $\rho_s = -0,824$ ($p < 0,001$); (iv) MIRKA: $\rho_s = -0,894$ ($p < 0,0001$).

medidas	binario
Alumno: SONIA	
DET	001000111111
D	001010101000
LAM vertical	100000000111
LAM horizontal	001010111111
TT vertical	000100110001
TT horizontal	001000111100
SUMA:	104120544434
Alumno: NATI	
DET	00010100000011
D	00011000100001
LAM vertical	00101100001111
LAM horizontal	00010010110011
TT vertical	00010100000001
TT horizontal	10011010100001
SUMA:	10153320311136
Alumno: JAKO	
DET	11100100100101
D	01100100000100
LAM vertical	10110110111111
LAM horizontal	10101100111001
TT vertical	01100100000110
TT horizontal	00100100001101
SUMA:	33611610323524
Alumno: MIRKA	
DET	100001010101
D	110101010000
LAM vertical	110101010001
LAM horizontal	100011010101
TT vertical	110101010000
TT horizontal	111010000001
SUMA:	641325050204

Cuadro 10.11: Discretización y suma de medidas de recurrencia

10.4.5. Recurrencias sobre instancias correctas y erróneas.

En esta sección se usaron dos series de tiempo (para cada término de la concordancia) con el alfabeto $A = \{0, 1\}$, “1” = error; “0” = correcto. El Cuadro que sigue muestra, para cross-recurrencia⁹, curvas de tasa de recurrencia entre: (i) instancias correctas: recurrencia en los pares [0 0]; (ii) instancia errónea en un par: recurrencia en los pares [0 1] o [1 0]; (iii) recurrencias entre errores: recurrencia en los pares [1 1]. Se observa que NATI conlleva la tasa más alta de recurrencia en errores [0 1] / [1 0], aunque cayendo en las últimas dos sesiones. La tasa de recurrencia en el error [1 1] es alta al principio y en la sesión ocho para NATI pero para MIRKA se mantiene al menos en un 0.5% en cuatro sesiones.

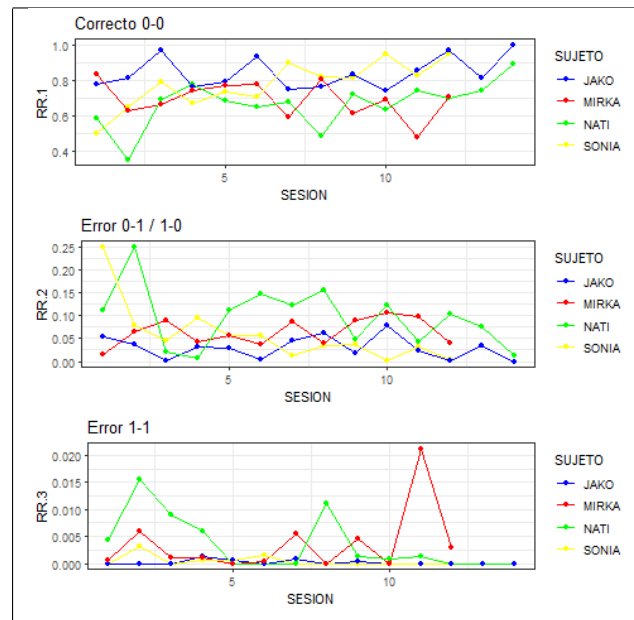


Figura 10.11: Recurrencias entre [0 0], [0 1] / [1 0], [1 1].

10.5. Indicadores de cambio de régimen.

En el ámbito de las ciencias del clima, Dakos et. al. (2012) propusieron una serie de indicadores capaces de anunciar transiciones en sistemas con “tipping points”; es decir, umbrales de bifurcación pasados los cuales se produce un cambio de régimen en el sistema a través de una transición. Lejos del “basin of attraction” del atractor alternativo al cual se cambiará el régimen la tasa de recuperación frente a perturbaciones es alta (por lo tanto el sistema tarda poco en volver cerca del equilibrio). En cambio cerca de éste la tasa de recuperación es baja (el sistema tarda en recuperarse de las perturbaciones para volver al atractor). Si el acercamiento al punto de bifurcación es *lento*, entonces se ha visto que cerca de éste la auto-correlación y la varianza aumentan (Sheffer et al., 2009). A continuación se describen algunos de los indicadores posibles de anunciar transiciones.

1. Auto-correlación, varianza. Cerca de la transición la tasa de retorno al equilibrio disminuye, entonces observaciones sucesivas se parecen más, aumentando así la auto-correlación en el *lag-1*. Esta se estima mediante el primer coeficiente de un modelo auto-regresivo (ajustado por cuadrados mínimos): $x_t = x_{t-1} + \varepsilon; \varepsilon \sim N(0, \sigma_\varepsilon^2)$. Por otra parte, las perturbaciones pueden llevar al sistema lejos del atractor y por tanto se espera que la varianza aumente. Esto se ha visto cuando la disminución de la tasa de retorno es lenta a medida que se está más cerca de la bifurcación (“critical slowing down”). Se ha observado también que la varianza aumenta (pero la correlación disminuye) en el fenómeno de *flickering*, en el cual fuertes perturbaciones pueden arrastrar al sistema hacia el atractor alternativo para volver luego al atractor anterior. Este ir y venir entre “basins of attractions” alternativos se sucede hasta la transición, luego de la cual

⁹Se usó la librería *crqa* de R.

el sistema se establece en el atractor alternativo estable (Wang et al., 2012). Además, la varianza y la correlación pueden aumentar en presencia de transiciones abruptas y cíclicas entre dos atractores inestables. Por otra parte, tanto en presencia de eventos extremos que llevan al sistema lejos del atractor pero que luego vuelven al atractor lentamente; como con cambios abruptos de un atractor a otro sin mediar una transición, no se ha visto que estos indicadores marquen cambio de régimen (Dakos et al., 2015).

2. Heteroscedasticidad condicional. Este término significa que, en una serie de tiempo, la varianza de una variable aleatoria en un determinado tiempo depende de (es condicional a) la varianza de la variable aleatoria en el (los) tiempo(s) anterior(es). Esto causa que períodos de alta varianza sigan a otros alta varianza; y, al contrario, períodos de baja varianza sigan a otros de baja varianza. Cerca de la transición se observará pues un *cluster* de alta varianza y lejos de ésta uno de varianza baja (Seekell et al., 2011). El fenómeno se detecta mediante un test que procede como sigue: (i) Ajustar un modelo auto-regresivo de orden n (seleccionado por ejemplo mediante el criterio de información de *Akaike*); (ii) tomar los residuos y elevarlos al cuadrado; (iii) retrasar los residuos un paso y hacer una regresión entre éstos como predictor y los residuos del paso (ii) como variable respuesta; (iv) calcular el producto entre el coeficiente de determinación r^2 de la *regresión* del paso (iii) y el tamaño muestral n de dicha regresión: nr^2 ; (v) comparar este valor con una distribución $\chi^2_{(df=1)}$. Cerca de la transición los test rechazados tienden a acumularse.

3. Modelo de Drift-Diffusion-jump. Se presupone que los datos son generados por el siguiente modelo: $dx_t = f(x_t, \theta_t) dt + g(x_t, \theta_t) dW + dJ_t$; donde: x es la variable de estado, $f(\cdot)$ y $g(\cdot)$ denotan funciones, dW es ruido blanco y J es un proceso de salto (“jump process”). Los saltos son perturbaciones grandes de un paso temporal que no están correlacionadas entre sí. El sistema hace una transición en el valor del parámetro θ_c crítico. El primer término mide el *drift*: la tasa determinística de cambio local. El segundo término mide *diffusion*: perturbaciones pequeñas que ocurren en cada paso temporal. El tercer término mide la grandes perturbaciones que suceden esporádicamente. El modelo es aproximado mediante el modelo no paramétrico siguiente (Carpenter & Brock, 2011).

$$dx_t = \mu(x_{t-}, \theta_t) dt + \sigma_D(x_{t-}, \theta_t) dW + d\left(\sum_{n=1}^{N_t} Z_n\right)$$

Aquí: $\mu(x_{t-}, \theta_t) dt$ aproxima el cambio determinístico en la serie de tiempo, $\sigma_D(x_{t-}, \theta_t)$ aproxima el desvío típico local de las pequeñas perturbaciones y el tercer término aproxima las grandes perturbaciones locales en la serie temporal mediante la varianza $\sigma_Z^2(x, \theta_t)$. Además los saltos Z_n se distribuyen como normal: $Z_n(0, \sigma_Z^2(\theta_t))$. Además la probabilidad de que un salto arribe en un tiempo incremental pequeño dt es $\lambda(x_t, \theta_t)$. Las funciones se estiman por medio de regresión no paramétrica con *kernels* normales. En concreto, para n observaciones y paso temporal Δ en un período de tiempo $H = n\Delta$, se computan los momentos $j = 1, 2, 4, 6$ del estimador con *kernel* pesados como:

$$\hat{M}_n^j(x) = \frac{n \sum_{i=1}^{n-1} K\left[\frac{x_{i\Delta}-x}{h}\right] (x_{(i+1)\Delta} - x_{i\Delta})^j}{H \sum_{i=1}^{n-1} K\left[\frac{x_{i\Delta}-x}{h}\right]}$$

donde: $K\left[\frac{x_{i\Delta}-x}{h}\right]$ es el kernel gaussiano de amplitud de banda h . Obsérvese que se computa sobre las *primeras diferencias* de los datos. Las funciones se estiman como: (i) función de *drift* en cada x : $\hat{\mu}(x, \theta_t) \approx \hat{M}_T^1(x)$; (ii) varianza local total en cada x : $\hat{\sigma}_T(x, \theta_t) \approx \hat{M}_T^2(x)$; (iii) varianza del salto en cada x : $\hat{\sigma}_Z(x, \theta_t) \approx \frac{1}{5n} \sum_{i=1}^n \frac{\hat{M}_i^6(x_{i\Delta})}{\hat{M}_i^4(x_{i\Delta})}$; (iv) función de frecuencia de arribo de un salto en cada x : $\hat{\lambda}(x, \theta_t) \approx \frac{\hat{M}_T^4(x)}{3[\hat{\sigma}_Z^2(x, \theta_t)]^2}$; (v) la varianza de la difusión es la diferencia: $\hat{\sigma}_D^2(x, \theta_t) \approx \hat{\sigma}_T^2(x, \theta_t) - \hat{\lambda}(x, \theta_t) \hat{\sigma}_T(x, \theta_t)$. Además se puede calcular la “varianza condicional”, que es la diferencia entre el segundo momento y el cuadrado del primer momento: $\hat{S}_n(x_i; \Delta_n) \approx \{\hat{M}_n^2(x_i; \Delta_n)\} - \{\hat{M}_n^1(x_i; \Delta_n)\}^2$ [Dakos et. al., 2012].

4. Análisis de Potenciales. Se trata de una técnica para derivar el potencial de un sistema y revelar la cantidad de “pozos” o atractores (Lenton & Livina, 2016; Levina et al. 2011). Se asume que el potencial de una serie temporal puede aproximarse mediante la ecuación diferencial estocástica siguiente: $dZ = -\frac{dU}{dz}dt + \sigma dW$. La idea detrás del método es que los valores giran con pequeñas perturbaciones alrededor de un atractor, por tanto, en un histograma aparecerán todos agrupados en el mismo *bin*. Es decir que un potencial con un solo atractor se parecerá a una densidad de probabilidad con una sola moda y lados suaves. Entonces se hace, en primer lugar, un histograma de los datos. Luego, se saca la densidad de probabilidad empírica de modo no paramétrico mediante un kernel gaussiano. El potencial se estima como:

$$U(z) = -\frac{\sigma^2}{2} \log[p_d(z)]$$

donde $p_d(z)$ es la densidad de probabilidad empírica. A continuación se ajustan polinomios de grado *par* creciente (grado dos para un atractor, cuatro para dos, etc.) y el grado más alto antes de obtener un coeficiente principal negativo se toma como la cantidad de atractores del sistema.

10.5.1. Cambios de régimen mediante ventanas móviles para recurrencias.

Se calculó la tasa de recurrencia (RR) para cada alumno¹⁰, pasando sobre los datos una ventana móvil de ancho $w = 50$, corriéndola de a un paso. Sobre esta serie se calcularon los indicadores.

10.5.1.1. Coeficiente AR(1) y desvío típico.

En primera instancia, se llevó a cabo un *detrending* mediante un suavizado no paramétrico con kernel gaussiano. Sobre los residuos se calcularon: (i) la estimación del primer coeficiente de un modelo autorregresivo (ajustado por cuadrados mínimos); (ii) el desvío típico; de acuerdo a un valor de ancho de ventana móvil ws y de ancho de banda para el kernel h , expresado como porcentaje del total de los datos. Luego se calculó sobre las curvas obtenidas el estadístico de correlación no paramétrico *tau* de Kendall, con el fin de evaluar la tendencia. Como los valores de dicho estadístico dependen de los valores ws y h que se elijan, se practicó un análisis de sensibilidad variando los valores de ws de a 10 unidades y los de h , de a 20 unidades. Los datos se encuentran en el apéndice del capítulo. Usando esto como guía, se seleccionaron los siguientes valores para los parámetros: $ws = [SONIA = 44, NATI = 55, JAKO = 50, MIRKA = 44]$; $h = [SONIA = 28, NATI = 99, JAKO = 10, MIRKA = 4]$. Por ejemplo, para SONIA, se empleó un ancho de ventana del 44% de los datos y un ancho de banda del 28% de los datos. Con el fin de evaluar la significancia del estadístico *tau* se hizo un test de *surrogates* como está especificado en Dakos et al. (2012). A tal fin, se ajustó el mejor modelo $ARMA(p, q)$ [autoregressive movil average] elegido según el AIC de los residuales de un *detrending* con kernel gaussiano. Luego se generaron 2000 conjuntos de datos del mismo largo que el conjunto de datos original a partir del modelo $ARMA$. Para cada conjunto de datos simulado se calculó la tendencia (*tau*) del la ventana móvil de las medidas del coeficiente del modelo $AR(1)$ y $\hat{\sigma}$. A continuación, se calcularon los percentiles $\frac{\alpha}{2}$ y $1 - \frac{\alpha}{2}$ y se definió un intervalo de confianza del $100 - \alpha\%$ como el rango entre dichos percentiles. Se declaró al *tau* observado significativo si se encontraba *fuera* de dicho intervalo. Se estableció $\alpha = 0,05$.

Los Cuadros que siguen muestran, para cada alumno: (i) los datos de recurrencia (el suavizado practicado señalado como un curva de color rojo) y los residuos; (ii) las dos medidas indicadoras calculadas; (iii) los resultados del test de *surrogates*, las líneas punteadas indican el cuantil del 2.5% y 97.5%. Con respecto a SONIA se observa un fuerte aumento de la pendiente del coeficiente $AR(1)$ hasta el tiempo 165, luego la pendiente se hace más moderada. Esto se corresponde al período $[165, 165+50]$ entre las sesiones 7 y final de la sesión 8. Lo mismo se observa para el desvío típico. Sin embargo, y a diferencia del coeficiente $AR(1)$, éste no resulta significativo en el test de *surrogates*. En cuanto a NATI, ningún indicador observa tendencia significativa, con lo cual no hay indicios de transiciones. Para JAKO, ambos indicadores evidencian tendencia significativa aunque *negativa*. Es decir que la tasa de retorno es cada vez más rápida y hay cada vez menos variabilidad. Las series de indicadores parecen comenzar a caer a partir del tiempo

¹⁰Se usó la librería *crqa* de R.

250 (período [250, 300] en la serie original, entre las sesiones 8 y 9). O sea que, a partir de dicho tiempo el sistema *no* hace transiciones sino que profundiza cada vez más el atractor en el que ya se encuentra. En lo que atañe a MIRKA, ambos indicadores tienen pendiente positiva y con más pendiente hasta alrededor del tiempo 570. O sea que la transición se produce entre los tiempos [570, 570+50] de la serie original, lo cual coincide con el período entre la sesión 11 y principios de la sesión 12.

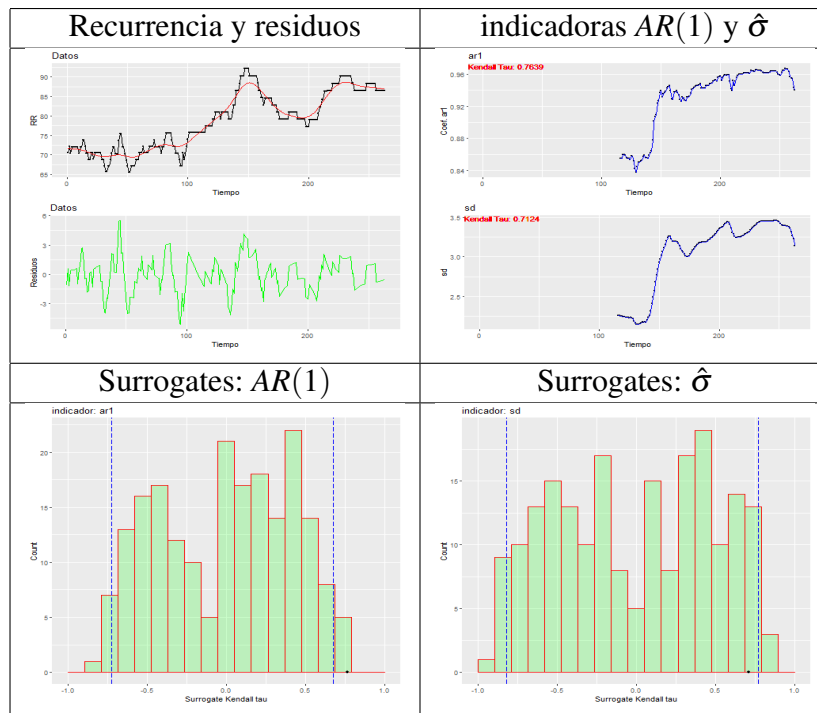


Figura 10.12: SONIA. Arriba izquierda: Recurrencia (detrending) y residuos; arriba derecha: Medidas indicadoras AR(1) y $\hat{\sigma}$; debajo: test de *surrogates*.

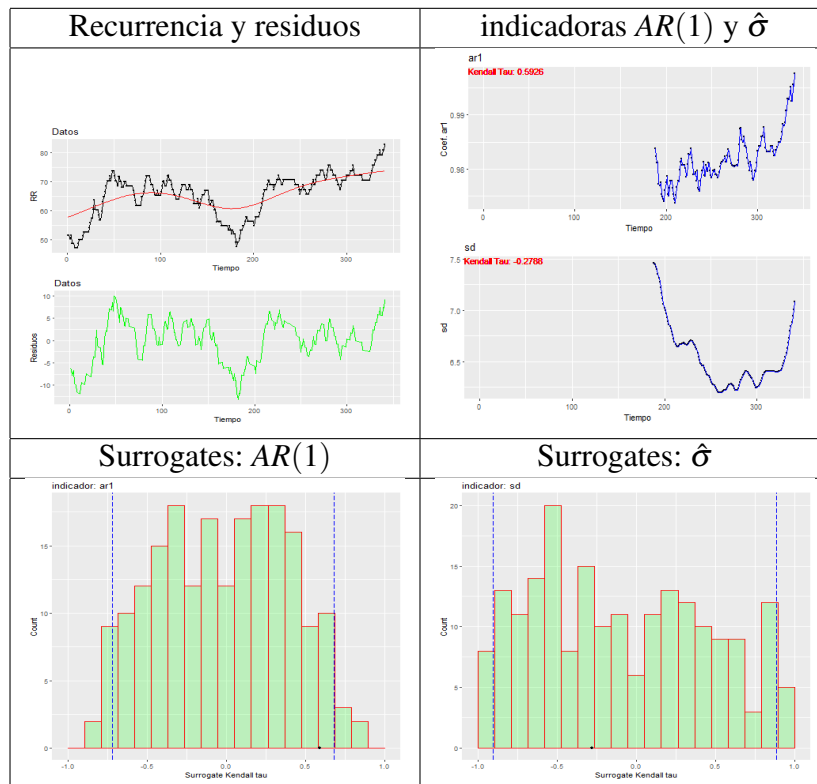


Figura 10.13: NATI. Arriba izquierda: Recurrencia (detrending) y residuos; arriba derecha: Medidas indicadoras AR(1) y $\hat{\sigma}$; debajo: test de *surrogates*.

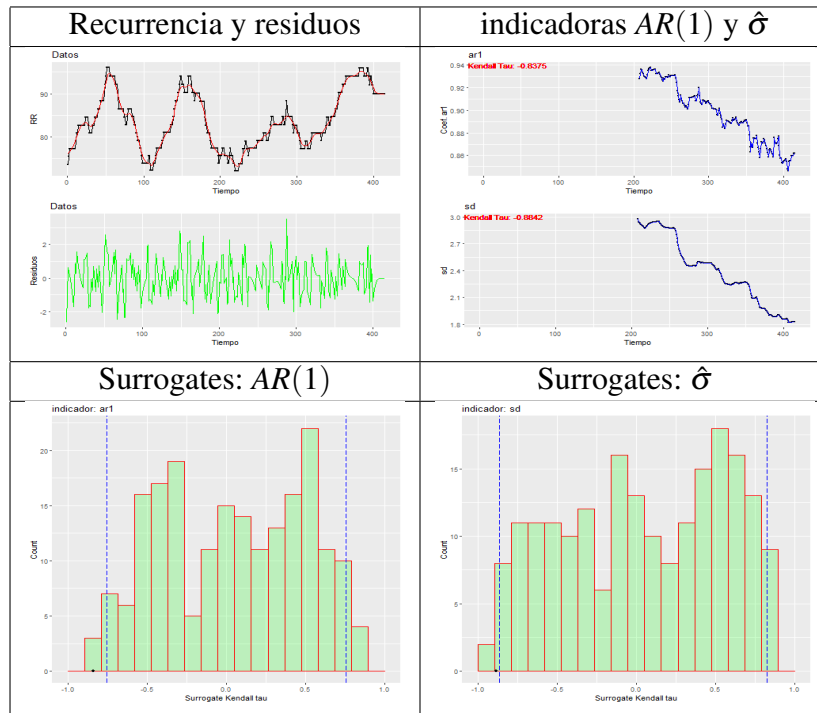


Figura 10.14: JAKO. Arriba izquierda: Recurrencia (detrending) y residuos; arriba derecha: Medidas indicadoras $AR(1)$ y $\hat{\sigma}$; debajo: test de *surrogates*.

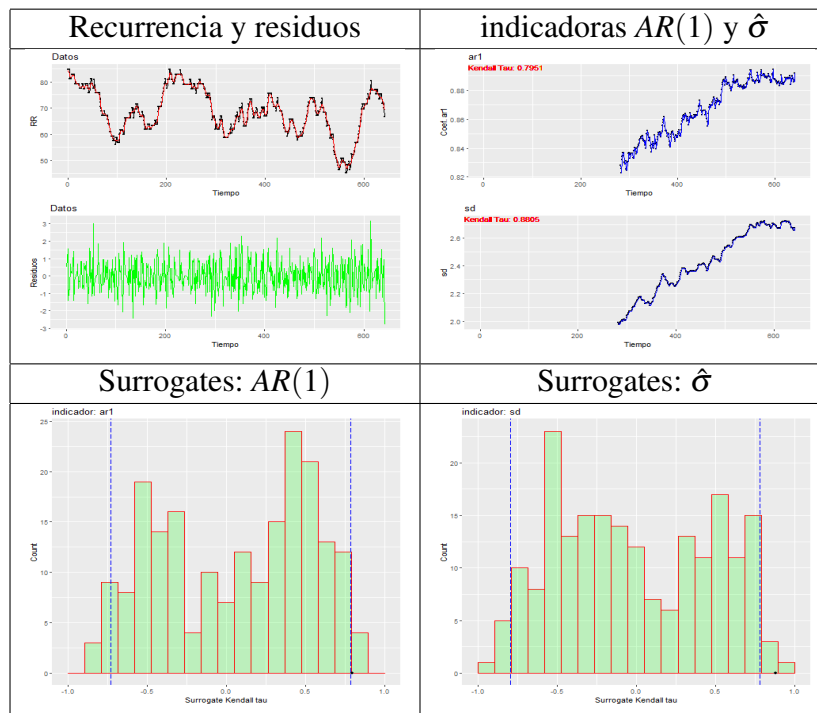


Figura 10.15: MIRKA. Arriba izquierda: Recurrencia (detrending) y residuos; arriba derecha: Medidas indicadoras $AR(1)$ y $\hat{\sigma}$; debajo: test de *surrogates*.

10.5.1.2. Heteroscedasticidad condicional.

Los Cuadros siguientes muestran, para cada alumno, dos gráficos: (i) el coeficiente de determinación según el tiempo (aquellas pruebas significativas son las que se ubican arriba de las líneas rojas); (ii) la acumulación de tests significativos en el tiempo. Cerca de la transición, se debería observar un clustering de tests significativos y un aumento de cantidad de test significativos. Se fijó $\alpha = 0,1$, la ventana en $w = 20$. Para SONIA la mayor acumulación de tests significativos se da hacia el tiempo 150 de la serie RR (periodo

entre sesiones 6 y 8 de la serie original). En lo que respecta a NATI parecen haber dos acumulaciones importantes, una hacia el tiempo 100 de la serie de *RR* (periodo entre sesiones 6 y 7 de la serie original); la otra hacia el tiempo 306 de la serie de *RR* (periodo entre sesiones 11 y 13 de la serie original). En lo que atañe a JAKO, no se observan acumulaciones importantes. Para MIRKA hay una acumulación inicial hacia el tiempo 124 (durante la sesión 3 de la serie original), una segunda hacia el tiempo 398 (periodo entre sesiones 7 y 9 de la serie original) y una última hacia el final, en el tiempo 642 (sesión 12 de la serie original).

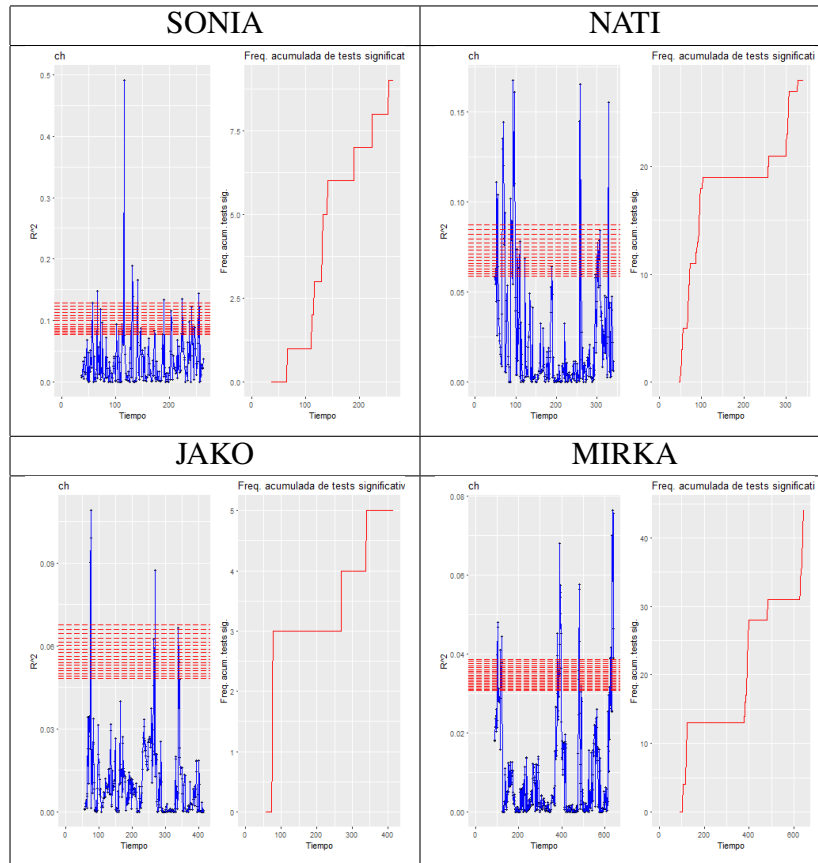


Figura 10.16: Heteroscedasticidad condicional. En cada alumno, izquierda: coeficiente de determinación según el tiempo (significativos sobre la línea roja); derecha: frecuencia acumulada de *tests* significativos en el tiempo.

10.5.1.3. Modelo de Drift-Diffusion-jump.

Para este indicador se transformaron los valores de *RR* con el logaritmo. No se usaron ventanas móviles. En primer lugar se muestran los gráficos con los datos transformados y sus primeras diferencias. Los siguientes muestran, para cada alumno, dos grupos de gráficos. El de la izquierda detalla la varianza condicional [arriba izquierda], la varianza total [arriba derecha], la varianza de la difusión [abajo izquierda] y la frecuencia de arribo de los saltos [abajo derecha], versus $a = \log(RR + 1)$, con *suavizado*. El gráfico de la izquierda muestra los mismos indicadores versus el tiempo.

En lo que respecta a SONIA, se observan dos regímenes diferentes separados hacia el tiempo 150 (sesión 8) [Cuadros de la derecha]. El drift muestra un pico en 150 y a partir de 200, lo cual coincide con el segundo régimen. La varianza condicional de la izquierda muestra un primer régimen entre los valores aproximados de 4.25 a 4.35 y un segundo régimen entre 4.4 y 4.5. Respecto de NATI, los indicadores versus el tiempo identifican dos regímenes separados hacia el tiempo 200 (entre las sesiones 8 y 10). Los indicadores versus la variable muestran un primer régimen entre los valores 3.9 y 4.2; y otro a partir de 3.3. En JAKO los Cuadros de la derecha indican que el sistema parece fluctuar entre los valores [4.3,4.4] y entre [4.4,4.55]. En los Cuadros de la derecha, la varianza condicional, total y la intensidad de arribos parecen indicar estabilización hacia el tiempo 350 (sesiones 11 / 12). En cuanto a MIRKA, los indicadores versus

el tiempo indican una transición hacia el tiempo 550 (sesiones 10 / 11); que se corresponde a valores entre 3.9 y 4.1.

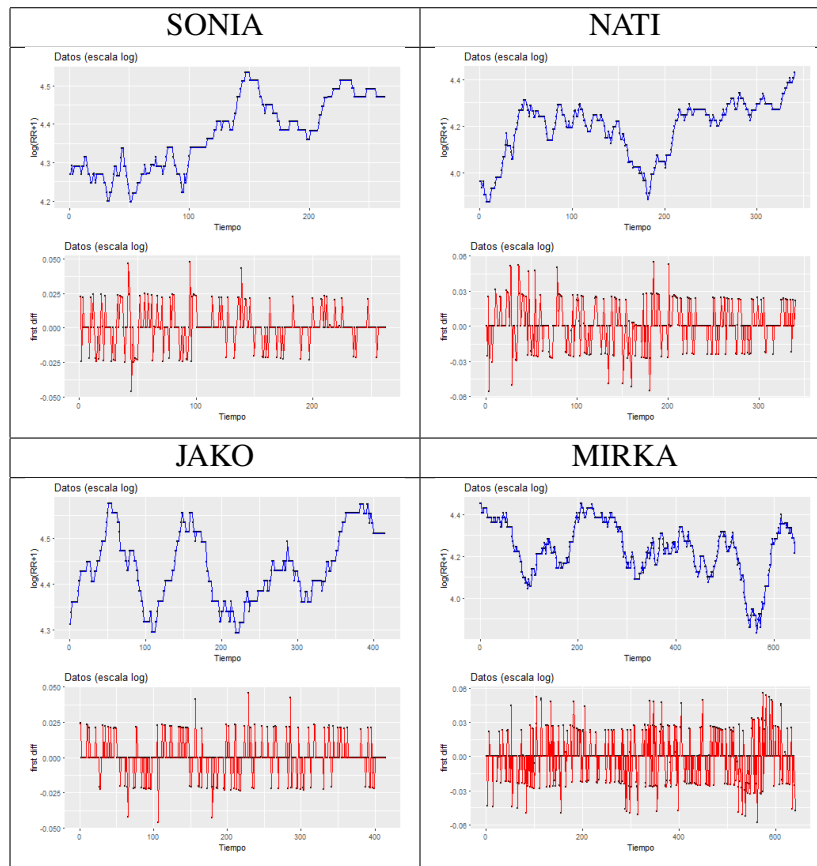


Figura 10.17: Datos transformados según $\log(RR + 1)$ y sus primeras diferencias.

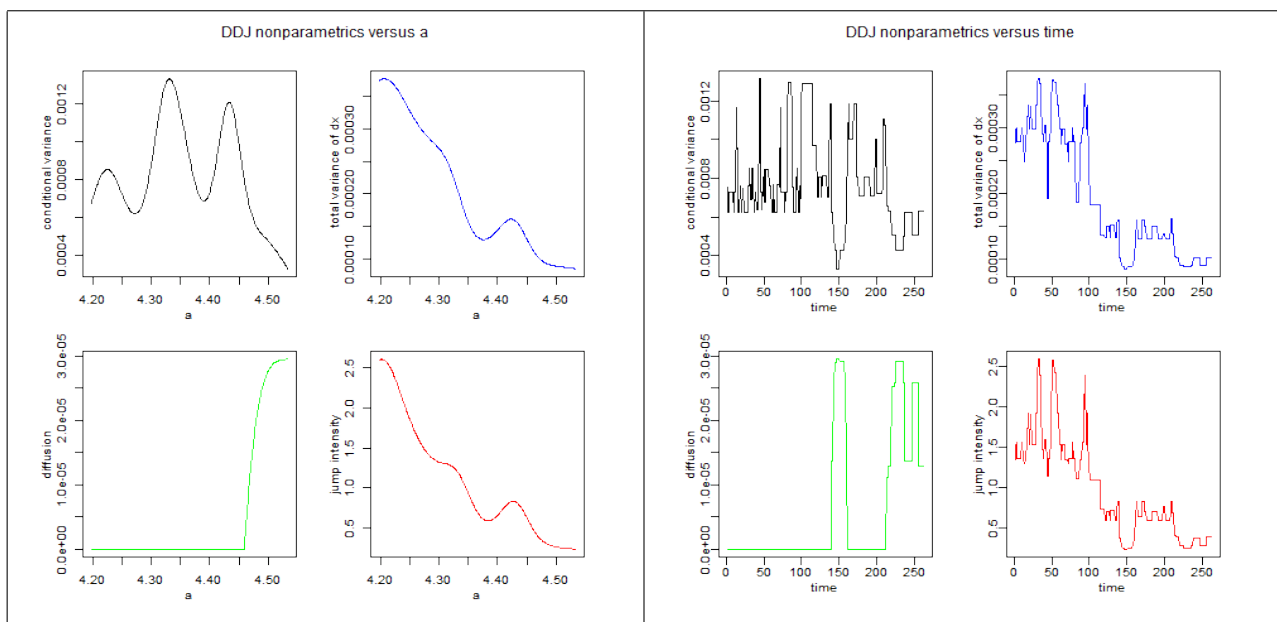


Figura 10.18: SONIA. Izquierda: la varianza condicional, la varianza total, la varianza de la difusión y la frecuencia de arribo los saltos, versus $a = \log(RR + 1)$. Derecha: mismos indicadores versus el tiempo.

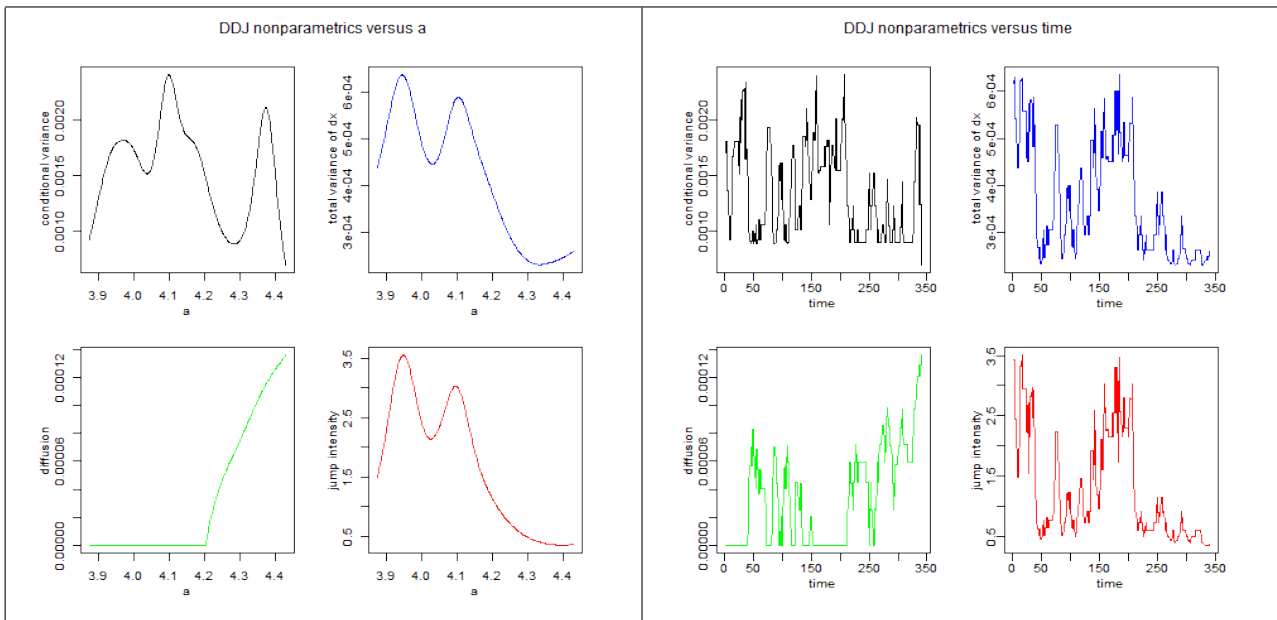


Figura 10.19: NATI. Izquierda: la varianza condicional, la varianza total, la varianza de la difusión y la frecuencia de arribo los saltos, versus $a = \log(RR + 1)$. Derecha: mismos indicadores versus el tiempo.

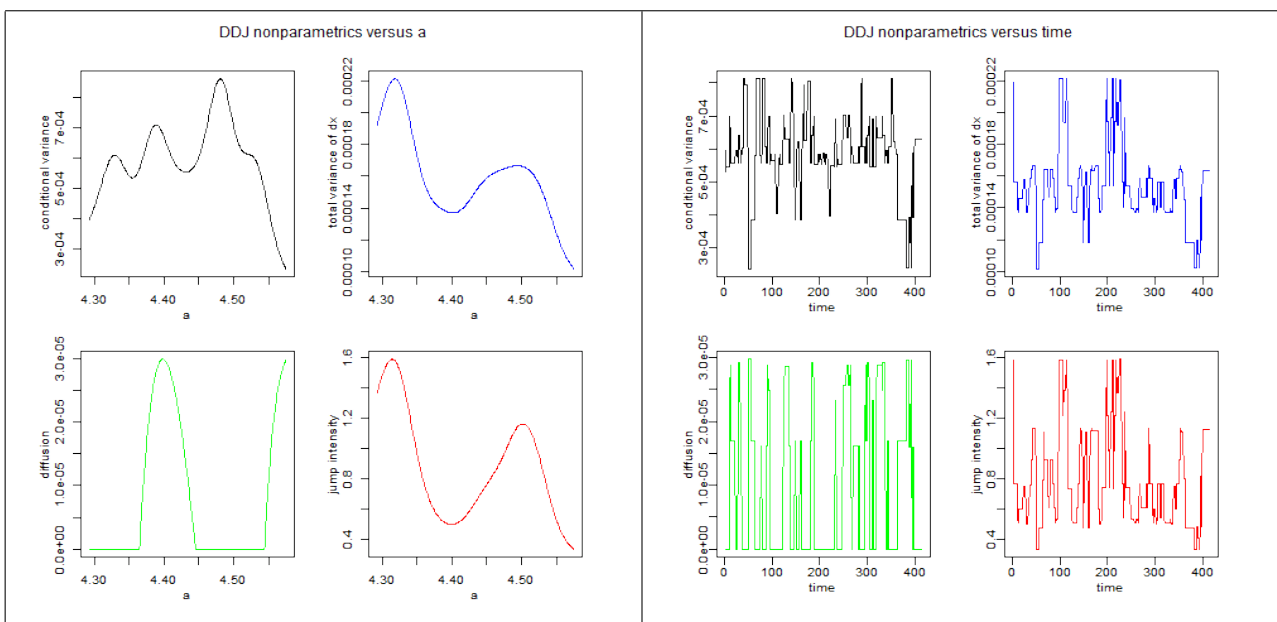


Figura 10.20: JAKO. Izquierda: la varianza condicional, la varianza total, la varianza de la difusión y la frecuencia de arribo los saltos, versus $a = \log(RR + 1)$. Derecha: mismos indicadores versus el tiempo.

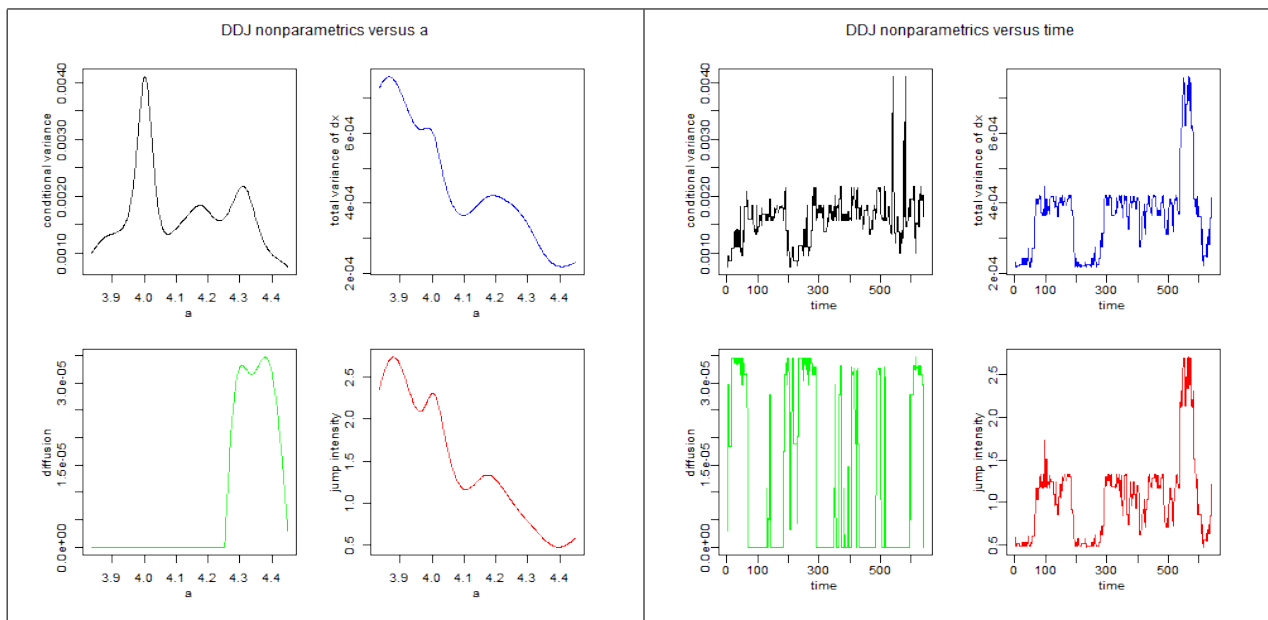


Figura 10.21: MIRKA. Izquierda: la varianza condicional, la varianza total, la varianza de la difusion y la frecuencia de arribo los saltos, versus $a = \log(RR + 1)$. Derecha: mismos indicadores versus el tiempo.

10.5.1.4. Análisis de Potenciales.

Se calculó el potencial para cada alumno haciendo correr una ventana mínima de $w = 0,1$ y una máxima de $w = 0,5$ (en porcentaje del total de los datos); aumentándola de a 10 puntos. No se consideraron como modas a las densidades menores a 0.1 . En las gráficas de la Figura 22 se muestran *raster plots* el ancho de la ventana versus el tiempo. Los colores indican la cantidad de “pozos” en el potencial en cada ventana, es decir, el número de atractores. En todos se identifican dos atractores, sin importar el ancho de la ventana.

10.5.1.5. Resumen de resultados de los indicadores.

El Cuadro siguiente resume la ubicación aproximada (períodos entre sesiones) de transiciones en los cuatro alumnos, según el resultado arrojado por cada indicador.

Indicador	SONIA	NATI	JAKO	MIRKA
coef. ar1; sd	7-8	no concluyente	¿ sin transición ?	11-12
Heterosced. cond.	6-7	6-7 ; 11-13	no concluyente	3;7-9;12
drift-diffusion-jump	8	8-10	11-12	10-11
potenciales: nro. de atractores	2	2	2	2

Cuadro 10.12: Resultado de los indicadores.

Con respecto a SONIA los indicadores coinciden en una transición hacia las sesiones 7-8. En el caso de NATI, los resultados del modelo de *drift-diffusion-jump* marcan claramente dos regímenes, por tanto se toman los resultados de éste y se ubica la transición hacia las sesiones 8-10. JAKO parece oscilar entre dos niveles del mismo atractor o bien entre dos atractores. Como el resultado de los potenciales da que hay dos atractores se elige esta posibilidad y se marca la transición hacia las sesiones 11-12. En MIRKA, el primer y el cuarto indicador coinciden en la sesión 11 para ubicar la transición. Sin embargo, en MIRKA la transición es hacia un atractor de menor competencia (más baja recurrencia). Por el contrario en JAKO y SONIA es hacia uno de más competencia (más alta recurrencia). En NATI la recurrencia aumenta hacia el final de la serie, con lo cual parece ir hacia una competencia más alta pero sin estabilizarse.

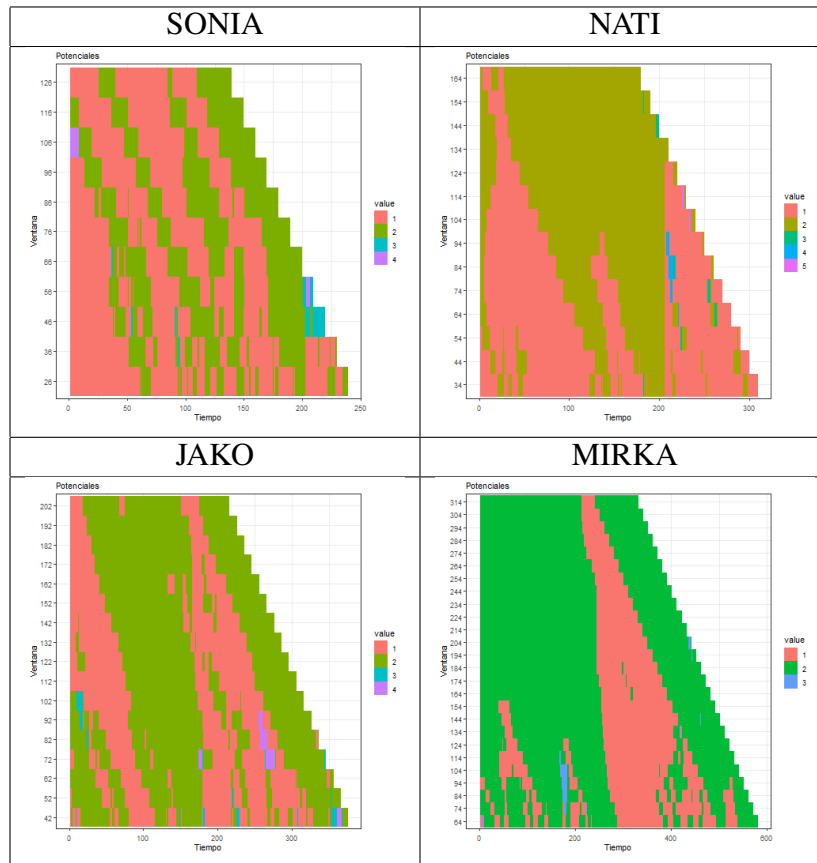


Figura 10.22: Análisis de potenciales: ancho de la ventana contra el tiempo. Los colores indican la cantidad de “pozos” del potencial.

10.6. Conclusiones.

El siguiente Cuadro muestra los resultados surgidos de la aplicación de las técnicas, que responden a las preguntas hechas al principio del capítulo. Respecto de los *motifs*, JAKO y SONIA poseen más sesiones en el tópic 2 del “atractor correcto”. En cambio, para NATI y MIRKA cerca de la mitad de las sesiones están en el “atractor fuerte del error” (tópico 3). En el plano *Entropía-Complejidad*, tanto JAKO como SONIA están pasando del régimen de TRANSICIÓN al de NO_ERROR, con JAKO más avanzado. En contraste, para NATI y MIRKA se hallan más sesiones en el tránsito del régimen ERROR al de TRANSICIÓN. El análisis con *RQA* reveló más sesiones con cambio de dinámica (por ejemplo, más períodos de laminaridad) en SONIA y MIRKA. O sea que MIRKA, aunque comete más errores, estos están más lejos en el tiempo durante más sesiones que en NATI. La evidencia de transiciones arrojó coincidencia entre los indicadores y el análisis de *change point* hacia la sesión 8 en SONIA, y hacia la 11 en JAKO. En NATI parece haber un cambio entre las sesiones 7-8, según ambas técnicas. En cuanto a MIRKA, los indicadores mostraron una transición hacia la sesión 11. Por último, el análisis de potenciales apuntó hacia dos atractores, justificando la posibilidad de transiciones entre ellos. El principal problema del análisis de este capítulo fue que las técnicas utilizadas requieren gran cantidad de datos, con lo cual se tuvo que recurrir a ventanas móviles de larga duración. Esto dificultó a su vez la identificación más precisa de particulares sesiones donde se hallen las transiciones posibles.

Pregunta	Técnica	SONIA	NATI	JAKO	MIRKA
¿Sesiones de dinámica similar?	Motifs	Tópico 1,2(↑)	Tópico 1,2,3(↑)	Tópico 1,2(↑)	Tópico 1,2,3(↑)
¿Regímenes?	Plano E-C	TRANS. → NO_ERROR	ERROR → TRANS.	TRANS. → NO_ERROR	ERROR → TRANS.
¿Sesiones con cambio de dinámica?	RQA	3,7,8,9,10,12	4,14	3,6,12,14	1,2,6,8,12
¿Transiciones?	Entropía	5-8	5-7	9-11	2-3 (?)
	Indicadores	7-8	8-10	11-12	11

Cuadro 10.13: Resultados de las técnicas. “↑” indica más sesiones dentro del tópico; (?) indica “no concluyente”.

10.7. Apéndice de capítulo.

	sesiones	freq.rel	freq.abs	n	temas	ref.
4	1.2	0.33	6.00	18	1	1.00
5	1.3	0.21	4.00	19	2	1.00
6	1.4	0.33	14.00	42	1	1.00
7	1.5	0.26	11.00	42	2	1.00
8	1.6	0.28	7.00	25	1	1.00
9	1.7	0.11	3.00	28	2	2.00
10	1.8	0.19	7.00	37	2	2.00
11	1.9	0.19	5.00	26	1	2.00
1	1.10	0.05	1.00	19	2	2.00
2	1.11	0.18	6.00	34	1	2.00
3	1.12	0.06	1.00	18	2	2.00
12	2.1	0.40	12.00	30	3	3.00
18	2.2	0.62	10.00	16	3	3.00
19	2.3	0.24	5.00	21	1	1.00
20	2.4	0.15	2.00	13	2	2.00
21	2.5	0.33	6.00	18	3	1.00
22	2.6	0.38	18.00	47	3	3.00
23	2.7	0.35	13.00	37	1	3.00
24	2.8	0.50	19.00	38	3	3.00
25	2.9	0.26	7.00	27	1	1.00
13	2.10	0.38	14.00	37	3	3.00
14	2.11	0.24	7.00	29	1	1.00
15	2.12	0.32	9.00	28	1	1.00
16	2.13	0.27	6.00	22	1	1.00
17	2.14	0.11	3.00	27	2	2.00
26	3.1	0.23	6.00	26	1	1.00
32	3.2	0.19	10.00	52	2	2.00
33	3.3	0.03	1.00	30	2	2.00
34	3.4	0.21	6.00	28	1	1.00
35	3.5	0.20	8.00	41	2	2.00
36	3.6	0.07	2.00	30	2	2.00
37	3.7	0.24	9.00	37	1	1.00
38	3.8	0.25	8.00	32	1	1.00
39	3.9	0.16	8.00	51	2	2.00
27	3.10	0.28	5.00	18	1	1.00
28	3.11	0.15	7.00	47	2	2.00
29	3.12	0.03	1.00	32	2	2.00
30	3.13	0.19	5.00	27	1	2.00
31	3.14	0.00	0.00	13	2	2.00
40	4.1	0.15	12.00	80	2	2.00
44	4.2	0.33	13.00	39	3	1.00
45	4.3	0.33	21.00	64	1	1.00
46	4.4	0.24	15.00	62	1	1.00
47	4.5	0.24	8.00	34	1	1.00
48	4.6	0.21	10.00	47	1	1.00
49	4.7	0.37	30.00	81	3	3.00
50	4.8	0.20	7.00	35	2	3.00
51	4.9	0.36	27.00	74	3	3.00
41	4.10	0.33	17.00	52	3	1.00
42	4.11	0.46	22.00	48	3	3.00
43	4.12	0.25	19.00	75	1	1.00

Cuadro 10.14: Clasificación de tópicos. sesiones: sesión por cada alumno; freq.rel: frecuencia relativa de error; freq. abs: frecuencia absoluta de error; n: número de instancias por sesión; temas: tópico asignado a cada sesión; ref.: tópico de referencia.

	PE.Sonia	C.Sonia	PE.Nati	C.Nati	PE.Jako	C.Jako	PE.Mirka	C.Mirka
1	0.918	0.061	0.74	0.16	0.53	0.22	0.473	0.221
2	0.874	0.09	0.84	0.11	0.52	0.22	0.689	0.182
3	0.503	0.22	0.72	0.17	0.22	0.16	0.818	0.124
4	0.759	0.153	0.41	0.22	0.76	0.15	0.715	0.172
5	0.712	0.173	0.52	0.22	0.61	0.20	0.799	0.134
6	0.738	0.163	0.67	0.19	0.36	0.21	0.667	0.19
7	0.503	0.22	0.85	0.10	0.65	0.20	0.849	0.106
8	0.65	0.195	0.75	0.16	0.77	0.15	0.431	0.22
9	0.634	0.199	0.78	0.14	0.53	0.22	0.783	0.142
10	0.31	0.197	0.85	0.10	0.67	0.19	0.714	0.173
11	0.684	0.184	0.68	0.19	0.56	0.21	0.879	0.087
12	0.323	0.2	0.83	0.12	0.21	0.15	0.671	0.188
13			0.70	0.18	0.71	0.18		
14			0.39	0.21	0.00	0.00		

Cuadro 10.15: Medidas de Entropía de Permutación (PE) y Complejidad (C) para cada aprendiziente.

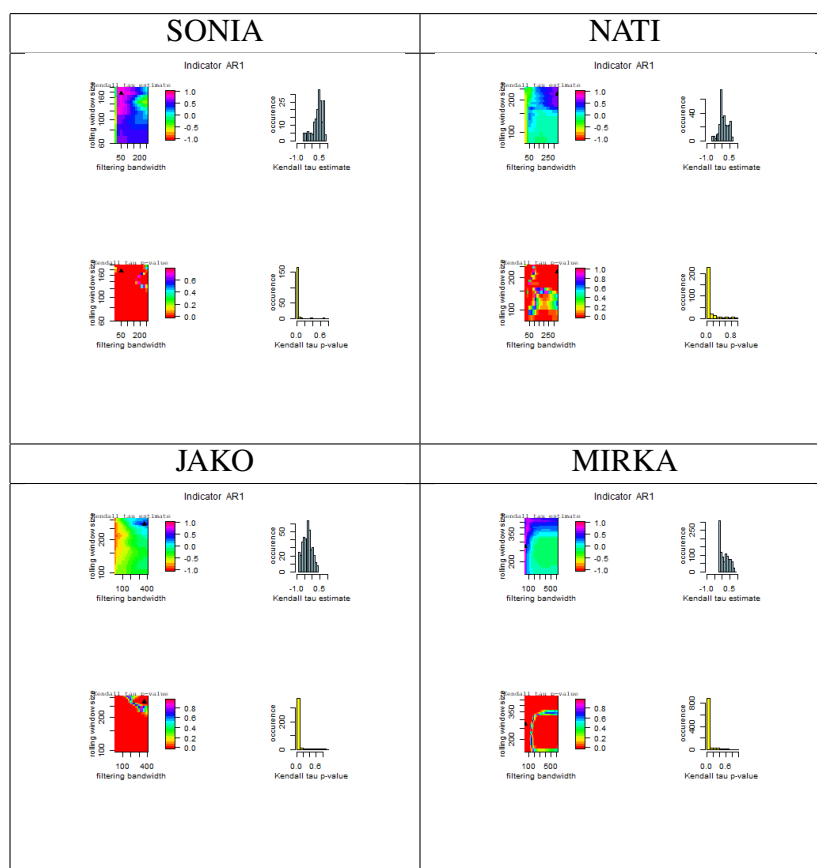


Figura 10.23: Análisis de sensibilidad para los parámetros de AR(1) [auto-correlación]. Izquierda: gráficos de valores de τ (arriba) y su p-valor (abajo) según diferentes valores de largo de ventana (eje vertical) versus ancho de banda (eje horizontal). Derecha: histograma de los valores de τ (arriba) y de sus p-valor (debajo).

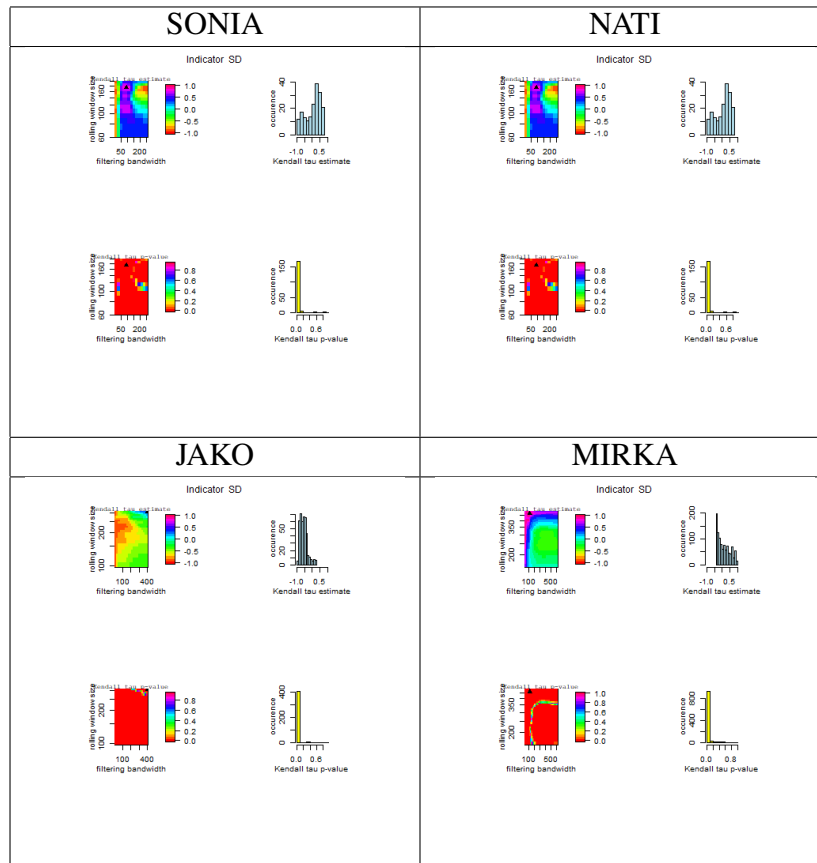


Figura 10.24: Análisis de sensibilidad para los parámetros de SD [desvío típico]. Izquierda: gráficos de valores de τ (arriba) y su p-valores (abajo) según diferentes valores de largo de ventana (eje vertical) versus ancho de banda (eje horizontal). Derecha: histograma de los valores de τ (arriba) y de sus p-valores (debajo).

Capítulo 11

Dinámica en redes complejas y predicción de error.

El presente capítulo introduce nociones de redes complejas. Se describen las redes creadas y se analizan sus características estructurales. Luego, se presenta el denominado “modelo estocástico de agentes”, que permite llevar a cabo un análisis estadístico temporal sobre redes. Además, se muestra una serie de atributos diseñados para caracterizar los enlaces de las redes y se aplican modelos de minería de datos para clasificar el estatus de error de las últimas cuatro sesiones de cada alumno. Por último, se aplica un *clustering* por mixtura de distribuciones, se seleccionan atributos relevantes mediante un modelo mixto con regularización y se ajusta a los datos de cada *cluster* un modelo mixto generalizado.

11.1. Nociones sobre grafos.

En esta sección se seguirá a Barabási & Pósfai (2016) y Newman (2010).

Red. Una red (grafo) es una colección de nodos conectados por medio de enlaces. Sea V el conjunto de nodos: $V = \{n_1, n_2, \dots, n_N\}$; y E el conjunto de enlaces: $E = \{e_1, e_2, \dots, e_M\}$ (N nodos totales y M enlaces totales). Un grafo simple $G(V, E)$ es un par de conjuntos V y E donde $V \neq \{\emptyset\}$. En la versión *dirigida* los enlaces son pares *ordenados* de elementos distintos de V , en cambio; en la *no dirigida* son pares *no-ordenados*. En este grafo los enlaces tienen pesos igual a 1; pero se puede asignar un peso a cada enlace según una función de peso $w : E \times \mathbb{R}$. Un grafo G es *conexo* si existe un camino entre cualquier par de vértices del mismo. Una *componente* es un subconjunto de vértices de un grafo que: (i) definen un (sub)grafo conexo y (ii) no están incluidos en ningún otra estructura conexa. La componente más grande de un grafo se denomina *componente gigante*.

Matriz de adyacencia. Una forma de representar un grafo es mediante una matriz de adyacencia. Si se denota al enlace entre los nodos i y j como (i, j) , la matriz de adyacencia A es aquella cuyos elementos a_{ij} cumplen:

$$a_{ij} \begin{cases} w & \text{si hay un enlace entre los nodos } i \text{ y } j \\ 0 & \text{caso contrario} \end{cases}$$

donde $w = 1$ en la versión no pesada y $A = A^T$ ($a_{ij} = a_{ji}$) en la versión no dirigida. Si hay bucles, entonces $a_{ii} = w$, sino $a_{ii} = 0$. En la versión no dirigida se cuentan $M_{max} = N^2$ enlaces como máximo; por otro lado, en la no dirigida hay como máximo: $M_{max} = \binom{N}{2} = \frac{N(N-1)}{2}$.

Grado. El grado de un nodo i -ésimo es el número de enlaces adyacentes a éste: $k_i = \sum_{j=1}^N A_{ij}$. Ya que cada enlace posee dos extremos se tiene que $2M = \sum_{i=1}^N k_i$. El grado medio será:

$\langle k \rangle = \frac{1}{N} \sum_{i=1}^N k_i = \frac{2M}{N}$. En el caso de las redes dirigidas cada enlace puede tener un extremo de entrada (“in”) y otro de salida (“out”), con lo cual se calcula el grado para ambos: $k_i^{in} = \sum_{j=1}^N A_{ij}$; $k_i^{out} = \sum_{i=1}^N A_{ij}$. En este caso: $M = \sum_{i=1}^N k_i^{in} = \sum_{i=1}^N k_i^{out} = \sum_{ij} A_{ij}$. El grado medio para cada tipo será: $\langle k^{in} \rangle = \frac{1}{N} \sum_{i=1}^N k_i^{in}$ y $\langle k^{out} \rangle = \frac{1}{N} \sum_{i=1}^N k_i^{out}$. Y el grado medio total: $\langle k \rangle = \frac{1}{N} \sum_{ij} A_{ij} = \frac{M}{N}$.

Para el caso de redes dirigidas, “strength” suma, para cada nodo i -ésimo, el peso (valor de un atributo) de las j -aristas que entran y/o salen de dicho nodo: $s_i^{in} = \sum_{j=1}^N w_{ij} A_{ij}$, $s_i^{out} = \sum_{i=1}^N w_{ij} A_{ij}$.

Diámetro. El diámetro de un grafo es el camino más corto entre dos nodos de longitud máxima; es decir, es la distancia máxima entre dos pares de nodos. Se denota como d_{max} , donde d es la distancia geodésica. Por otro lado, la distancia media, $\langle d \rangle$, es el promedio de las distancias entre todos los pares de nodos de la red. Reza:

$$\langle d \rangle = \frac{1}{N(N-1)} \sum_{i,j=1, N; i \neq j} d_{ij}$$

Densidad. La densidad del grafo es la cantidad de enlaces efectivamente presentes sobre la máxima cantidad de enlaces que el grafo podría tener. Se calcula como:

$$\rho = \frac{M}{M_{max}} = \frac{M}{\binom{N}{2}} = \frac{2M}{N(N-1)} = \frac{\langle k \rangle}{N-1}$$

Si una red crece en el tiempo; o sea si: $N = N(t)$ y $M = M(t)$, puede suceder que a medida que $N \rightarrow \infty$, la densidad vaya a cero $\rho \rightarrow 0$ (la red es rala) o bien, hacia una constante $\rho \rightarrow b$ (la red es trófica).

Coficiente de Clustering. El coeficiente de *clustering local* cuantifica en qué medida los vecinos del nodo i -ésimo son vecinos entre sí (o sea que forman triángulos). Se trata de una medida de densidad local. Reza:

$$C_i = \frac{2L_i}{k_i(k_i-1)}; L = \sum_{j,k \in \text{Neigh}(i); j \neq k} a_{jk}$$

donde L es el número de pares de vecinos enlazados (y que a su vez, enlazados éstos al nodo i forman un triángulo). Promediando sobre todos los nodos, el coeficiente de *clustering medio* indica en qué medida pares de vecinos de un nodo *tomado al azar* en la red son vecinos entre sí: $\langle C \rangle = \frac{1}{N} \sum_{i=1}^N C_i$. Por otra parte, el coeficiente de *clustering global* muestra en qué medida es posible encontrar triángulos (clausura transitiva) en la red. Es: $C_{\Delta} = 3 \frac{\Delta}{T}$; donde Δ es el número de tripletes conectados (triángulos) y T el número de tripletes.

Coficiente de Asortatividad. El coeficiente de *assortative mixing* mide la posible segregación en la red. Es positivo si dos nodos con atributos de *la misma* categoría tienden a conectarse (red asortativa / homofilia) y es negativo cuando tienden a diferir (red disortativa). Defínase Q , el coeficiente de modularidad como:

$$Q = \frac{1}{2m} \sum_{ij} \left(A_{ij} - \frac{k_i k_j}{2m} \right) \delta(c_i, c_j)$$

donde: c_i y c_j es la categoría a la que pertenece el atributo del nodo i / j , y $\delta(c_i, c_j) = 1$ si $c_i = c_j$ sino es *ceros*. Si hay asortatividad entonces $0 < Q < 1$ y si $Q < 0$ la red es disortativa. El segundo término de la resta indica el número esperado de enlaces en los que ambos extremos tienen la misma categoría *por azar*. Con el objetivo de asegurar que $Q = 1$ en el caso de que la red sea perfectamente asortativa, se divide Q por su máximo: $Q_{max} = \frac{1}{2m} \sum_{ij} \left(2m - \frac{k_i k_j}{2m} \delta(c_i, c_j) \right)$, con lo cual se obtiene el coeficiente de asortatividad:

$$A = \frac{Q}{Q_{max}}.$$

Para el caso continuo: $R = \frac{\sum_{ij} i j \left(A_{ij} - \frac{k_i k_j}{2m} \right) x_i x_j}{\sum_{ij} i j \left(k_j \delta_{ij} - \frac{k_i k_j}{2m} \right) x_i x_j} \in [0, 1]$; donde el par (x_i, x_j) son los valores (continuos) de

los atributos en los nodos en los extremos de una arista (ij). Se trata de una medida análoga a la correlación de Pearson. Si ambos valores (x_i, x_j) tienden a crecer o decrecer, entonces la correlación será positiva (asortatividad) y si varían en direcciones opuestas, será negativa (disortatividad).

Distribución de grado. La distribución del grado de los nodos indica la probabilidad de que un nodo tomado al azar tenga grado k : $p_k = \frac{n_k}{N}$, donde n_k es el número de nodos con grado k (por supuesto: $\sum_k p_k = 1$).

Modelos teóricos.

Grafos aleatorios [“random graphs”]. Un *grafo aleatorio* es un grafo obtenido muestreando de una colección dada de grafos. Sea $\mathcal{G}(N_v, N_e)$ la colección de grafos $G(V, E)$ con $|V| = N_v$ (número de nodos) y $|E| = N_e$ (número de vértices). Un grafo aleatorio se crea asignando una probabilidad *uniforme* $\mathbb{P}(G) = \left(\frac{N}{N_e} \right)^{-1}$ a cada grafo $G \in \mathcal{G}(N_v, N_e)$, donde $N = \binom{N_v}{2}$ es el número total de enlaces. Una posibilidad alternativa (más usada en la práctica) es definir la colección de grafos $\mathcal{G}(N_v, p)$ como aquella que contiene todos los grafos de orden N_v que se pueden obtener asignando cada enlace a cada par de vértices con probabilidad $p \in (0, 1)$. Un grafo aleatorio se forma pues; (i) o bien poniendo L enlaces al azar entre los pares de nodos; (ii) o bien poniendo un enlace con probabilidad p . Se producen grafos con coeficiente de *clustering* promedio bajo y dependiente del grado. La distribución de grado sigue una distribución *Poisson* de parámetro $\langle k \rangle$ si N es grande y p pequeño.

Grafos “Small World”. El grafo se construye a partir de un círculo de nodos, donde cada nodo está conectado a c vecinos inmediatos. Luego, un número pequeño de enlaces se re-cablean con probabilidad $p \in [0, 1]$ conectando dos pares de nodos elegidos al azar. Si $p = 1$ se obtiene un grafo aleatorio (y si $p = 0$ se vuelve al grafo circular original). La red obtenida posee coeficiente de *clustering* alto y diámetro bajo (efecto “small world”). Con $p = 0$ la distribución de grado tiene toda la masa en un solo punto; en cambio si $p = 1$ se obtiene la distribución de la red aleatoria. El diámetro crece según $\ln N$.

Grafos de “preferential attachment”. Este modelo crea redes con distribuciones de grado con colas pesadas, que aproximan una ley de potencia. Así, unos pocos nodos tienen grado alto y la mayoría grado bajo. Se comienza con m_0 nodos, cuyos enlaces se eligen arbitrariamente, con tal de que cada nodo quede con al menos un enlace. En cada tiempo t se agrega un nuevo nodo con m ($m \leq m_0$) enlaces que conectan los m nodos presentes en la red. La probabilidad $P(k)$ de que un enlace de un nuevo nodo se conecte con otro nodo i de grado k_i depende del grado del nodo i como: $P(k_i) = \frac{k_i}{\sum_j k_j}$. O sea que se pone en funcionamiento un mecanismo de “los ricos se vuelven más ricos”. Debido a esto se forman *hubs* en la red, y la distribución

de grado obtenida sigue una ley de potencia. Se obtiene una con $N = t + m_0$ nodos y $M = m_0 + mt$ enlaces. El coeficiente de *clustering* promedio es independiente del grado pero decrece con N .

El Cuadro que sigue compara el modelo de “Preferential attachment” con el de “random graph” según sus distribuciones de grado [P_k], coeficiente de clustering promedio [$\langle C \rangle$], distancia promedio [$\langle d \rangle$] y grado medio [$\langle k \rangle$].

	R.G.	P.A.
P_k	$p_k = e^{-\langle k \rangle} \frac{\langle k \rangle^k}{k!}$ [$N \rightarrow \infty; p \rightarrow 0$]	$p_k = k^{-\gamma}$
$\langle C \rangle$	$\frac{\langle k \rangle}{\ln \langle k \rangle}$	$\frac{(\ln N)^2}{N}$
$\langle d \rangle$	$\frac{\ln N}{\ln \langle k \rangle}$	$\frac{\ln N}{\ln(\ln N)}$
$\langle k \rangle$	$p(N-1)$	$2m$

Cuadro 11.1: Modelos teóricos de redes: RG = “random graph”; PA = “Preferential Attachment”.

11.2. Redes de concordancia.

Se crearon cuatro redes acumulando los casos de concordancia de las sesiones. Las redes eran dirigidas, donde la dirección indicaba la relación de asimetría del controlador en la concordancia plural. Se asignaron pesos a las aristas de los grafos, indicando la cantidad de errores en la que la concordancia entre las palabras (ij) tenía un error (0, 1, 2, ...). Además, se asignó a los nodos los atributos descritos en el Cuadro 2. Por otra parte las Figuras [1 y 2] muestran las redes y una ampliación de la red SONIA hasta la sesión 3. El Cuadro 3 señala características globales de las redes. Se trata de redes jerárquicas, de diámetro bajo y coeficiente de *clustering* bajo. Los *hubs* corresponden a las palabras funcionales. En consecuencia, la distribución de grado tiene la forma de una distribución Pareto (“power law”) con mucha masa de probabilidad en dichos *hubs* y poca masa en el resto [Figuras 3 y 4]. El exponente de las “power laws” [$p_k = k^{-\gamma}$] se encuentra entre $2 < \gamma < 3$, y resulta coherente con un mecanismo generador de “preferential attachment”. La correlación de grado es negativa, ilustrando el hecho de que los *hubs* no tienden a conectarse entre sí, sino que las conexiones son entre nodos de grado alto y bajo; y entre nodos de grado bajo. La medida “strength” media (S) muestra la suma en promedio de los pesos de las aristas que entran en un nodo; es decir, indica la cantidad media de errores en los que está involucrado un controlador. Por otro lado la medida de asortatividad (R) señala la correlación entre los valores de cantidad de error de los nodos de las concordancias. Como es negativo, significa que la red es disortativa respecto del error. Ambos términos de la concordancia varían en direcciones opuestas en cuanto a la cantidad de errores en los que se hallan involucrados.

Variable	Descripción	Clase	Niveles
MOD	Tipo de modificador del controlador	Cualitativa	0 = no es controlador, 1 = artículo definido; 2 = artículo indefinido; 3 = determinante; 4 = adjetivo
ERROR	Cantidad de errores de concordancia en los que interviene la palabra.	Cuantitativa	[0, 1, 2, ...]
ES	¿Tiene la palabra plural con “e” epentética?	Cualitativa	0 = NO, 1 = SI
ANIM	¿Es la palabra controlador animado?	Cualitativa	0 = no es controlador; 1 = controlador inanimado; 2 = controlador animado.
FAM.LEX	Índice PCA de familiaridad y frecuencia léxica del controlador	Cualitativa	0 = no es controlador; 1 = baja; 0 = alta
IMA.CONC	Índice PCA de imaginabilidad y concreción léxica del controlador	Cualitativa	0 = no es controlador; 1 = baja; 0 = alta
ESP1	Tipos de terminaciones del “objetivo”	Cualitativa	1 = “us” [Ej.: <u>sus</u>]; 2 = “is” [Ej.: <u>mis</u>]; 3 = “os” [ej.: <u>rojos</u>]; 4 = “as” [ej.: <u>blancas</u>]; 5 = “es” [ej.: <u>azules</u>]; 6 = “es*” [ej.: <u>útiles</u> (-e- epentética)]
ESP2	Tipos de terminaciones del “controlador”	Cualitativa	1 = “us” [Ej.: <u>sus</u>]; 2 = “is” [Ej.: <u>mis</u>]; 3 = “os” [ej.: <u>rojos</u>]; 4 = “as” [ej.: <u>blancas</u>]; 5 = “es” [ej.: <u>azules</u>]; 6 = “es*” [ej.: <u>útiles</u> (-e- epentética)]
MORF	Similaridad de la palabra entre terminaciones de Español e Italiano	Cuantitativa	[0,6;2]
STEM	Similaridad de la palabra entre las raíces léxicas de Español e Italiano	Cuantitativa	[0; 7,6]

Cuadro 11.2: Atributos de los nodos.

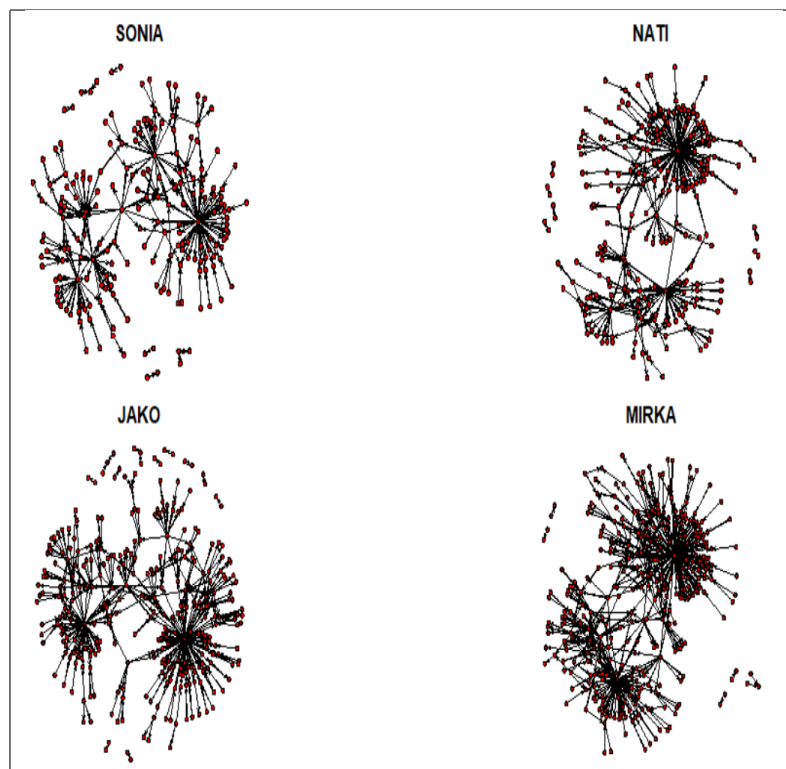


Figura 11.1: Visualización de las Redes.

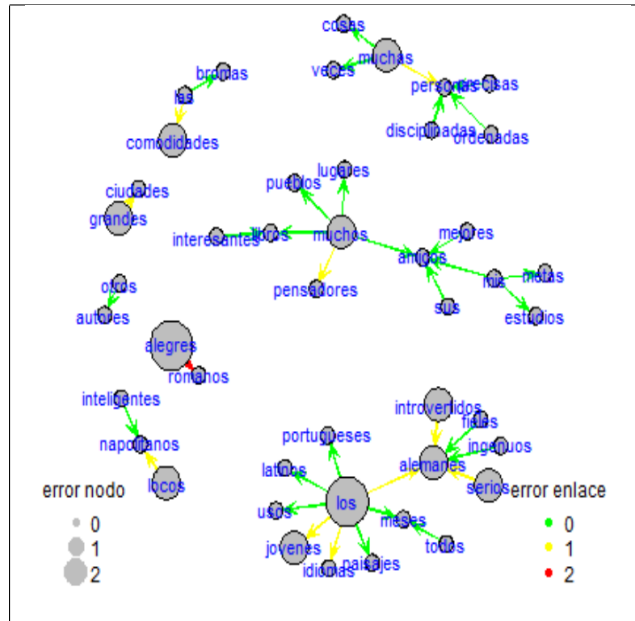


Figura 11.2: Visualización de las Redes (SONIA: hasta sesión 3).

	K	K_max	N	E	D	C	L	S	gama	R_deg	R
SONIA	3.23	87.00	193.00	312.00	0.01	0.04	1.03	0.35	2.02	-0.33	-0.32
NATI	3.44	119.00	227.00	390.00	0.01	0.08	1.04	0.58	2.21	-0.35	-0.17
JAKO	2.97	129.00	312.00	464.00	0.00	0.00	1.02	0.24	2.35	-0.28	-0.35
MIRKA	4.03	252.00	343.00	691.00	0.01	0.08	1.16	0.59	2.09	-0.32	-0.23

Cuadro 11.3: Medidas globales de las redes. K = grado medio; K.max = grado máximo, N = número de nodos; E = número de aristas; D = densidad; C = coeficiente de clustering medio; L =diámetro medio; S = strenght medio; gama = exponente de power law; R = coeficiente (cuantitativo) de asortatividad; R.deg = correlación de grado

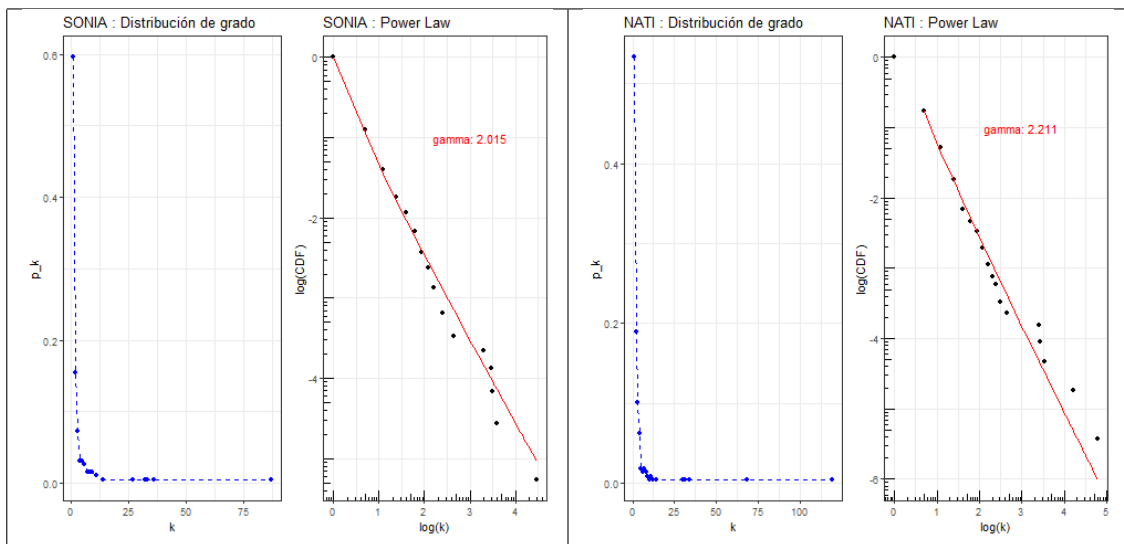


Figura 11.3: Distribución de grado: SONIA, NATI.

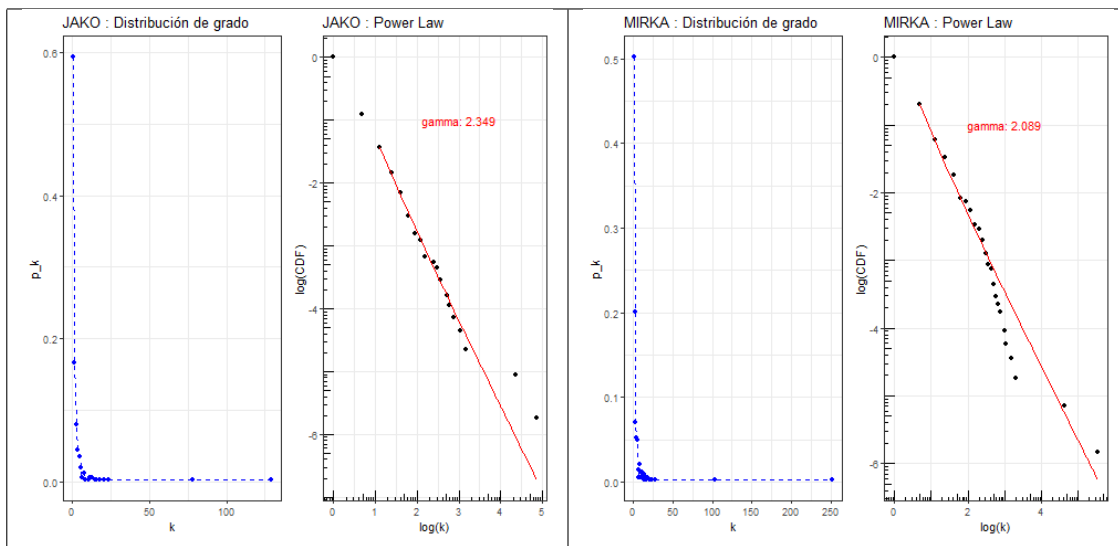


Figura 11.4: Distribución de grado: JAKO, MIRKA.

Ahora bien, la Figura 5 muestra la evolución de las medidas “S” (strength medio) y “R” (asortatividad) según van creciendo las redes. Resulta interesante ver donde se estabiliza el (de)crecimiento, identificando puntos de corte de posibles regímenes de dinámica diferente. A tal efecto se aplicó un análisis de cambio de punto para examinar un posible cambio en la locación. Dicha metodología sirve para estimar el punto de una serie en el cual cambian las propiedades estadísticas de dicha serie. Se utilizó el método no paramétrico propuesto por Dehling *et al.* (2018), útil para series de datos no independientes y asimétricos. Para cada punto k se calcula un estadístico basado en el estimador de *Hodges-Lehmann*¹ entre la serie segmentada entre dos grupos $[1, k]$ y $[k + 1, n]$ y se elige el máximo en valor absoluto. El estimador rechaza la hipótesis nula (de que el proceso es estacionario: la locación no cambia en el tiempo) para valores (negativos o positivos) grandes del estadístico. Los resultados se muestran en los Cuadros que siguen. En lo que respecta a “S” habría un cambio de régimen a partir de la sesión 6 para NATI y MIRKA (SONIA y JAKO no rechazan la hipótesis nula). En lo que atañe a la asortatividad, NATI cambiaría de régimen a partir de la quinta sesión; JAKO y MIRKA a partir de la tercera.

	point	estadístico	p.valor
SONIA	4.00	0.99	0.28
NATI	6.00	1.73	0.01
JAKO	7.00	0.87	0.43
MIRKA	6.00	2.04	0.0005

Cuadro 11.4: Análisis de cambio de punto: Strength medio.

	point	estadístico	p.valor
SONIA	3.00	1.25	0.0884
NATI	5.00	2.72	p<0.0001
JAKO	3.00	1.43	0.0328
MIRKA	3.00	2.69	p<0.0001

Cuadro 11.5: Análisis de cambio de punto: Asortatividad.

¹Dados dos grupos independientes de observaciones X_1, \dots, X_{n_1} y Y_1, \dots, Y_{n_2} el estimador de *Hodges-Lehmann* se define como la mediana de sus diferencias: $med \{ (Y_j - X_i) : 1 \leq i \leq n_1, 1 \leq j \leq n_2 \}$.

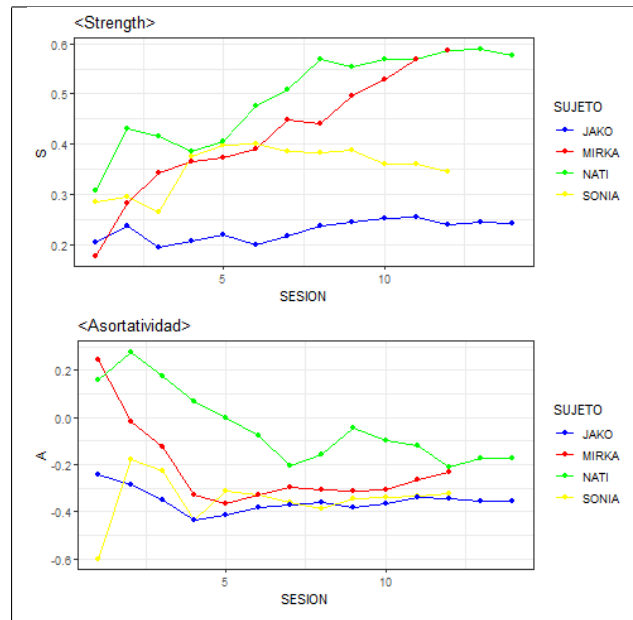


Figura 11.5: Evolución de “Strength medio” y assortatividad.

11.3. Análisis estadístico de evolución temporal.

Se tomaron, de las redes anteriores, solamente los casos de concordancia de error, es decir que se trata ahora de redes de crecimiento de error. Se agruparon las sesiones en tres etapas (de crecimiento acumulado) según el siguiente esquema:

1. SONIA y MIRKA: (i) hasta la sesión 4, (ii) hasta la sesión 8, (iii) hasta la sesión 12.
2. NATI y JAKO: (i) hasta la sesión 5, (ii) hasta la sesión 9, (iii) hasta la sesión 14.

También eran grafos dirigidos. Sin embargo, ahora la dirección indicaba el *origen* del error (Ej.: “la casas” recibiría la representación $las \rightarrow casas$, mostrando que el error se ubica en el artículo). Si el error estaba en ambos términos entonces la flecha era bidireccional. La red no tenía pesos en las aristas. Los atributos de los nodos se mantuvieron como indicado en el Cuadro 2. La Figura 6 muestra las redes de error hasta la última sesión, para cada alumno. Observar que las redes mantienen constantes los nodos / palabras y lo que se agregan son las aristas.

Se aplicó el llamado *Actor-Based Stochastic Model* de Snijders (1996; Snijders et al., 2007, 2010), un modelo estadístico de evolución temporal de redes.

El modelo trata a cada nodo de una red como un actor. Ya que los enlaces de la red se tratan como si fueran estados, se modeló el cambio como un proceso de Markov (en cada punto del tiempo el estado actual de la red determina probabilísticamente el estado de la red en el instante sucesivo, sin influencia de estados pasados). El proceso es continuo, es decir que la red observada en cada punto es el resultado de micro-cambios continuos realizados en mini-pasos. En cada mini-paso, dicho actor debe decidir si preserva, elimina o establece enlaces con otros actores. Este cambio se halla influenciado por factores endógenos a la red (por ejemplo, la tendencia a la reciprocidad de los enlaces: “si soy tu amigo, sos mi amigo”); o exógenos a ésta: covariables del actor (los atributos del nodo) o covariables diádicas (interacciones entre atributos de dos nodos conectados por un enlace, como presencia de la característica en ambos nodos). Además en la evolución temporal pueden cambiar características de los actores influenciados por la estructura de la red². Por lo tanto se pueden seguir dos dinámicas: una de la red y otra de características de comportamiento de los actores. Las relaciones entre los actores están representadas en la matriz de adyacencia $X_{n \times n} = (x_{ij})$ (n actores); con $x_{ij} = 1$ si hay un enlace entre los actores i y j , y $x_{ij} = 0$ si no lo hay³. Los

²Un estudiante que fuma puede ser amigo de otro que fuma [homofilia]; o bien no fumar inicialmente pero comenzar a hacerlo si la mayoría de su grupo de amigos es fumador [Asimilación].

³Redes con pesos categóricos se tratan como modelos con redes separadas.

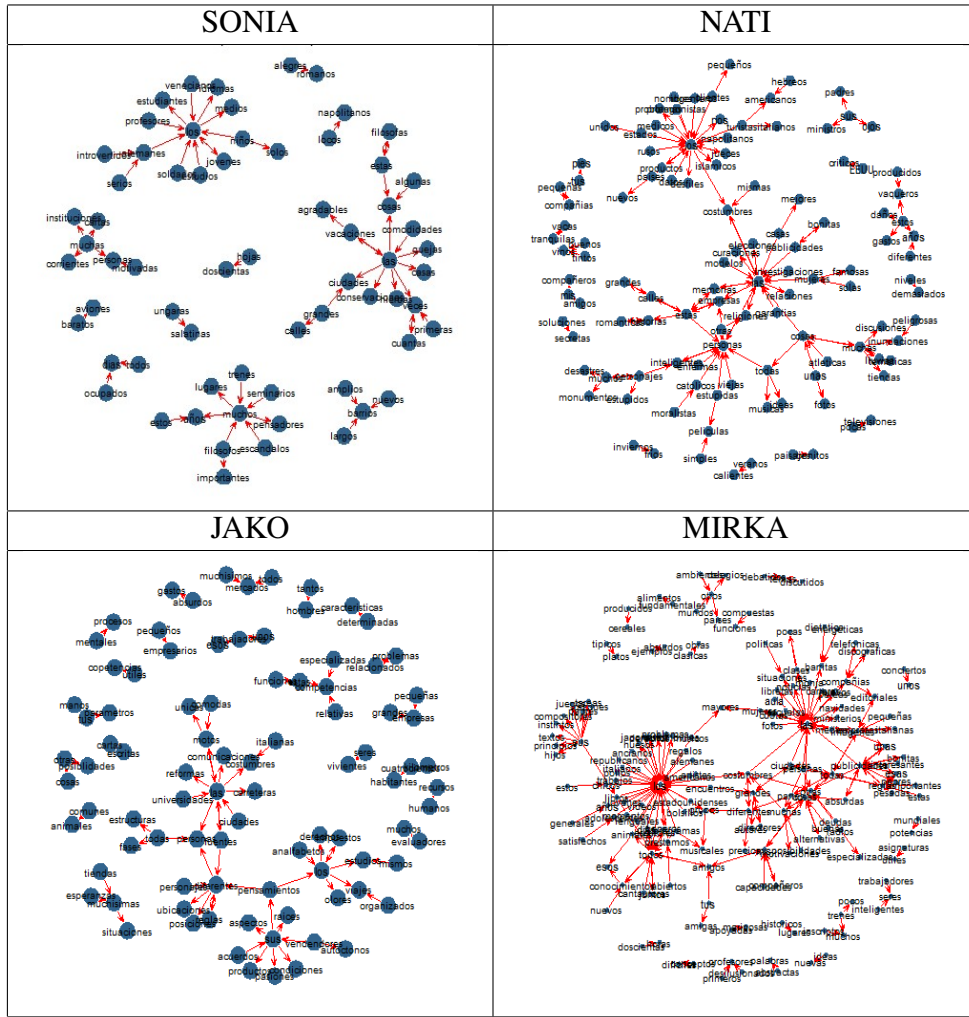


Figura 11.6: Redes de crecimiento de error.

atributos de comportamiento de los actores deben ser categóricos u ordinales y se los denotará como z_{hi} ($h = 1, \dots, H$), el valor del h -ésimo atributo del i -ésimo actor. Por lo tanto, cada actor tendrá las observaciones $Y(t) = (x_{ij}(t), z_{1i}(t), \dots, z_{Hi}(t))$, en cada tiempo discreto t . El proceso de cambio se descompone en dos sub-procesos. El primero modela la frecuencia con la que suceden los cambios (de enlaces y del comportamiento de los actores) en los mini-pasos. Es una función de tasa de arribo que depende el efecto de la tasa del periodo m , de efectos de covariables del actor y de efectos de posición del actor en la red (por ejemplo, el grado). Son: $\lambda_i^X(Y, m) = \rho_m^X \left\{ \exp \left(\sum_k \alpha_k^X a[x]_{ki}(Y(t)) \right) \right\}$; $\lambda_i^{Z_h}(Y, m) = \rho_m^{Z_h} \left\{ \exp \left(\sum_k \alpha_k^{Z_h} a[Z_h]_{ki}(Y(t)) \right) \right\}$. Por tanto, los tiempos de espera entre d mini-pasos de cada actor sigue una distribución exponencial de parámetro λ_i ($d \sim \varepsilon(\lambda_i)$). Si hay tasa constante: $\lambda_i = \rho, \forall i$. Por otra parte, se modela el subproceso de cambio determinístico; es decir, en cuál enlace o atributo comportamental se produce el cambio. Los atributos de comportamiento cambian siempre en una unidad (crecen o decrecen). La función se descompone en tres: una función de evaluación, una función de “satisfacción” y un término de error (modelado a partir de la distribución de Gumbel): $f_i + g_i + \varepsilon_i$. La segunda función g_i penaliza la pérdida de un enlace o el decrecimiento de un valor comportamental como pérdida de satisfacción. No será usado dicho componente en el análisis. La parte fundamental es la función de evaluación f_i ; y especifica los efectos de estructura o de comportamiento. El actor seleccionará el cambio que maximiza dicha función. Al igual que con la función de tasa de arribo, hay dos; una para cada dinámica:

- $f_i^X(\beta, x) = \sum_{k=1}^n \beta_k^X s_{ik}^X(x)$;
- $f_i^{Z_h}(\beta, z) = \sum_{k=1}^n \beta_k^{Z_h} s_{ik}^{Z_h}(x, z)$;

donde: β_k son parámetros, $s_{ik}(x)$ son funciones que dependen de la estructura de la red y $s_{ik}(x, z)$ son funciones que dependen del comportamiento del actor pero también del comportamiento o características estructurales (de red) de sus vecinos.

Por fin, la probabilidad de que un actor cambie de estado (enlace o comportamental) se define como:

$$P_{ij}(x(i \rightarrow j), x(t), z(t)) = \frac{\exp\{[f+g]_i^X(\beta^X, \gamma^X, x(i \rightarrow j)(t))\}}{\sum_k \exp\{[f+g]_i^X(\beta^X, \gamma^X, x(i \rightarrow k)(t))\}};$$

$$P_{i\delta}(z(i \uparrow_h \delta) | x(t), z(t)) = \frac{\exp\{[f+g]_i^Z(\beta^{Z_h}, \gamma^{Z_h}, z(i \uparrow_h \delta)(t))\}}{\sum_{\tau \in \{-1, 0, 1\}} \exp\{[f+g]_i^Z(\beta^{Z_h}, \gamma^{Z_h}, z(i \uparrow_h \delta \tau)(t))\}}$$

A continuación se describen los efectos de estructura incluidos en el modelo que nos ocupa.

1. *Outdegree-Popularity*. Es el número de enlaces salientes de los nodos *conectados* a un nodo *i*-ésimo.
 $s_{i1}(x) = \sum_j x_{ij} x_{j+}; x_{j+} = \sum_h x_{jh}$.
2. *Indegree-Popularity*. Es el número de enlaces entrantes de los nodos *conectados* a un nodo *i*-ésimo.
 $s_{i2}(x) = \sum_j x_{ij} x_{+j}; x_{+j} = \sum_h x_{hj}$.
3. *Reciprocity*. Es el número de enlaces recíprocos entre dos nodos. $s_{i3}(x) = \sum_j x_{ij} x_{ji}$.
4. *Anti isolates*. Efecto de conexión con otros nodos, los cuales estarían aislados sin dicha conexión.
 $s_{i4}(x) = \sum_j I\{x_{+j} \geq 1, x_{j+} = 0\}$.
5. *Isolate - popularity*. Efecto de estar conectado con actores cuyo grado de enlaces entrantes es 1.
 $s_{i5}(x) = \sum_j x_{ij} I\{x_{+j} = 1, x_{j+} = 0\}$.
6. *out-in degree^(1/2) assortativity*. La tendencia de que los nodos con alto grado de enlaces salientes estén conectados a nodos con alto grado de enlaces entrantes. $s_{i6}(x) = \sum_j x_{ij} x_{i+}^{1/2} x_{+j}^{1/2}$.
7. *Balance*. Similaridad entre los enlaces *salientes* del nodo *i*-ésimo y los enlaces *salientes* de los otros nodos con los cuales se conecta dicho nodo *i*-ésimo. $s_{i7}(x) = \sum_{j=1}^n x_{ij} d_{ij}; d_{ij} = \sum_{h=1; h \neq i, j}^n (b_0 - |x_{ih} - x_{jh}|)$.
8. *In-struct equivalence*. Similaridad entre los enlaces *entrantes* del nodo *i*-ésimo y los enlaces *entrantes* de los otros nodos con los cuales se conecta dicho nodo *i*-ésimo. $s_{i8}(x) = \sum x_{ij} d_{ij}; d_{ij} = \sum_{h=1; h \neq i, j}^n (b_0 - |x_{hi} - x_{hj}|)$.
9. *SameX*. Mismo comportamiento entre los actores (nodos) *i* y *j*. $s_{i9}(x) = \sum_j x_{ij} I(v_i = v_j)$. [homofilia respecto de atributos cualitativos]
10. *sameXRecip*. Mismo comportamiento (homofilia) en enlaces bidireccionales: $s_{i10}(x) = \sum_j x_{ij} x_{ji} I(v_i = v_j)$.
11. *SimX*. Suma de las similitudes centradas (respecto de un determinado atributo *v*: sim_{ij}^v) entre *i* y los otros actores *j* con los que está conectado. $s_{i11}(x) = \sum_j x_{ij} (sim_{ij}^v - \widehat{sim}^v)$. [Asortatividad para atributos cuantitativos].
12. *SimRecipX*. Similaridad respecto de enlaces bidireccionales. $s_{i12}(x) = \sum_j x_{ij} x_{ji} (sim_{ij}^v - \widehat{sim}^v)$.
13. *EgoX*. El número de enlaces salientes pesados por un valor de covariable del actor. $s_{i13}(x) = v_i x_{i+}$.
14. *AltX*. La suma de la covariable sobre todos los actores con los cuales *i* tiene un enlace. $s_{i14}(x) = \sum_j x_{ij} v_j$.
15. *EffFrom*. Dependencia de un atributo de otras variables de comportamiento h' ($h \neq h'$). $s_{i1}(x, z_h) = z_{ih} z_{ih'}$.

Entonces, las funciones de evaluación para la dinámica de la red y la del comportamiento se definen como:

$$f_i^X(\beta, x) = \exp\{\beta_1 s_{i1}(x) + \beta_2 s_{i2}(x) + \beta_3 s_{i3}(x) + \beta_4 s_{i4}(x) + \beta_5 s_{i5}(x) + \beta_6 s_{i6}(x) + \beta_7 s_{i7}(x) + \beta_8 s_{i8}(x) + \left(\sum_v \beta_9 s_{i9}(x) + \beta_{10} s_{i10}(x) + \beta_{11} s_{i11}(x) + \beta_{12} s_{i12}(x) + \beta_{13} s_{i13}(x) + \beta_{14} s_{i14}(x)\right)\}$$

$$f_i^{Z_h}(\beta, z_h) = \exp\left\{\sum_{z_h} \beta_1 s_{i1}(x, z_h)\right\}$$

Las tasas de arribo son constantes (sin efectos de covariables o de estructura): $\lambda_i^X = \rho_X$; $\lambda_i^{Z_h} = \rho_{Z_h}$.

Para la dinámica de comportamiento se considera la cantidad de errores de una palabra / nodo⁴. Se utilizó *RSiena* (Ripley et al., 2018). Se testearon las siguientes hipótesis:

- (i) El nodo (palabra) es origen de error, según un determinado atributo [*EgoX*].
- (ii) El nodo (palabra) no es origen de error, según un determinado atributo [*AltX*].
- (iii) Un determinado atributo es similar en ambos nodos (palabra) [*SimX*, *SameX*].
- (iv) El error es recíproco, según un determinado atributo [*SimRecipX*, *sameXRecip*].
- (v) La acumulación de errores en una palabra crece, según un determinado atributo [*EffFrom*].

Los Cuadros 6 a 9 muestran los resultados significativos del modelo estimado, para cada aprendiz⁵. Las palabras de controlador con alta imaginabilidad y concretud; así como aquellas que funcionan como adjetivos y las que conllevan -e- epentética tienden a ser origen del error. Las palabras con controlador de familiaridad / frecuencia léxica alta tienden tanto a generar como a recibir errores. Por otra parte, las palabras terminadas en “es” (ej.: “cantantes”, “amables”), las que conllevan -e- epentética y las que tienen alta similaridad de raíz / desinencia con el italiano tienden a tener valores similares con el otro término de la concordancia errónea. Solamente en el caso de MIRKA se halló un efecto de influencia en la acumulación de error, respecto de controladores con valores imaginabilidad / concretud altos. No hubo efectos significativos de reciprocidad de errores. El Cuadro que sigue resume lo apuntado:

Hipótesis	SONIA	NATI	JAKO	MIRKA
I	FAM.LEX	MOD=”4”; ES	IMA.CONC	-
II	FAM.LEX	-	-	-
III	ES; ESP=”es”; STEM	ESP=”es”	MORF	ESP=”es”; ES
IV	-	-	-	-
V	-	-	-	IMA.CONC

Cuadro 11.6: Resumen de resultados para cada hipótesis.

Una manera de evaluar la bondad de ajuste del modelo consiste en comparar el modelo observado con otros resultados posibles simulados a partir del modelo ajustado. Se simularon 1000 redes generadas a partir de los parámetros del modelo y se calculó para cada una la probabilidad de cada grado entrante [“in degree”], tomando de *cero* a *tres* grados. Los resultados se muestran mediante un “violin plot” para cada grado mostrando la distribución sobre los mil valores obtenidos. El punto rojo indica el valor observado. El p-valor indica la probabilidad de observar la distribución de grado observada en las mil simulaciones. Cuanto más alto, mejor ajuste (o sea que la hipótesis nula es “el modelo ajusta bien” y *no* debe rechazarse [$p < 0,05$]). Se puede notar que ningún modelo ajustado rechaza la hipótesis nula [Figura 7].

⁴Los nodos que todavía no son parte de la sesión tienen valor faltante “NA”.

⁵Se considera una convergencia aceptable para el ajuste de cada coeficiente si $Conv < 0,1$ y el ratio de convergencia máximo del modelo es menor a 0,25

	Function	Effect	Coef	StEr	Pval	Sig	Conv
1	red	constant red rate (period 1)	1.162	0.239	p<0.001	***	0.086
2	red	constant red rate (period 2)	0.238	0.103	0.021	*	0.045
3	red	reciprocity	3.366	1.483	0.0232	*	0.070
4	red	indegree - popularity	0.761	0.249	0.0022	**	0.019
5	red	anti isolates	3.938	1.429	0.0058	**	0.007
6	red	stem similarity	2.256	1.056	0.0326	*	0.075
7	red	same esp_3	4.235	1.705	0.013	*	0.060
8	red	same es	1.873	0.903	0.038	*	0.089
9	red	fam.lex alter	0.695	0.329	0.0343	*	0.001
10	red	fam.lex ego	12.922	3.065	p<0.001	***	0.058
11	error	rate error (period 1)	0.604	0.206	0.0034	**	0.011
12	error	rate error (period 2)	0.310	0.102	0.0025	**	0.035

Cuadro 11.7: SONIA. Maximum Convergence Ratio: 0.236

	Function	Effect	Coef	StEr	Pval	Sig	Conv
1	red	constant red rate (period 1)	2.741	0.783	5e-04	***	0.076
2	red	constant red rate (period 2)	0.591	0.202	0.0034	**	0.018
3	red	reciprocity	2.628	0.517	p<0.001	***	0.038
4	red	outdegree - popularity	0.625	0.121	p<0.001	***	0.036
5	red	isolate - popularity	2.478	0.536	p<0.001	***	0.009
6	red	mod_4 ego	1.569	0.699	0.0248	*	0.050
7	red	same esp_3	1.748	0.523	8e-04	***	0.060
8	red	es ego	45.032	4.495	p<0.001	***	0.008
9	error	rate error (period 1)	1.401	0.249	p<0.001	***	0.031
10	error	rate error (period 2)	0.355	0.085	p<0.001	***	0.002

Cuadro 11.8: NATI. Maximum Convergence Ratio: 0.126

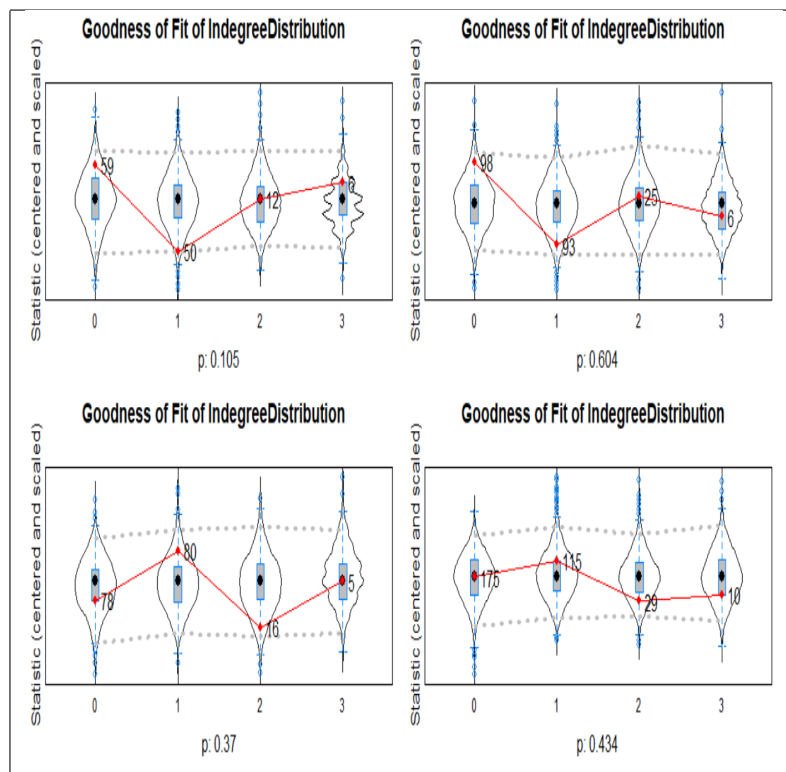


Figura 11.7: Bondad de Ajuste de los modelos ajustados.

	Function	Effect	Coef	StEr	Pval	Sig	Conv
1	red	constant red rate (period 1)	1.523	0.485	0.0017	**	0.017
2	red	constant red rate (period 2)	0.359	0.120	0.0027	**	0.009
3	red	reciprocity	2.575	0.680	2e-04	***	0.009
4	red	out-in degree ^(1/2) assortativity	2.411	0.440	p<0.001	***	0.029
5	red	in-struct equivalence	0.683	0.353	0.0532	.	0.065
6	red	morf similarity	2.672	1.000	0.0075	**	0.013
7	red	ima.conc ego	3.288	0.881	2e-04	***	0.005
8	error	rate error (period 1)	0.556	0.168	9e-04	***	0.006
9	error	rate error (period 2)	0.184	0.068	0.0071	**	0.036

Cuadro 11.9: JAKO. Maximum Convergence Ratio: 0.107

	Function	Effect	Coef	StEr	Pval	Sig	Conv
1	red	constant red rate (period 1)	0.849	0.113	p<0.001	***	0.015
2	red	constant red rate (period 2)	0.893	0.104	p<0.001	***	0.086
3	red	reciprocity	3.309	0.387	p<0.001	***	0.007
4	red	balance	0.533	0.142	2e-04	***	0.014
5	red	anti in-isolates	0.769	0.451	0.0884	.	0.023
6	red	same esp_3	2.728	0.864	0.0016	**	0.007
7	red	same es	0.801	0.295	0.0065	**	0.006
8	red	ima.conc alter	0.212	0.136	0.1192		0.008
9	error	rate error (period 1)	0.568	0.122	p<0.001	***	0.025
10	error	rate error (period 2)	0.821	0.146	p<0.001	***	0.042
11	error	error: effect from ima.conc	0.479	0.206	0.0199	*	0.029

Cuadro 11.10: MIRKA. Maximum Convergence Ratio: 0.387

11.4. Predicción de estatus de error.

En esta sección se intenta predecir el estatus de error de concordancia [correcto / incorrecto] en las últimas cuatro sesiones de cada aprendiente a partir de predictores que involucran características de las concordancias, dinámica de error y atributos de los grafos de concordancia ya descriptos. Específicamente, los predictores fueron los siguientes, calculados con las concordancias pertenecientes a la *componente gigante* de los grafos descriptos en la sección 11.2.

Atributos de las concordancias. Los atributos utilizados que describen las concordancias (ya descriptos) son:

"MOD", "LDA", "GRAMS", "ES", "MORF.f", "STEM.f", "CUMRES", "EST1", "EST2", "EST3", "EST4", "EST5", "EST6", "EST7", "IMA.CONC.f", "FAM.LEX.f", "ANIM", "ESP1", "ESP2".

Atributos basados en los grafos. Los siguientes atributos se calcularon mediante la librería de R “*link-prediction*” (Bojanowski & Chrol, 2019). Sean: (i) $A = [a_{xy}]$: la matriz de adyacencia del grafo; (ii) n : el número de nodos; (iii) k_x : el grado del nodo x ; (iv) $paths_{xy} < l >$: conjunto de todos los caminos de largo l del nodo x al nodo y ; (v) D , la matriz diagonal de los grados; (vi) $L = D - A$: el laplaciano del grafo; (vii) L^+ : la pseudo-inversa de Moore-Penrose del laplaciano; (viii) λ_1 : el primer autovalor de la matriz A ; (ix) m : el número de enlaces. Se definen entonces las siguientes medidas, que caracterizan los enlaces del grafo (o sea, las concordancias) en términos de la proximidad / similaridad $S = [s_{xy}]$ entre los nodos conectados por el enlace.

- “preferential attachment” [pa]: $s_{xy} = k_x \times k_y$;

- “Kantz index” [*ka*]: Conteo de todos los caminos que pasan por un determinado par de nodos. $S = (I - \beta A)^{-1} - I$, donde el parámetro β debe satisfacer que: $\beta < \frac{1}{\lambda_1}$.
- “L cosine” [*cos_l*]: Coseno del ángulo entre las columnas de L^+ . $s_{xy} = \frac{l_{xy}^+}{\sqrt{l_{xx}^+ l_{yy}^+}}$.
- “laplaciano inverso” [*l*]: $S = L^+$ también proporciona una medida de proximidad entre los nodos.
- “random walk with restart” [*RWR*]: considérese un caminante aleatorio que comienza en el nodo x y regresa al mismo nodo con probabilidad α . Sea q_x la distribución estacionaria de una *cadena de Markov*, es decir: $q_x = (1 - \alpha) P^T q_x + \alpha e_x$; donde: e_x es un vector con 1 en la posición del nodo x y P es la matriz de transición: $P_{xy} = \frac{1}{k_x}$ si $A_{xy} = 1$. La solución para todos los nodos en simultáneo reza: $q = \alpha (I - (1 - \alpha) P^T)^{-1}$. Con lo cual se define: $s_{xy} = q_{xy} + q_{yx}$.
- “Matrix Forest Index” [*mf*]: $S = (I + L)^{-1}$.
- “Average Commute Time” [*act*]: Sea $r(x, y)$ el número de pasos promedio necesario para que un caminante aleatorio vaya del nodo x al nodo y . Para lograr simetría la medida se toma en ambas direcciones de los enlaces: $n(x, y) = r(x, y) + r(y, x) = 2m (l_{xx}^+ + l_{yy}^+ + 2l_{xy}^+)$. Entonces: $s_{xy} = \frac{1}{n(x, y)}$.
- “Normalized Average Commute Time” [*act_n*]: Medida anterior pero ahora tomando en cuenta el grado de los nodos. $s_{xy} = \frac{1}{m(x, y)\pi_y + m(y, x)\pi_x}$; donde π es la distribución estacionaria de una *cadena de Markov* que describe un caminante aleatorio en el grafo.

También se incluyó la *frecuencia del enlace* (cuántas veces aparece una determinada concordancia).

Atributos dinámicos. Se calculó la cantidad de errores acumulados hasta la instancia de concordancia anterior a la instancia n -ésima. Sobre dicho vector se aplicó una ventana móvil de largo⁶ $w = 45$ y se calcularon las medidas que se describen a continuación.

- Atributos basados en estadística descriptiva⁷. Se calcularon los siguientes estadísticos sobre la distribución de valores de la serie: (a) mediana; (b) media; (c) media podada (5%); (d) desvío estándar; (f) kurtosis; (g) asimetría; (h) coeficiente de variación; (i) valor máximo; (j) valor mínimo; (k) «*Burtisness*»: $(\sigma - \mu) / (\sigma + \mu)$; (l) «*CV_diff*»: el desvío estándar de la primera diferencia de la serie: $\sigma(\text{diff}(x))$; (m) «*Ben_dist*»: la media de la porción de la serie que está arriba de la media general relativa a la porción que está debajo de la media general: $\mu(x_i > \mu) / \mu(x_i < \mu)$; (n) “lumpiness”: varianza de las varianzas; (ñ) “stability”: varianza de las medias; “FE”: tiempo hasta el valor cuya primera diferencia no es nula; (o) “crossing_points”: número de veces en que la serie cruza su valor medio; (p) “Max_level_shift”: el cambio en media más grande entre dos ventanas consecutivas.
- Atributos basados en autocorrelación: Se calcula la función de autocorrelación «ACF» (detecta correlaciones lineales) y se calcula su valor en el primer *lag* y el tiempo hasta que se produce el primer mínimo y se cruza el primer *cero*.
- “Hurst”: índice H de dependencia a largo plazo de la serie (“memoria”); $H = 0$ implica incrementos independientes y $H \in [0, 5, 1]$ implica “memoria” en la serie.
- “Motiv”: Se binariza la serie de manera que los valores mayores a la media son iguales a *uno* y los menores son *cero*. Luego se calcula la entropía de las “palabras” de largo igual a *tres*.

⁶Es decir que las primeras 44 observaciones recibieron datos faltantes.

⁷Tanto estos atributos como varios de los que siguen fueron implementaciones hechas en R de los scripts para *Matlab* propuestos en: www.comp-engine.org. También se usó la librería de R “tsfutures”.

- **Atributos basados en RQA:** Se calculan algunas medidas de *Recurrence Quantification Analysis* (Webber & Marwan, 2015), una técnica que permite medir propiedades de atractores en el espacio de fase [phase space] reconstruido. RQA mide el grado de nuevas visitas del sistema en estas regiones del espacio de fase y computa ciertas medidas basadas en dichas recurrencias. Específicamente se calculan: (a) recurrencia (“RR”): porcentaje de puntos que recurren; (b) determinismo (“DET”): mide la predictibilidad del sistema; (c) Trapping time (“TT”): tiempo promedio en el cual el sistema se encuentra atrapado en un mismo estado de recurrencias; (d) Laminaridad (“LAM”): relacionada con segmentos de tiempo en los cuales no hay intermitencia en el sistema [ver cap. 10].
- **Atributos basados en Entropía de Permutación:** Con la función «PE» se calculó una medida de entropía (H) y otra de complejidad (C) para la serie. La entropía de permutación fue originariamente propuesta por Bandt & Pompe (2002) [ver cap. 10].
- **Atributos basados en «onditas» (wavelets):** Se aplicó el algoritmo *wavMODWT* (librería de R: *wmtsa*), forma no-decimada de la transformada discreta de *wavelets* (Percival & Walden, 2008; Nason, 2008). Dados $n = 2^J$ observaciones ($J = \log_2(n)$), se obtienen como resultado los coeficientes s_{jk} y d_{jk} para los niveles $j = J, J-1, \dots, 1$ y los valores de cada nivel $k = 1, \dots, 2^{j-1}$. Los s_{jk} son los coeficientes de escala (o padre), en cambio los d_{jk} son los coeficientes de *wavelets madre* (aquí se usa el *Haar*) y se denominan «detalle» [*detail*] porque representan datos de más alta frecuencia. Formalmente:

$$d_{j,k} = \sum_{l=-\infty}^{\infty} g_l y_{2k-l}$$

$$\text{donde: } y = (y_1, y_2, \dots, y_n) \text{ es el vector de datos y } g_l = \begin{cases} 2^{-1/2} & \text{si } l = 0 \\ -2^{-1/2} & \text{si } l = 1, \\ 0 & \text{si no} \end{cases}$$

igualmente para los coeficientes de escala:

$$s_{j,k} = \sum_{l=-\infty}^{\infty} h_l y_{2k-l}, \text{ donde: } h_l = \begin{cases} 2^{-1/2} & \text{si } l = 0 \\ 2^{-1/2} & \text{si } l = 1 \\ 0 & \text{si no} \end{cases}$$

A partir del conjunto de coeficientes para cada nivel, se calculó la energía y la varianza explicada (en escala logarítmica).

Frecuencia de nodos que intervienen en los enlaces de la red. Se creó una matriz $W_{d \times t}$ de filas t = cantidad de nodos (palabras) de la red y columnas d = cantidad de concordancias (enlaces de la red). En dicha matriz la celda $w_{t,d}$ contaba la frecuencia de la palabra t en la concordancia d . Se normalizaron las celdas según el esquema:

$$w_{t,d} = tf \times idf = tf_{t,d} \times \log_2 \left(\frac{N}{df_t} + 1 \right)$$

Donde: N es el número de documentos, df_t es el número de concordancias que contienen el término t , $tf_{t,d}$ es el número de ocurrencias del término t en la concordancia d . Se tomaron las 50 palabras más frecuentes como predictores.

En total hubo 126 predictores. El siguiente Cuadro muestra el total de observaciones asignadas al conjunto de entrenamiento (70% de los datos) y al de validación (30% de los datos), para cada alumno. Los datos faltantes de las primeras 44 observaciones de los predictores “dinámicos” fueron reemplazados por *cero*. Los atributos continuos fueron estandarizados ($(x - \mu_x) / \sigma_x$). La variable respuesta fue $y \in \{0, 1\}$ donde 1 era concordancia «incorrecta» y 0, «correcta». Entre paréntesis se detalla la cantidad de observaciones por categoría de respuesta.

	SONIA	NATI	JAKO	MIRKA
ENTRENAMIENTO	207 (0: 158, 1: 49)	278 (0: 178, 1: 100)	336 (0: 276, 1: 60)	438 (0: 322, 1: 116)
VALIDACIÓN	96 (0: 83, 1: 13)	105 (0: 81, 1: 24)	115 (0: 104, 1: 11)	247 (0: 163, 1: 84)
TOTAL	303	383	451	685

Cuadro 11.11: Conjuntos de entrenamiento y validación para cada alumno.

El desbalance de clase muestra la relación de la clase minoritaria respecto a la mayoritaria ($IR = \frac{\#clase1}{\#clase0} \in [0, 1]$, $IR = 1$ si están balanceadas). Para el conjunto de entrenamiento es: (i) SONIA: $IR \approx 0,31$; NATI: $IR \approx 0,56$; JAKO: $IR \approx 0,21$; MIRKA: $IR \approx 0,36$. Ya que el desbalance puede afectar el desempeño de los clasificadores, se usaron las siguientes estrategias para corregir el desbalance (o sea obtener $IR \approx 0,5$ para el conjunto de entrenamiento): (i) “undersampling” [US]: tomar al azar un subconjunto de observaciones de la clase *mayoritaria*; (ii) “oversampling” [OS]: duplicar al azar un conjunto de observaciones de la clase minoritaria; (iii) “weighting” [W]: pesar las clases (aumentando la importancia de la categoría minoritaria); (iv) “synthetic minority over-sampling” (SMOTE): generar observaciones de la clase minoritaria como combinaciones aleatorias convexas de los vecinos de las observaciones.

Se aplicaron los siguientes clasificadores: (i) *Support Vector Machine* [SVM]; (ii) *Random Forest* [RF]; (iii) *Recursive Partitioning and Regression Trees* [RPART]; (iv) *Gradient Boosting Machine* [GBM]; (v) *eXtreme Gradient Boosting* [XGB]; (vi) *logistic regression* [LOGREG].

En el Cuadro que sigue se indica el rango de los parámetros de cada uno para afinar. También se indica el rango de afinación de los parámetros de los métodos de corrección de desbalance. Cada clasificador se aplicó tanto con dichos métodos de corrección como sin ellos.

Método	Parámetros. Rango: (inferior, superior)
SVM	$C = (2^{-8}, 2^{15})$; $\sigma = (2^{-15}, 2^3)$
RF	$mtry = (3, 10)$; $ntree = (50, 500)$; $nodesize = (10, 50)$
RPART	$cp = (0,0001, 0,1)$; $minsplit = (1, 10)$; $minbucket = (5, 50)$
XGB	$nrounds = (200, 600)$; $maxdepth = (3, 20)$; $\lambda = (2^{-10}, 2^{10})$
GBM	$n.trees = (100, 5000)$; $interaction.depth = (2, 10)$; $bag.fraction = (0,7, 1)$ $shrinkage = (0,001, 0,5)$; $n.minobsinnode = (5, 15)$
LOGREG	–
US	$rate = (0,6 \times IR)$
OS	$rate = (1,5 \times IR^{-1})$
W	$weight\ class\ 1 = (1, 10)$; $weight\ class\ 0 = 1$
SMOTE	$rate = (1,5 \times IR^{-1})$

Cuadro 11.12: Parámetros de los métodos de clasificación y de los de corrección de desbalance.

Se llevó a cabo una primera selección de 80 atributos por medio de un *ensemble* de medidas de selección, a saber:

1. Área bajo la curva ROC (AUC): Gráfica en la que el eje «y» contiene la proporción de verdaderos positivos y el eje «x», la proporción de falsos positivos. Si el predictor separara perfectamente las clases habría un corte de probabilidad de asignación de la clase cuyo valor de verdaderos positivos sería 1 y de falsos positivos sería 0; por lo tanto, el área total bajo la curva sería $AUC = 1$ también. Por otro lado, si un predictor no separa bien las clases, el área se acercaría a $AUC = 0,50$. Se debe maximizar.
2. Relief Score (Kira & Rendell, 1992; ver también: Kuhn & Johnson, 2013, cap. 18) : el algoritmo selecciona un punto al azar y luego busca los puntos más cercanos de ambas clases (denominados «error» [miss] y «acierto» [hit]). Luego se calcula una medida de *score* de la diferencia entre el punto elegido y los errores y aciertos. Esto se repite con m puntos y el *score* general es una suma de estas

diferencias. La intuición radica en que un predictor que separa bien las clases tendrá los «aciertos» cerca y los «errores» lejanos; por eso, *scores* generales altos indicarán predictores influyentes.

3. *Pesos de Random Forest* [RF]: Se calculan los pesos para cada atributo utilizando un modelo de *Random Forest*.
4. *praznik_JMI*: La Información mutua indica la reducción en la incertidumbre de una variable si se sabe el valor de otra⁸: $I(X, Y) = H(X) - H(X | Y)$. El método selecciona la variable de máxima información mutua entre un predictor X y la variable respuesta Y . Luego se añaden uno a uno el resto de los predictores X tratando de maximizar la función: $J(X) = \sum_{W \in S} I(X, W; Y)$, donde S es el conjunto de las predictoras ya seleccionadas.

Si se tiene la tabla de confusión, donde: TP = “true positives”; TN = “true negatives”; FP = “false positives” [falsa alarma]; FN = “false negatives” [“miss”: negativos clasificados como positivos]; se pueden derivar las medidas descritas en el Cuadro 6, utilizadas para verificar el desempeño de los modelos entrenados con el conjunto de evaluación. A éstas, se agregó el área bajo la curva ROC [*auc*] y una medida de costo [c] de clasificación. Esta última multiplica por un peso w a cada observación predicha (1 ó 0) y luego saca un promedio. Las matrices de costo para cada alumno fueron las que siguen, según el arreglo:

OBS / PRED	0	1
0	0	costo
1	costo	costo

SONIA: $\begin{bmatrix} 0 & 1 \\ 6 & 0 \end{bmatrix}$; NATI: $\begin{bmatrix} 0 & 2 \\ 6 & -2 \end{bmatrix}$; JAKO: $\begin{bmatrix} 0 & 1 \\ 10 & 0 \end{bmatrix}$; MIRKA: $\begin{bmatrix} 0 & 1 \\ 2 & -1 \end{bmatrix}$.

Medidas	Fórmula	Definición
balanced accuracy (bac)	$\frac{TP}{FP+FN} + \frac{TN}{FP+TN} / 2$	proporción de clasificados correctamente del total (para datos desbalanceados)
sensitivity (sens) [recall]	$\frac{TP}{TP+FN}$	Proporción de positivos clasificados correctamente y que lo son en el patrón de referencia
positive predictive value (ppv) [precisión]	$\frac{TP}{TP+FP}$	Proporción de positivos clasificados correctamente (TP) de todos los clasificados como positivos (TP+FP)
F1	$2 \times \frac{\text{precision} \times \text{recall}}{\text{precision} + \text{recall}}$	media armónica entre precisión y <i>recall</i>
False positive rate	$\frac{FP}{FP+TN}$	probabilidad de falsa alarma
False negative rate	$\frac{FN}{FN+TP}$	probabilidad de “miss”

Cuadro 11.13: Medidas derivadas de la tabla de confusión.

A continuación se aplicó, para cada método, una segunda selección de atributos (en el rango [1, 80]) basado en la medida *praznik_JMI* en conjunto con el entrenamiento de cada modelo de clasificación. El conjunto de entrenamiento se dividió a su vez en dos conjuntos de entrenamiento y validación aplicando *validación cruzada* con cinco *folds*. Se evaluó el desempeño de los modelos mediante la minimización de: (i) tasa de falsos positivos; (ii) tasa de falsos negativos; (iii) medida de costo de clasificación.

Usualmente si $P(x = \text{“correcto”}) > 0,5$ se le asigna la categoría 1 = *correcto* y si no la de 0 = *incorrecto*; sin embargo, aquí se usó la medida de costo para optimizar el punto de corte de la probabilidad para la decisión de asignar los valores a la categoría “correcto” u “error”.

Se llevó a cabo el análisis usando la librería *mlr* de *R* (Bischl et al., 2016). Los Cuadros 15 a 18 muestran los resultados de los *mejores* clasificadores obtenidos, por alumno. La fila denotada “ensemble” se refiere a un clasificador que usa como predictores los valores predichos del resto de los clasificadores de la tabla. Las filas se hallan ordenadas según los valores de *balanced accuracy* [bac].

⁸Recuérdense las definiciones de: (i) entropía (incertidumbre en una variable): $H(X) = \sum_{x \in X} p(x) \log [p(x)]$; (ii) entropía condicional (incertidumbre de una variable sabiendo el valor de otra): $H(X | Y) = \sum_{x \in X} \sum_{y \in Y} p(x, y) \log [p(x | y)]$.

En general no se logró superar el 80% de *accuracy*. En el caso de SONIA los dos mejores clasificadores resultaron ser SVM con *undersampling* y RF con *smote*. Para NATI, RPART con *class weighting* y SVM con *oversampling*. Respecto de JAKO, SVM con *undersampling* y XGB con *class weighting*. En lo que atañe a MIRKA, fueron SVM con *oversampling* y RPART con *smote*. Con lo cual SVM resultó el clasificador de mejor desempeño. SONIA y JAKO alcanzaron mejores niveles de *accuracy* que NATI y MIRKA. Por otro lado, el “ensemble”, aunque fue el valor más alto de *accuracy* en el caso de SONIA y MIRKA, representó una mejora muy leve respecto del mejor de los clasificadores para ambos casos.

De los clasificadores con mejor desempeño apuntados; RF, RPART y XGB permiten ordenar los predictores seleccionados asignando pesos que indican la contribución, dentro del modelo, de cada predictor para clasificar la variable respuesta. Por lo tanto es posible ordenar los predictores según su importancia. El Cuadro 7 muestra esto mismo, con los diez predictores más importantes en cada modelo. En lo que respecta a SONIA, resultaron de importancia: la imaginabilidad y concretud altas del controlador, que haya un término en la concordancia con “e” epentética; la animicidad del controlador y que esté presente un artículo definido. Las otras variables fueron dinámicas (autocorrelación [lag-2] y *wavelets* [“D5”]) y de los enlaces del grafo (“rwr”, “pa”, “cos_l”). En el caso de NATI resultaron de importancia para predecir la respuesta: la frecuencia del artículo “los”; variables dinámicas (autocorrelación [lag-1, lag-2], *wavelets* [“D4”], *kurtosis* y desvío típico) y variables de enlaces del grafo (“mf”, “cos_l”, “act_n”). Notar que no hubo variables de atributos de concordancias en este caso. En lo que concierne a JAKO, fueron importantes: la terminación en “as” del primer término de la concordancia, la similaridad (de las terminaciones) media y alta con el italiano, los modificadores que son adjetivos, la similaridad media con el italiano de la raíz léxica, que haya “e” epentética en un término de la concordancia, la frecuencia de los enlaces del grafo y la frecuencia de las palabras “lugares”, “competencias” y “particulares”. En lo que atañe a MIRKA solamente seis atributos fueron importantes: la frecuencia de enlace y los atributos de enlace del grafo (“mf”, “ka”, “act”, “rwr”, “act_n”). Nótese que los atributos derivados del grafo fueron seleccionados para todos los alumnos, lo cual aboga por la inclusión de datos de grafos en tareas de clasificación.

	IMPORTANCIA
SONIA [RF]	ima.conc="1" (6); rwr (3.71), l (2.35), es="1" (2.25), ACF_2 (2.24), D5 (2.04), cos_l (2), anim="1" (1.87), pa (1.80), mod="0" (1.73)
NATI [RPART]	mf (29.75), l (27.15), cos_l (20.28), act_n (20.07), “los” (19.15), D4 (18.73), SD (16.71), ACF_1 (16.17), ACF_2 (15.47), kurt (13.58)
JAKO [XGB]	esp1="as" (0.16); morf.f="1" (0.09); “lugares” (0.07), “competencias” (0.066), morf.f="2" (0.063), mod="3" (0.055), stem.f="1" (0.051), es="1" (0.048), “particulares” (0.044), freq_enlace (0.043)
MIRKA [RPART]	mf (33.81), freq_enlace (6.0433), ka (6.0434), act (4.41), rwr (3.92), act_n (2.94)

Cuadro 11.14: Los diez primeros predictores más importantes. Pesos entre paréntesis.

	clasif	tipo	bac	pr	rec	f1	auc	c
1	ensemble		0.800	0.417	0.769	0.541	0.841	0.104
2	svm	under	0.788	0.385	0.769	0.513	0.726	0.354
3	rf	smote	0.785	0.324	0.846	0.468	0.794	0.365
7	gbm	cw	0.772	0.306	0.846	0.449	0.718	0.385
5	xgb	cw	0.754	0.282	0.846	0.423	0.678	0.417
6	logreg	cw	0.687	0.286	0.615	0.390	0.652	0.521
4	rpart	cw	0.684	0.250	0.692	0.367	0.652	0.531

Cuadro 11.15: Clasificadores [SONIA]: support vector machine (svm), random forest (rf), rpart, Gradient Boosting Machine (gbm), eXtreme Gradient Boosting (xgb), logistic regression (logreg), ensemble; (2) tipos de pesos: undersampling (under), oversampling (over), smote, class weight (cw), base (sin pesos)

	clasif	tipo	bac	pr	rec	f1	auc	c
4	rpart	cw	0.671	0.395	0.625	0.484	0.654	0.667
2	svm	over	0.664	0.328	0.833	0.471	0.598	0.629
5	xgb	cw	0.624	0.302	0.792	0.437	0.553	0.762
6	logreg	base	0.621	0.284	0.958	0.438	0.596	0.724
7	gbm	cw	0.621	0.326	0.625	0.429	0.592	0.819
3	rf	cw	0.618	0.297	0.792	0.432	0.614	0.781
1	ensemble		0.590	0.267	0.958	0.418	0.706	0.400

Cuadro 11.16: Clasificadores [NATI]: support vector machine (svm), random forest (rf), rpart, Gradient Boosting Machine (gbm), eXtreme Gradient Boosting (xgb), logistic regresion (logreg), ensemble; (2) tipos de pesos: undersampling (under), oversampling (over), smote, class weight (cw), base (sin pesos)

	clasif	tipo	bac	pr	rec	f1	auc	c
2	svm	under	0.784	0.257	0.818	0.391	0.768	0.400
5	xgb	cw	0.760	0.225	0.818	0.353	0.685	0.443
3	rf	smote	0.736	0.292	0.636	0.400	0.701	0.496
6	logreg	smote	0.717	0.250	0.636	0.359	0.646	0.530
1	ensemble		0.710	0.316	0.545	0.400	0.786	0.148
4	rpart	under	0.702	0.151	1.000	0.262	0.725	0.539
7	gbm	over	0.674	0.155	0.818	0.261	0.629	0.600

Cuadro 11.17: Clasificadores [JAKO]: support vector machine (svm), random forest (rf), rpart, Gradient Boosting Machine (gbm), eXtreme Gradient Boosting (xgb), logistic regresion (logreg), ensemble; (2) tipos de pesos: undersampling (under), oversampling (over), smote, class weight (cw), base (sin pesos)

11.5. Mixtura de Bernoulli y regularización.

Otra forma de separar los datos en dos conjuntos de dinámica diferente consiste en hacer un *clustering* por mezcla de distribuciones de Bernoulli. Esta sección adopta dicha estrategia. Luego se aplica a los datos de cada conjunto un modelo mixto de regularización para seleccionar las doce primeras variables. A continuación se selecciona el mejor modelo mixto logístico compuesto por dichas variables, para cada *cluster*. En primer lugar se detalla brevemente qué se entiende por un modelo de mixturas de distribuciones y uno mixto por regularización.

11.5.1. Clustering por mixtura de distribuciones.

Esta sección introduce la técnica de *clustering* por mixtura de distribuciones y seguirá mayormente a Piccard (2007) y Gómez (2014). Sea $Y = (Y_1 = y_1, Y_2 = y_2, \dots, Y_j = y_j, \dots, Y_n = y_n)$ una muestra aleatoria de n variables donde Y_j es la variable aleatoria j -ésima e y_j su realización observada⁹, cuya función de densidad de probabilidad es $f(y_j)$. La probabilidad de que la observación j -ésima provenga de la i -ésima distribución se denomina “mixtura de g componentes” y viene indicada por la función de probabilidad:

$$f(y_j | \psi) = \sum_{i=1}^g \pi_i f_i(y_j | \theta_i)$$

donde: $\psi = (\pi_1, \dots, \pi_g, \theta_1, \dots, \theta_g)$ es el vector de parámetros del modelo; y $f_i(y_j | \theta_i)_{i=1, \dots, g}$ es la densidad de la i -ésima componente del modelo, π_i su peso y θ_i el vector de parámetros de dicha distribución. Las componentes pueden pertenecer a diferentes familias paramétricas. Los pesos son proporciones con lo cual: $0 \leq \pi_i \leq 1$ y $\sum_{i=1}^g \pi_i = 1$.

⁹En principio, las variables aleatorias Y_j son vectores de dimensión q . Sin embargo aquí se considerará el caso unidimensional.

	clasif	tipo	bac	pr	rec	f1	auc	c
1	ensemble		0.674	0.467	0.845	0.602	0.715	0.146
2	svm	over	0.647	0.447	0.810	0.576	0.674	0.194
4	rpart	smote	0.644	0.449	0.786	0.571	0.644	0.206
5	xgb	smote	0.602	0.416	0.738	0.532	0.536	0.279
6	logreg	base	0.579	0.385	0.881	0.536	0.530	0.259
3	rf	over	0.559	0.371	0.940	0.532	0.581	0.263
7	gbm	under	0.554	0.371	0.857	0.518	0.524	0.300

Cuadro 11.18: Clasificadores [MIRKA]: support vector machine (svm), random forest (rf), rpart, Gradient Boosting Machine (gbm), eXtreme Gradient Boosting (xgb), logistic regresion (logreg), ensemble; (2) tipos de pesos: undersampling (under), oversampling (over), smote, class weight (cw), base (sin pesos)

Cada dato recibirá una “etiqueta” que designe la pertenencia más probable a una de las g componentes de la mezcla. Pero dichas etiquetas se desconocen a priori, con lo cual, la estimación de su valor puede considerarse un problema análogo al de datos faltantes. Este problema de estimación se puede resolver aplicando el algoritmo de *expectation-maximization* [EM]. Sea $y = (y_1, \dots, y_n)$ una muestra aleatoria observada de tamaño n (realización de Y) y será denominado vector de datos *incompletos*. Sea también el vector de variables aleatorias *latentes* Z_j con realizaciones $z_j: Z = (Z_1 = z_1, \dots, Z_g = z_g)$; con lo cual $z = (z_1, \dots, z_g)$ se denominará vector de datos *no observados*. Juntándolos, $X = (Y, Z)$ es la muestra de variables aleatorias completas con realizaciones $x_1 = (y_1, z_1), \dots, x_n = (y_n, z_n)$. Aquí Z_j es una variable indicadora de dimensión g , cuyo elemento i -ésimo indica la pertenencia al componente k de la mezcla, señalado por un *uno* y *ceros* en las otras posiciones. O sea que cada dato completo es de la forma $x_j = (y_j, z_{j1}, \dots, z_{jg})$. La variable latente $Z = (Z_1, \dots, Z_g)$ sigue una distribución multinomial de parámetros $n=1$ y $\pi = (\pi_1, \dots, \pi_g)$:

$$P(Z_j = z_j) = \frac{1}{z_{j1}! z_{j2}! \dots z_{jg}!} \pi_1^{z_{j1}} \pi_2^{z_{j2}} \dots \pi_g^{z_{jg}} = \prod_{i=1}^g \pi_i^{z_{ji}}$$

donde: $\sum_{i=1}^g z_{ji} = 1$ y $\sum_{j=1}^n \sum_{i=1}^g z_{ji} = n$. La función de densidad conjunta de una variable aleatoria “completa” se escribirá, teniendo en cuenta la relación entre distribuciones condicionales y marginales, como: $f(X_j) = f(Y_j = y_j, Z = z_j) = f(Y_j = y_j | Z_j = z_j) P(Z_j)$. Ya que $Z_j = 1$ indica que la j -ésima observación pertenece al i -ésimo componente de la mezcla, se tiene que $f(Y_j = y_j | Z_{ji} = 1) = f_i(y_j | \theta_i)$. Entonces, desarrollando $f(X_j)$ se obtiene:

$$\begin{aligned} f(Y_j = y_j, Z = z_j) &= f(Y_j = y_j, Z_{j1} = z_{j1}, \dots, Z_{jg} = z_{jg}) = f(Y_j = y_j | Z_{j1} = z_{j1}, \dots, Z_{jg} = z_{jg}) P(Z_{j1} = z_{j1}, \dots, Z_{jg} = z_{jg}) \\ &= \left\{ \prod_{i=1}^g [f_i(y_j | \theta_i)^{z_{ji}}] \right\} \times \left\{ \prod_{i=1}^g \pi_i^{z_{ji}} \right\} = \prod_{i=1}^g [\pi_i f_i(y_j | \theta_i)]^{z_{ji}} \end{aligned}$$

Los parámetros de la mezcla se estiman por *máxima verosimilitud*. Con lo cual, la función de verosimilitud para los datos completos (suponiendo independencia de éstos) reazará: $L(\psi | y, z) = \prod_{j=1}^n \prod_{i=1}^g [\pi_i f_i(y_j | \theta_i)]^{z_{ij}}$; con: $\psi = (\pi_i, \theta_i)_{i=1, \dots, g}$. Tomando logaritmo se obtiene la log-verosimilitud:

$$\mathcal{L}(\psi | y, z) = \log L(\psi | y, z) = \sum_{j=1}^n \sum_{i=1}^g z_{ij} [\log \pi_i + \log f_i(y_j | \theta_i)]$$

El valor estimado del vector de parámetros será pues el máximo de esta última: $\hat{\psi} = \underset{\psi}{\text{máx}} \log L(\psi | y, z)$. Ahora bien, la i -ésima proporción de la mezcla puede interpretarse como la probabilidad *a priori* de que una observación de la muestra provenga de la g -ésima población. Con lo cual, una vez hallado $\hat{\psi} = (\hat{\pi}, \hat{\theta})$, se puede calcular las probabilidades posteriores $\hat{\tau}_{1j}, \hat{\tau}_{2j}, \dots, \hat{\tau}_{gj}$ de que la observación y_j pertenezca a cada una de las i componentes de la mezcla. Para ello se usa la regla de *Bayes*:

$$\hat{\tau}_{ij} = P(Z_{ij} = 1 | Y_j = y_j) = \frac{f_i(Y_j = y_j | Z_{ij} = 1) P(Z_{ij} = 1)}{\sum_{i=1}^g f_i(Y_j = y_j | Z_{ij} = 1) P(Z_{ij} = 1)} = \frac{\hat{\pi}_i f_i(y_j | \hat{\theta}_i)}{\sum_{i=1}^g \hat{\pi}_i f_i(y_j | \hat{\theta}_i)}; i = 1, \dots, g; j = 1, \dots, n$$

Por último, la asignación de la observación j -ésima al i -ésimo cluster se lleva a cabo eligiendo la mayor de las probabilidades a posteriori:

$$\hat{z}_{ij} = 1 \text{ si } i = \underset{i}{\text{máx}} (\hat{\tau}_{ij}); \text{ o bien } \hat{z}_{ij} = 0, \text{ si no.}$$

En el caso de mezclas de distribuciones *Bernoulli*, se tiene que: $f_i(y_j | \theta_i) = \theta_i^{y_j} (1 - \theta_i)^{1-y_j}$ con $\theta_i = p_i$. Con lo cual la distribución mixta, eligiendo $g = 2$ será: $f(y_j | \psi) = \pi_1 p_1^{y_j} (1 - p_1)^{1-y_j} + \pi_2 p_2^{y_j} (1 - p_2)^{1-y_j}$. La log-verosimilitud del modelo será:

$$\mathcal{L}(\psi | y, z) = \sum_{j=1}^n \sum_{i=1}^2 z_{ij} \left\{ \log \pi_i + \log \left[\pi_i p_i^{y_j} (1 - p_i)^{1-y_j} \right] \right\} = \sum_{j=1}^n \sum_{i=1}^2 z_{ij} [\log \pi_i + y_j \log p_i + (1 - y_j) \log (1 - p_i)]$$

El algoritmo *EM* maximiza dicha log-verosimilitud. Consta de dos pasos que se suceden en cada iteración: (i) Paso “E”: Partiendo del vector de parámetros $\psi^{(m)}$ en la m -ésima iteración, se calcula la esperanza condicional de la función de log-verosimilitud de los datos completos, dado el dato observado y utilizando los parámetros $\psi^{(m)}$: $Q(\psi | \psi^{(m)}) = E \left[\mathcal{L}(\psi | y, Z) | Y = y, \psi^{(m)} \right]$. (ii) Paso “M”: se maximiza $Q(\psi | \psi^{(m)})$ con respecto a ψ : $\psi^{(m+1)} = \underset{\psi}{\text{max}} Q(\psi | \psi^{(m)})$. De este modo comenzando en $m = 0$ se itera hasta la convergencia (según un criterio de parada): $\psi^{(0)} \rightarrow \psi^{(1)} \rightarrow \dots \rightarrow \psi^{(t)} = \hat{\psi}$. Se utilizará la implementación del paquete *flexmix* de R (Grün & Leisch, 2007).

11.5.2. Modelo mixto de regularización.

Se usará el paquete *glmLasso* de R (Groll & Tutz, 2014). Allí se implementa un modelo mixto generalizado con penalización L_1 que permite realizar una selección de variables mediante *shrinkage* de coeficientes. La regresión LASSO («Least Absolute Selection and Shrinkage Operator») fue propuesta por R. Tibshirani (1996; ver también: Hastie, Tibshirani & Friedman, 2017) y; al igual que la regresión *Ridge*, impone una penalización en el tamaño de los coeficientes para seleccionar variables. Dada una regresión lineal con x_{ij} predictoras e y_i valores de la variable respuesta ($i = 1, \dots, N$ y $j = 1, \dots, p$), *Lasso* resuelve el siguiente problema de optimización minimizando:

$$\hat{\beta}^{Lasso} = \left\{ \min_{\beta} \frac{1}{N} \sum_{i=1}^N w_i \mathcal{L}(y_i, \eta) + \lambda \sum_{j=1}^p |\beta_j| \right\}$$

donde $\beta = (\beta_0, \beta_1, \dots, \beta_p)$ es el vector de parámetros y $\mathcal{L}(y_i, \eta)$ el logaritmo negativo de la verosimilitud para la observación i -ésima; $\eta = \beta_0 - \sum_{j=1}^p x_{ij} \beta_j$ es el predictor lineal; y w_i son posibles pesos para las observaciones. El parámetro positivo $\lambda \geq 0$ controla el grado de penalización impuesto: cuanto más alto, más penalización y, entonces, los coeficientes caen hacia *cero*; cuanto más bajo, los coeficientes se acercan a una estimación por OLS [*Ordinary Least Squares*]. Es decir que $\hat{\beta}^{Lasso} = \hat{\beta}^{OLS}$ si $\lambda = 0$; y $\hat{\beta}^{Lasso} = 0$ si $\lambda = \infty$.

Considérese ahora el modelo lineal generalizado mixto [GLMM]. Sea y_{it} la observación t en el cluster i ($i = 1, \dots, n$; $t = 1, \dots, T_i$); y agrupando dichas observaciones en un cluster i -ésimo, se tiene el vector: $y_i^T = (y_{i1}, \dots, y_{iT_i})$. Sea el vector de covariables asociado de efectos fijos $x_{it}^T = (1, x_{it1}, \dots, x_{itp})$ y el vector de efectos aleatorios $Z_{it}^T = (z_{it1}, \dots, z_{itq})$. Además, $\beta^T = (\beta_0, \beta_1, \dots, \beta_p)$ es el vector de coeficientes de efectos fijos y $b^T = (b_1, \dots, b_q)$, el vector de coeficientes de los efectos aleatorios. Las observaciones y_{it} se suponen variables aleatorias independientes condicionalmente (a los efectos aleatorios) con esperanzas $\mu_{it} = E[y_{it} | b_i, x_{it}, z_{it}]$ y varianzas $\text{var}(y_{it} | b_i) = \phi v(\mu_{it})$, donde ϕ es el parámetro de escala y $v(\cdot)$ una función de varianza conocida. Entonces, el GLMM tiene la forma:

$$g(\mu_{it}) = x_{it}^T \beta + z_{it}^T b_i$$

donde g es la función de enlace (continua y monótona) y los factores aleatorios se distribuyen como $b_i \sim N_q(0, Q)$, o sea con una normal multivariada con matriz de varianzas y covarianzas Q de orden $q \times q$ y vector de medias nulo. Ahora bien, la función de densidad de las observaciones condicionales a los predictores fijos y aleatorios pertenece a la familia exponencial. Es decir:

$$f(y_{it} | x_{it}, b_i) = \exp \left\{ \frac{(y_{it} \theta_{it} - \kappa(\theta_{it}))}{\phi} + c(\theta_{it}, \phi) \right\}$$

donde: $\theta_{it} = \mu(\theta_{it})$ denota el parámetro natural; ϕ es el parámetro de escala; $\kappa(\theta_{it})$ es una función específica al tipo de familia exponencial y $c(\theta_{it}, \phi)$ es una constante de normalización. Una forma de maximizar la log-verosimilitud del modelo consiste en aplicar *penalized quasi-likelihood* (PQL). La matriz de varianzas y covarianzas depende de un vector de parámetros desconocido ρ , es decir, $Q(\rho)$. En el método apuntado, dicho vector ρ de parámetros de la estructura de covarianza se junta con el parámetro de escala en $\gamma^T = (\rho^T, \phi)$ y también se juntan los vectores de los coeficientes $\delta^T = (\beta^T, b^T)$. Con éstos, la log-verosimilitud del modelo puede aproximarse mediante:

$$\mathcal{L}^{app}(\gamma, \delta) \approx \sum_{i=1}^n \log(f(y_i | \delta, \gamma)) - \frac{1}{2} b^T Q(\rho)^{-1} b$$

Para el método de regularización consiste en incluir el término de penalización $\lambda \sum_{i=1}^p |\beta_i|$ en la expresión anterior $\mathcal{L}^{app}(\gamma, \delta)$. Con lo cual: $\mathcal{L}^{pen}(\gamma, \delta) = \mathcal{L}^{app}(\gamma, \delta) - \lambda \sum_{i=1}^p |\beta_i|$. Una vez hallado el estimador $\hat{\gamma}$ el problema de hallar los estimadores de los coeficientes $\hat{\delta}$ se reduce a maximizar la log-verosimilitud penalizada mediante un algoritmo basado en *full gradient*.

$$\hat{\delta} = \underset{\delta}{\operatorname{argmax}} \mathcal{L}^{pen}(\gamma, \delta) = \underset{\delta}{\operatorname{argmax}} [\mathcal{L}^{app}(\gamma, \delta) - \lambda \sum_{i=1}^p |\beta_i|]$$

11.5.3. Resultados.

En primera instancia se recategorizó la variable “ESP1”, con los siguientes niveles: 1 = “us”, “is”; 3 = “os”, 4 = “as”, 5 = “es”, 6 = “*es” (porque la segunda categoría incluía pocos casos, lo cual generaba problemas de estimación). Se aplicó un *clustering* de distribuciones Bernoulli, con datos agrupados por sesión. Es decir que los clusters agrupan las observaciones *dentro* de las sesiones. El siguiente Cuadro muestra los dos clusters obtenidos. Se observa que el primero agrupa las sesiones en torno a una media de error del 34% (rango [0,32;0,62]); por otro lado, el segundo las agrupa entorno a un error del 17% (rango [0;0,27]). Con lo cual se obtuvo un conjunto de error alto y otro de error medio-bajo.



Figura 11.8: Clusters obtenidos mediante mixturas de Bernoulli.

Luego se aplicó un modelo mixto de regularización a cada conjunto de datos. El parámetro de regularización λ se obtuvo mediante validación cruzada. Los factores aleatorios fueron las sesiones. El Cuadro que sigue muestra las primeras doce variables seleccionadas para cada cluster.

Cluster	λ	Selección
1	15	"esp1" "morf.f" "mod" "est5" "lda" "est2" "C" "anim" "est6" "esp2" "skew" "stem"
2	5	"es" "esp1" "morf.f" "est6" "fam.lex" "lda" "stem" "est4" "est1" "est7" "mod" "esp2"

Cuadro 11.19: Selección de variables mediante regularización mixta.

A continuación se define el modelo logístico mixto condicional para la concordancia i en el grupo j (el grupo está definido como la sesión k anidada en el alumno g). Para el cluster 1, los índices son: $k = 1, \dots, 4, 6, \dots, 12$; $g = 1, 2, 4$; $i = 1, \dots, 631$. Para el cluster 2 son: $k = 1, \dots, 14$; $g = 1, \dots, 4$; $i = 1, \dots, 1191$.

Distribuciones: $y_{ij} | v_{0ij}, w_{ij} \sim \text{Binomial}(N_{ij}, \pi_{ij})$, $N = 1$; $v_{0ij} \sim N_q(\mathbf{0}, \mathbf{G}_0)$; $w_i \sim N_{n_i}(\mathbf{0}, \mathbf{R}_i)$, donde: $w_i = w_{j1}, \dots, w_{jn_i}$ ($n_i = 4, \dots, 81$); y con $q = 1, \dots, 15$ para el cluster 1 y para el 2 es $q = 1, \dots, 37$.

Matrices de varianza: $G_0 = \sigma_{v_0}^2 I$; y $R_i = \sigma_{\omega}^2 I$.

Función de enlace: $\text{logit} = g(E[y_{ij} | v_{0ij}, w_{ij}]) = \log \left[\frac{\pi_{ij}}{1 - \pi_{ij}} \right]$

Función inversa: $\pi_{ij} = g^{-1}(X\beta + Z_v v + Z_\omega w)$

Predictor lineal: $\eta_{ij} = \varphi_{ij} + v_{0ij} + w_{ij}$. Donde: φ_{ij} representa los efectos fijos; v_{0ij} son los efectos aleatorios «entre»; w_{ij} son los efectos aleatorios «dentro».

Modelo teórico del Cluster 1:

$$g(E[y_{ij} | v_{0ij}, w_{ij}]) = \log \left[\frac{\pi_{ij}}{1 - \pi_{ij}} \right] = (\beta_0 + v_{0i}) + w_{ij} + \beta_1 \text{ANIM} + \beta_2 \text{C} + \beta_3 \text{ESP1}_3 + \beta_4 \text{ESP1}_4 + \beta_5 \text{ESP1}_5 + \beta_6 \text{ESP1}_6 + \beta_7 \text{EST5} + \beta_8 \text{MOD}_1 + \beta_9 \text{MOD}_2 + \beta_{10} \text{MOD}_3 + \beta_{11} \text{MORF.f}_1 + \beta_{12} \text{MORF.f}_2 + \beta_{13} \text{SKEW} + \beta_{13} \text{EST5} + \beta_{14} \text{ESP2}_2 + \beta_{15} \text{ESP2}_3 + \beta_{16} \text{ESP2}_4 + \beta_{17} \text{EST2} + \beta_{18} \text{LDA} + \beta_{19} \text{STEM}$$

Modelo teórico del Cluster 2:

$$g(E[y_{ij}|v_{0ij}, w_{ij}]) = \log \left[\frac{\pi_{ij}}{1-\pi_{ij}} \right] = (\beta_0 + v_{0i}) + w_{ij} + \beta_1 ESP1_3 + \beta_2 ESP1_4 + \beta_3 ESP1_5 \\ + \beta_4 ESP1_6 + \beta_5 ESP2_2 + \beta_6 ESP2_3 + \beta_7 ESP2_4 + \beta_8 EST1 + \beta_9 FAM.LEX + \beta_{10} MOD_1 \\ + \beta_{11} MOD_2 + \beta_{12} MOD_3 + \beta_{13} STEM + \beta_{14} EST6 + \beta_{15} ES_1 + \beta_{16} ES_2 + \beta_{17} LDA + \beta_{18} EST7 \\ + \beta_{19} EST4 + \beta_{20} MORF.f_1 + \beta_{21} MORF.f_2$$

Los efectos aleatorios (condicionales) son: (i) v_{0i} : la desviación del grupo i de la ordenada al origen; (ii) w_{ij} : efectos aleatorios «dentro» de cada grupo, con $w^T = [w_{j1}, w_{j2}, \dots, w_{jni}]$. Los efectos fijos (marginales) son: (i) β_0 es la media basal marginal; (ii) β_j ($j = 1, 2, \dots$) son los efectos de las variables independientes. La matriz de varianza-covarianza de los efectos aleatorios «entre» como aquella de los efectos aleatorios «dentro» son la identidad.

Se practicó selección de modelos basado en medidas de información (Burnham & Anderson, 2010). Primero, se ajustaron modelos logísticos mixtos con las predictoras significativas seleccionadas. Fueron $2^{12} - 1 = 4095$ modelos, jerarquizados mediante la medida de información AIC (como: $\frac{n_{c,1}}{p} = \frac{631}{12} \approx 52 > 40$ y $\frac{n_{c,2}}{p} = \frac{1191}{12} \approx 99 > 40$, no se usó la versión AICc corregida por tamaño muestral). Luego se examinó la frecuencia de las predictoras en el conjunto completo de modelos, que da un panorama de la incerteza por la selección. A continuación se redujo la cantidad de modelos al subconjunto «de confianza» con la regla $\frac{W(i)}{W(1)} > \frac{1}{8}$. Sobre dicho subconjunto se llevó a cabo un promedio de coeficientes con la varianza calculada con «full average».

Las gráficas muestran, para ambos clusters, los *Odds Ratio* del mejor modelo hallado. Por otro lado, los Cuadros muestran los coeficientes promediados. En lo que respecta al primer cluster (error alto), el Cuadro de coeficientes promediados indica que resultan significativas las variables: $ANIM=1$, C , $ESP1="os"$, $ESP1="as"$, $ESP1="es"$, $ESP1="*es"$, $EST5=1$, $MOD=2$, $MOD=3$, $SKEW$. En lo que atañe al segundo cluster, las variables significativas fueron: $ESP1="os"$, $FAM.LEX=1$, $ESP2="as"$. Nótese que los *Odds Ratio* significativos del primer cluster coinciden con los coeficientes promediados. En cambio, el mejor modelo encontrado para el cluster 2 señala que también son significativas las variables: $ESP2="as"$, $EST1$, $MOD=2$, $STEM=1$. O sea que este modelo en particular no coincide con el promedio de los coeficientes de los modelos del subgrupo “de confianza”.

En resumen, la chance de cometer un error de concordancia ($Y = 1$) respecto de no cometerlo ($Y = 0$) *disminuye* cuando se pasa de la categoría de referencia a la k -ésima categoría de la predictoras para: (i) las palabras de primer término de la concordancia terminadas en “os” (respecto de la referencia “-is”, “-us”) para ambos clusters; (ii) las palabras que funcionan como controlador de familiaridad y frecuencia léxica alta (cluster 2); (iii) las palabras del primer término de la concordancia terminadas en “as”, “es”, “*es” (cluster 1); (iii) la estrategia 5 (cluster 1). Por otro lado, la chance de cometer un error de concordancia ($Y = 1$) respecto de no cometerlo ($Y = 0$) *aumenta* cuando se pasa de la categoría de referencia a la k -ésima categoría de la predictoras para: (i) los controladores animados; (ii) los modificadores que son adjetivos o determinantes (ambos en el cluster 1). Además, la chance de error *disminuye* a medida que aumenta la complejidad de errores acumulados precedentes medido por “C” (cluster 1). La chance *aumenta* a medida que aumenta la asimetría de la distribución de errores acumulados precedentes [SKEW] (cluster 1). Resulta de interés notar que las terminaciones de las palabras del primer término de la concordancia (ESP1) constituyeron siempre factor de protección; y los tipos de modificador (MOD) siempre factor de riesgo. La animicidad del controlador fue factor de riesgo para las concordancias de error alto pero dicha variable no resultó seleccionada en el caso de error medio-bajo.

Ya que, para el caso de las variables categóricas, los p-valores no están ajustados por tests múltiples, los gráficos 11 y 12 muestran las comparaciones entre los niveles ajustando los p-valores por el método de Tukey. En lo que atañe al cluster 1, resultan significativas las palabras del primer término terminadas en “-os” y “-*es” (-e- epentética) respecto de la referencia “-us”, “-is” [chance de error disminuye] y también hay diferencia significativa entre las palabras terminadas en “-as” con respecto a las terminadas en “-os” [chance de error aumenta]. En cuanto a la variable MOD del primer cluster solo resultó significativa la diferencia entre las concordancias con modificador adjetivo y el nivel de referencia (artículo definido) [chance de error aumenta]. En lo que respecta al segundo cluster se observaron las siguientes diferencias significativas: (i)

ESP1: las palabras terminadas en “-os” respecto de la referencia [chance de error disminuye]; y aquellas terminadas en “-as” y “-es” respecto de las terminadas en “-os” [chance de error aumenta]; (ii) ESP2: las palabras del segundo término de la concordancia terminadas en “-as” respecto de aquellas terminadas en “-os” (referencia) [chance de error disminuye] y aquellas terminadas en “-es” respecto de aquellas terminadas en “-as” [chance de error aumenta]; (iii) MOD: ninguna diferencia resultó significativa. En el apéndice se detallan los tests.

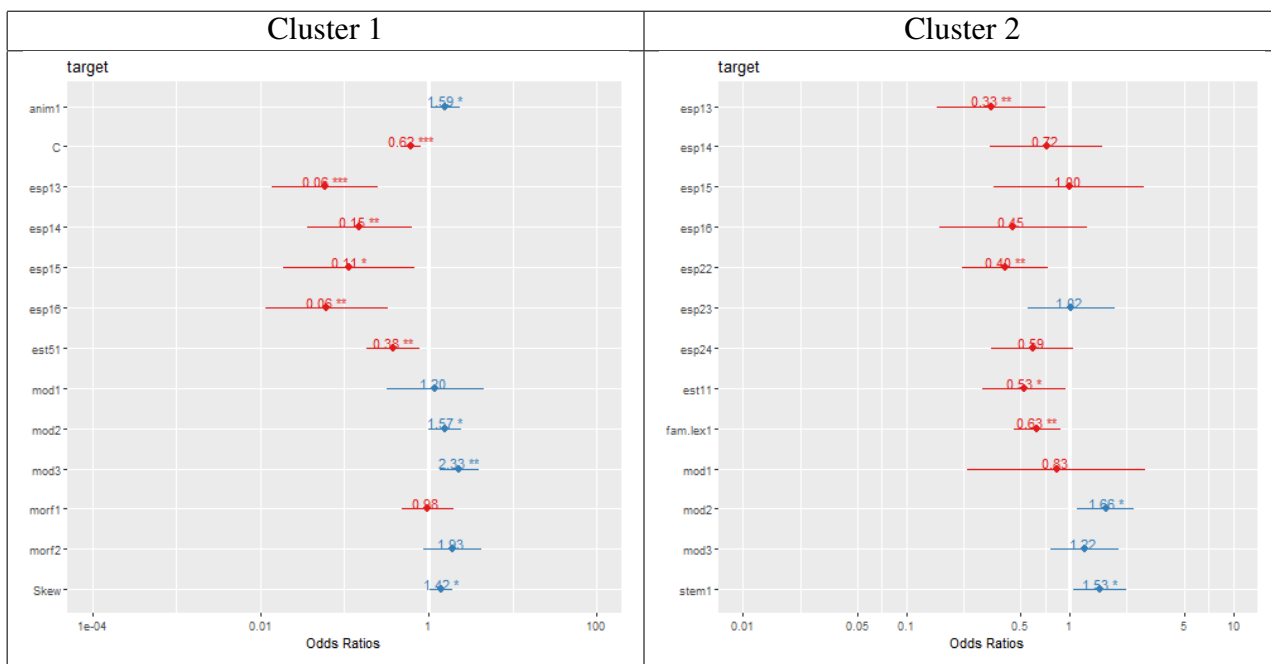


Figura 11.9: Odds Ratio de los mejores modelos para los *clusters* 1 y 2.

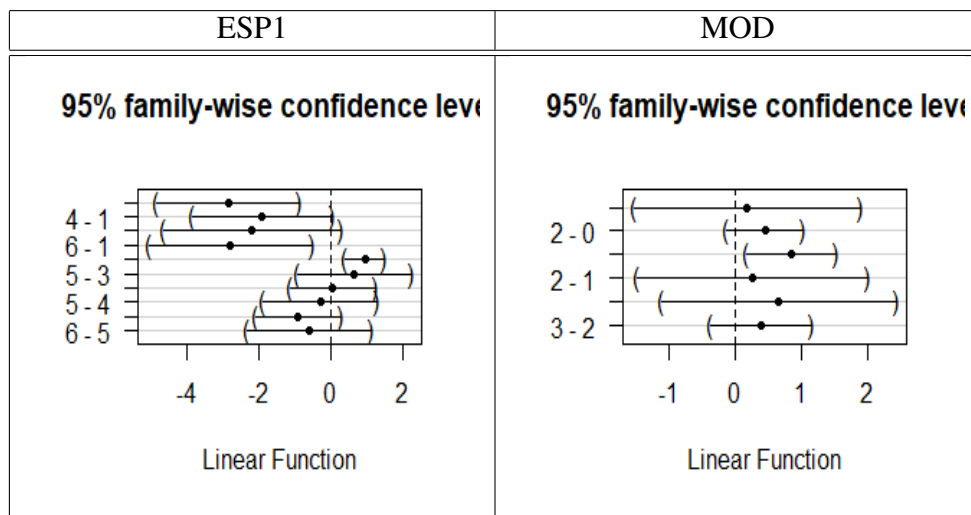


Figura 11.10: Cluster 1: Comparaciones múltiples (p-valores ajustados por *Tukey*)

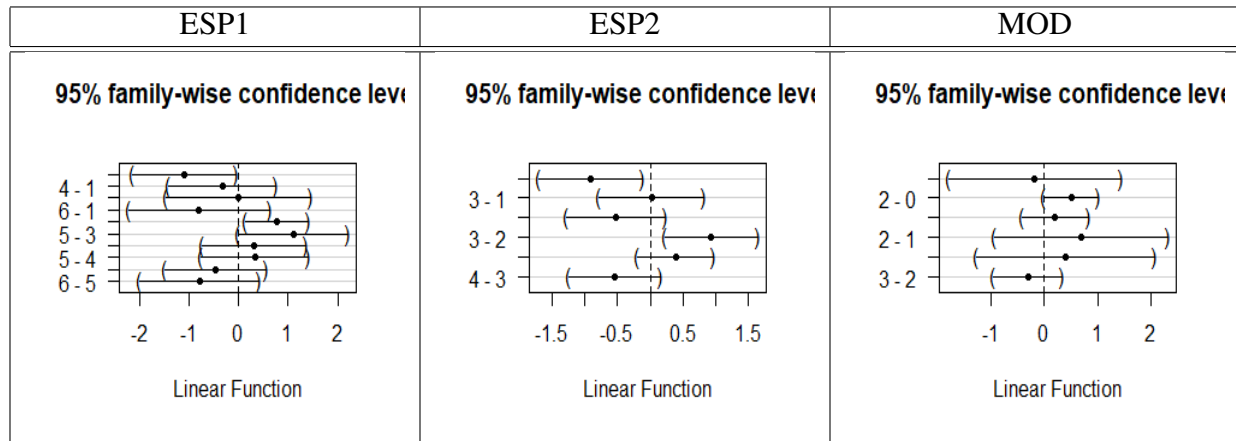


Figura 11.11: Cluster 2: Comparaciones múltiples (p-valores ajustados por *Tukey*)

	Estimate	Std. Error	Adjusted SE	z value	Pr(> z)
cond((Int))	1.0120	0.6891	0.6905	1.4657	0.1427
cond(anim1)	0.4388	0.2209	0.2213	1.9832	0.0473
cond(C)	-0.4866	0.1443	0.1446	3.3658	0.0008
cond(esp13)	-2.8435	0.7231	0.7245	3.9246	0.0001
cond(esp14)	-1.8624	0.7240	0.7254	2.5673	0.0102
cond(esp15)	-2.1784	0.9125	0.9143	2.3826	0.0172
cond(esp16)	-2.6827	0.8484	0.8500	3.1561	0.0016
cond(est51)	-0.9484	0.3816	0.3823	2.4807	0.0131
cond(mod1)	0.1902	0.6779	0.6792	0.2800	0.7795
cond(mod2)	0.4560	0.2290	0.2295	1.9871	0.0469
cond(mod3)	0.8497	0.2739	0.2744	3.0962	0.0020
cond(morf1)	0.0025	0.2973	0.2979	0.0085	0.9932
cond(morf2)	0.4115	0.4556	0.4561	0.9023	0.3669
cond(Skew)	0.3514	0.1671	0.1674	2.0993	0.0358
cond(est61)	-0.1008	0.2990	0.2993	0.3368	0.7362
cond(esp22)	0.0352	0.1743	0.1746	0.2014	0.8404
cond(esp23)	-0.0663	0.2140	0.2144	0.3093	0.7571
cond(esp24)	0.1974	0.3402	0.3404	0.5800	0.5619
cond(est21)	-0.0912	0.2354	0.2357	0.3869	0.6988
cond(lda1)	0.0376	0.2093	0.2096	0.1794	0.8576
cond(stem1)	-0.0141	0.1062	0.1063	0.1330	0.8942

Cuadro 11.20: Cluster 1: Coeficientes promediados.

	Estimate	Std. Error	Adjusted SE	z value	Pr(> z)
cond((Int))	-0.3006	0.5530	0.5533	0.5433	0.5869
cond(esp13)	-1.2090	0.4321	0.4324	2.7960	0.0052
cond(esp14)	-0.4047	0.4232	0.4236	0.9555	0.3393
cond(esp15)	-0.0989	0.5344	0.5350	0.1849	0.8533
cond(esp16)	-1.2133	1.0524	1.0531	1.1521	0.2493
cond(esp22)	-0.7216	0.4030	0.4032	1.7900	0.0735
cond(esp23)	0.0180	0.3308	0.3311	0.0543	0.9567
cond(esp24)	-0.8257	0.9948	0.9953	0.8296	0.4068
cond(est11)	-0.4401	0.3655	0.3657	1.2033	0.2289
cond(fam.lex1)	-0.4528	0.1662	0.1663	2.7223	0.0065
cond(mod1)	-0.1097	0.4853	0.4858	0.2258	0.8214
cond(mod2)	0.2717	0.2885	0.2886	0.9413	0.3466
cond(mod3)	0.1132	0.2122	0.2124	0.5331	0.5940
cond(stem1)	0.3272	0.2248	0.2249	1.4547	0.1458
cond(est61)	0.3250	0.4425	0.4427	0.7342	0.4628
cond(es1)	0.4645	0.9514	0.9519	0.4880	0.6255
cond(es2)	0.4471	1.6911	1.6925	0.2641	0.7917
cond(lda1)	0.0920	0.2547	0.2549	0.3611	0.7180
cond(est71)	0.0842	0.2439	0.2441	0.3449	0.7301
cond(est41)	-0.0054	0.1524	0.1525	0.0357	0.9715
cond(morf1)	-0.0231	0.1275	0.1276	0.1814	0.8560
cond(morf2)	-0.0093	0.1093	0.1093	0.0852	0.9321

Cuadro 11.21: Cluster 2: Coeficientes promediados.

11.6. Conclusiones.

A continuación se comentan los hallazgos de los diferentes métodos aplicados.

Análisis estadístico longitudinal de grafos. En el caso de SONIA: Las palabras con controlador de familiaridad / frecuencia léxica alta tienden tanto a generar como a recibir errores; las palabras terminadas en “es” (ej.: cantantes, amables), las que conllevan -e- epentética y las que tienen alta similaridad de raíz / desinencia con el italiano tienden a tener valores similares con el otro término de la concordancia errónea. En cuanto a NATI, tienden a generar errores las palabras que funcionan como adjetivos y aquellas con -e- epentética. Por otro lado, las palabras terminadas en “es” tienden a tener valores similares con el otro término de la concordancia errónea. Para JAKO tienden a generar error las palabras de controlador con alta imaginabilidad y concretud. Las palabras que tienen alta similaridad de desinencia con el italiano tienden a tener valores similares con el otro término de la concordancia errónea. En el caso de MIRKA dicho efecto de similaridad se observó con las palabras terminadas en “es” y aquellas con -e- epentética. También, para MIRKA, se halló un efecto de influencia en la acumulación de error, respecto de controladores con valores imaginabilidad / concretud altos.

Clasificadores para predecir el estatus de error. En general no se logró superar el 80% de accuracy. SVM resultó el clasificador de mejor desempeño. SONIA y JAKO alcanzaron mejores niveles de *accuracy* que NATI y MIRKA. Utilizar el “ensemble” redundó en una mejora muy leve. Los atributos derivados del grafo fueron seleccionados para todos los alumnos. En lo que respecta a la selección de los atributos de las concordancias y a la frecuencia del enlace en el grafo y de las palabras, se consideraron de importancia los siguientes: (i) SONIA: la imaginabilidad y concretud altas del controlador, que haya un término en la concordancia con -e- epentética; la animicidad del controlador y que esté presente un artículo definido; (ii) NATI: la frecuencia del artículo “los”; (iii) JAKO: la terminación en “as” del primer término de la concordancia, la similaridad (de las terminaciones) media y alta con el italiano, los modificadores que

son adjetivos, la similaridad media con el italiano de la raíz léxica, que haya -e- epentética en un término de la concordancia, la frecuencia de los enlaces del grafo y la frecuencia de las palabras “lugares”, “competencias” y “particulares”; (iv) MIRKA: la frecuencia de enlace. En el caso de SONIA y NATI también resultaron importantes para predecir los atributos dinámicos. Parece interesante señalar que ni NATI ni MIRKA tuvieron atributos de concordancia seleccionados como influyentes y con ellos los clasificadores tuvieron peor desempeño.

Mixtura de Bernoulli y regularización. Se aplicó un *clustering* por mezcla de distribuciones para hallar dos grupos de probabilidad de error diferentes. El primero agrupaba las sesiones en torno a una media de error del 34% (“error alto”) y el segundo, en entorno a un error del 17% (“error medio-bajo”). Luego de aplicar un GLMM con penalización L_1 , se eligieron las primeras doce variables y se procedió a la selección de un conjunto de “confianza” de modelos GLMM para cada cluster. Según los coeficientes promediados en dicho conjunto, a excepción de “Skew” y “C” el resto de las variables seleccionadas fueron atributos de las concordancias. En lo que respecta al primer cluster (error alto) resultaron significativas las variables: $ANIM=1$, C , $EST5=1$, $MOD=3$, $SKEW$. También fue significativa la diferencia entre: $ESP1=$ “os”, “as”, “es”, “*es” y la referencia ($ESP1=$ “us”, “is”), En lo que atañe al segundo *cluster*, las variables significativas fueron: $ESP1=$ “os”, $FAM.LEX=1$.

Por último, el Cuadro que sigue compara la selección de los atributos de las concordancias según el alumno y los métodos. Se desprende lo siguiente:

1. Inciden particularmente en el error de concordancia las palabras terminadas en “es”, en especial si pertenecen al primer término de la concordancia, que generalmente coincide con el «objetivo».
2. También aquellas concordancias con -e- epentética en ambos términos (efecto de homofilia en la variable ES).
3. Que la concordancia incluya un adjetivo y un determinante favorece el error (cluster de error “alto”).
4. En especial influyen en el error las palabras que tienden a no ser controladores (ESP1).
5. Los atributos del controlador (animicidad; imaginabilidad / concretud altas; familiaridad y frecuencia léxica altas) están relacionados con el error.

	SONIA	NATI	JAKO	MIRKA
GRAFOS	FAM.LEX;STEM ES; ESP=“es”	MOD=“4”,ES ESP=“es”	IMA.CONC MORF	IMA.CONC ES; ESP=“es”
CLASIFICADORES	IMA.CONC; ES=“1”; ANIM; MOD=“0”	-	ESP1=“as”; MORF.F=“1”, MORF.F=“2”, MOD=“3” ES=“1”; STEM.F=“1”	-
MIXTURA Y GLMM	CLUSTER 1	ANIM, EST5,	ESP1=“os”, “as”, “es”, “*es”,	MOD=“2”, “3”
	CLUSTER 2	FAM.LEX,	ESP1=“os”	

Cuadro 11.22: Atributos de concordancias seleccionados según los alumnos y los métodos empleados.

11.7. Apéndice de capítulo.

11.7.1. Comparaciones múltiples.

	lhs	rhs	estimate	std.error	statistic	p.value
1	3 - 1	0.00	-2.84	0.75	-3.81	0.00
2	4 - 1	0.00	-1.89	0.73	-2.58	0.06
3	5 - 1	0.00	-2.18	0.93	-2.36	0.11
4	6 - 1	0.00	-2.79	0.85	-3.27	0.01
5	4 - 3	0.00	0.95	0.20	4.75	0.00
6	5 - 3	0.00	0.66	0.59	1.11	0.77
7	6 - 3	0.00	0.06	0.44	0.13	1.00
8	5 - 4	0.00	-0.29	0.59	-0.49	0.99
9	6 - 4	0.00	-0.89	0.45	-2.00	0.23
10	6 - 5	0.00	-0.60	0.65	-0.93	0.87

Cuadro 11.23: Cluster 1. Comparaciones múltiples: ESP1

	lhs	rhs	estimate	std.error	statistic	p.value
1	1 - 0	0.00	0.19	0.68	0.27	0.99
2	2 - 0	0.00	0.45	0.23	1.98	0.18
3	3 - 0	0.00	0.84	0.27	3.13	0.01
4	2 - 1	0.00	0.27	0.69	0.39	0.98
5	3 - 1	0.00	0.66	0.70	0.94	0.77
6	3 - 2	0.00	0.39	0.30	1.32	0.53

Cuadro 11.24: Cluster 1. Comparaciones múltiples: MOD

	lhs	rhs	estimate	std.error	statistic	p.value
1	3 - 1	0.00	-1.10	0.39	-2.85	0.03
2	4 - 1	0.00	-0.33	0.40	-0.82	0.92
3	5 - 1	0.00	-0.00	0.54	-0.01	1.00
4	6 - 1	0.00	-0.80	0.53	-1.50	0.54
5	4 - 3	0.00	0.77	0.23	3.29	0.01
6	5 - 3	0.00	1.10	0.41	2.70	0.05
7	6 - 3	0.00	0.31	0.39	0.79	0.93
8	5 - 4	0.00	0.33	0.40	0.82	0.92
9	6 - 4	0.00	-0.46	0.37	-1.24	0.71
10	6 - 5	0.00	-0.79	0.45	-1.78	0.37

Cuadro 11.25: Cluster 2. Comparaciones múltiples: ESP1

	lhs	rhs	estimate	std.error	statistic	p.value
1	2 - 1	0.00	-0.91	0.31	-2.94	0.02
2	3 - 1	0.00	0.02	0.31	0.06	1.00
3	4 - 1	0.00	-0.52	0.30	-1.77	0.28
4	3 - 2	0.00	0.93	0.28	3.37	0.00
5	4 - 2	0.00	0.39	0.23	1.73	0.30
6	4 - 3	0.00	-0.54	0.27	-2.02	0.18

Cuadro 11.26: Cluster 2. Comparaciones múltiples: ESP2

	lhs	rhs	estimate	std.error	statistic	p.value
1	1 - 0	0.00	-0.19	0.64	-0.29	0.99
2	2 - 0	0.00	0.50	0.20	2.49	0.05
3	3 - 0	0.00	0.20	0.25	0.82	0.83
4	2 - 1	0.00	0.69	0.64	1.07	0.69
5	3 - 1	0.00	0.39	0.66	0.58	0.93
6	3 - 2	0.00	-0.30	0.25	-1.19	0.61

Cuadro 11.27: Cluster 2. Comparaciones múltiples: MOD

Capítulo 12

Comentarios didácticos y conclusiones.

En este último capítulo se introduce la noción de *feedback* y se la aplica a los errores de concordancia desde la perspectiva de los SDC. Se discuten los hallazgos de la presente tesis a la luz de los datos de la literatura. Se avanzan comentarios didácticos generales respecto de la enseñanza de la concordancia en español L2.

12.1. Feedback.

12.1.1. Interacción y Feedback.

La hipótesis del “notar” [*noticing hypothesis* (Schmidt, 2010)] afirma que aspectos de input no pueden adquirirse a no ser que se los note; es decir registrándolos pero “dándose cuenta”, poniéndolos en el foco de la atención. Sin embargo, dicho “darse cuenta” se refiere a la apercepción de instancias de lenguaje, no a las reglas abstractas (consciencia meta-lingüística). ¿Por qué es importante la instrucción de L2 orientada a la atención de instancias? Pues porque los alumnos sufren de *developmental sharpening* (Doughty, 2003). Esto significa que, aunque los niños inicialmente ponen atención en aspectos salientes del input, la experiencia con dicho input favorece una estrategia dominante de procesamiento. Una vez establecida la estrategia, rasgos del input nuevos dejan de ser detectados, al quedar fuera del rango del procesamiento ya afinado por la L1. Por ejemplo, si en una tarea de correspondencia con fotografías se le dice a un adulto inglés aprendiente de nivel inicial de español, la oración *lo sigue la madre* y se le muestra una figura de un hombre y otra de una mujer, posiblemente el hablante tenderá a elegir la del hombre porque en inglés el orden de palabras es una estrategia que se usa para identificar el sujeto. Esto no quiere decir que los hablantes sean completamente insensibles a las pistas [*cues*] del input, pero se los debe ayudar para “notar” dichas pistas cuando las estrategias afinadas de comprensión / producción que traen con el bagaje de su L1 resultan inadecuadas. Por lo tanto, el proceso de “darse cuenta” puede operar tanto en L1 como en L2, la cuestión radica en “qué cosa” logra notar el aprendiente. La situación es que éste necesita notar ciertos aspectos del input pero no lo logra por diferentes razones: (i) los aspectos a notar no interrumpen la comunicación (si no se dice determinada estructura uno se comunica igualmente); (ii) no son conceptualmente similares con su L1; (iii) no poseen saliencia perceptiva (porque se las percibe con los mecanismos afinados por la L1). Para remontar estos problemas entra en acción la intervención pedagógica. Y aquí es dónde viene a ayudar la instrucción “basada en la forma” [*focus on form instruction* (Long & Robinson, 1998)]. Esta implica atraer la atención de los aprendientes a la forma lingüística mientras se encuentran involucrados en actividades comunicativas. Entonces, aunque una clase orientada al significado (a la comunicación) sea muy deseable, el profesor debería dirigir la atención del aprendiente a elementos formales en tanto que aparecen durante la tarea comunicativa, dado que éstos constituyen un problema para la producción y/o comprensión. El *feedback* constituye claramente un método de “foco en la forma” ya que apunta a un “darse cuenta” del error. Póngase el acento en esto: es “darse cuenta” del error pero no “entenderlo”. Es decir se orienta al “cómo se dice algo” pero no al “por qué se dice algo” (reglas meta-lingüísticas).

Larsen Freeman (2003) define al *feedback* como “información evaluativa disponible para los aprendientes respecto de su desempeño comunicativo”. Propone algunos “consejos” para lograr un *feedback* efectivo.

En primera instancia el instructor debería concentrarse en los errores frecuentes. Estos constituyen atractores fuertes. Los errores que se cometen esporádicamente no serán influenciados por el *feedback*. En segundo lugar, habría que enfocarse en los errores en los cuales los alumnos saben lo que quieren decir pero no logran formular bien sus intenciones comunicativas. En una palabra, fomentar la “intersubjetividad”: ayudar a los alumnos infiriendo sus objetivos comunicativos para ayudarlos. Por último, sería aconsejable ofrecer *feedback* cuando los aprendientes necesiten información sobre sus propias producciones para poder eliminar hipótesis. Cuando los alumnos generalizan el uso, tienden a “sobre-generalizar” en estructuras compatibles con los datos disponibles. El *feedback* resultará mucho más efectivo si se interviene “online”, cuando la sobre-generalización está ocurriendo.

Swain (1985, 2005, 2008) postuló la “hipótesis de la producción” [*output hypothesis*]. Se define como un proceso de modificación de las producciones de los alumnos en el cual ellos crean versiones modificadas de instancias de lenguaje incorrecto producido previamente. Se plantean las siguientes funciones del *output*. En primer lugar, producir lenguaje ayuda a los alumnos a ser conscientes de “lagunas” en su propio conocimiento lingüístico. Esto los inducirá a tratar de resolver el problema creando nuevas formas basándose en el conocimiento de la interlengua. De este modo, ellos logran ver que existen diferencias entre lo que dicen y el lenguaje de los hablantes nativos (o el de sus compañeros). Por ejemplo, un aprendiente italiano que intenta producir la concordancia plural de *tren* [it. *treno*; pl. *tren-i*], podría decir *los tren*; dándose cuenta de que no sabe cómo producir el plural y luego reformular diciendo *los trenos*; forma incorrecta, pero que agrega el plural en *-s* al singular de la palabra italiana *treno* (consistente en la interlengua). La segunda función consiste en que la producción ofrece la oportunidad de poner a prueba hipótesis sobre el lenguaje. Luego de notar que hay lagunas respecto de un cierto aspecto de la lengua, los alumnos usan sus producciones para testear cuál podría ser la forma correcta. Por ejemplo, si no conocen el género para la concordancia plural de *mapa* [it. *mappa* (fem.)], podrían producir conjuntamente *las mapas* y *los mapas*, y esperar confirmación del interlocutor sobre la forma correcta. En tercer lugar, el *output* sirve para que los alumnos conversen entre sí sobre aspectos metalingüísticos de la lengua, resolviendo así problemas gramaticales (¿«*mapa*» es femenino o masculino?). Otra función es la de promover fluidez y automatismo. Una vez automatizada una rutina de uso, se requiere menos esfuerzo cognitivo para producirla. Por último, la producción ayuda a que el alumno se mueva de un uso de la lengua basado en el significado, cuyo principal objetivo es comunicarse y ser comprendido aunque se cometan errores en la forma; a otra etapa donde el uso se basa en la morfo-sintaxis; es decir, poniendo más atención en la forma.

En Gass & Mackey (2015) se comenta la función de la evidencia negativa en el proceso de aprendizaje. Se llama “evidencia negativa” a la información que los alumnos reciben sobre la no corrección de lo que producen. El *feedback* provee dicha evidencia negativa sobre el fracaso de los enunciados. Una vez que se da el *feedback* (evidencia negativa) el alumno debe escudriñar cuál fue el problema y cómo modificar el estado actual de su conocimiento lingüístico. Entonces, formula hipótesis sobre cuál debería ser la forma correcta. Para confirmar dichas hipótesis, una posibilidad es obtener más *input* y ver si la hipótesis se verifica en las instancias percibidas. Otra sería, como se dijo antes, producir lenguaje a partir de las hipótesis (re)formuladas y esperar confirmación del interlocutor.

En lo que atañe a los tipos de *feedback*, Lyster & Saito (2010) proponen los siguientes grupos [ver también: Gass & Mackey, 2015]:

Recasts. Consisten en la reformulación de todo o parte del enunciado proferido por el alumno sin incluir al error. Es decir que se reformula el enunciado corrigiendo el error, pero sin cambiar el significado global de la oración. Se le ofrece al alumno la forma esperada [*target form*] de lo que dijo. Se consideran más implícitos en el sentido de que no se incluyen frases como “debes decir ...”, aunque podrían ser más explícitos si se focalizan en una determinada palabra o involucran modificaciones gramaticales o léxicas. Pueden incluir evidencia positiva ofreciendo un modelo de estructura correcta; o bien negativa, ya que la forma correcta se yuxtapone con la incorrecta dicha por el alumno. Justamente dicha yuxtaposición realza la evidencia positiva.

(Recast)

Alumno: Cuando soy frío, me pongo un pulóver.

Profesor: Yo cuando tengo frío, tomo una sopa.

Correcciones explícitas. El instructor provee la forma correcta e indica que lo que el alumno acaba de decir fue erróneo.

(Corrección explícita)

Alumno: Este regalo es por mi madre.

Profesor: Ah para! querés decir para tu madre.

Prompts. A diferencia de los casos anteriores, las formas correctas *no* se proveen al alumno (*feedback* es implícito). Lyster & Ranta (1997) los consideran un tipo de “negociación por la forma”. Se le ofrece al alumno *feedback* correctivo, que puede llevar a la auto-reparación por parte del alumno; es decir que se involucra la precisión de la forma, no solo el hecho de la comprensibilidad (porque el error en cuestión quizás no haga fracasar la comunicación). Son los siguientes:

- **Pedidos de aclaración:** Se requiere una reformulación de lo dicho porque el enunciado no fue comprendido o está mal formado. Se usan frases del tipo *¿Perdón? ¿qué querés decir con X?*

(Pedido de aclaración)

Alumno: Quiero un jugo de melanzanas. [it. *mela*, esp. *manzana*; it. *melanzana*, esp. *berenjena*]

Profesor: ¿Qué querés decir con “melanzana”?

- **Pistas metalingüísticas:** proporcionar información (“*es masculino*”), comentarios (*hay un error, ¿podés encontrar el error?*) o preguntas (*¿es femenino? ¿seguro?*), sobre la buena formación de los enunciados pero sin mencionar la forma correcta.

(Pistas metalingüísticas)

Alumno: hay mucha casas en mi barrio.

Profesor: ¿mucha casas? ¿seguro se dice así?

- **Elicitaciones:** diferentes estrategias para elicitación de la forma correcta por parte del alumno. Por ejemplo, mediante una estrategia de “continuación”: *Es un ...*; o mediante preguntas: *¿Cómo se dice eso?*

(Elicitaciones)

Alumno: Es una calle larga [it. *largo*, esp. *ancho*]

Profesor: Es una calle an...

- **Repeticiones:** Repetir el error de manera aislada, subrayándolo con la entonación.

(Elicitaciones)

Alumno: los alemanos son ordenados.

Profesor: ¿alemanOS?

Lyster & Ranta (1997) siguieron seis cursos de inmersión en el área de Montreal y encontraron lo siguiente. Del total de los errores cometidos, el 62% recibieron *feedback*. De este 62%, alrededor del 55% produjeron algún tipo de respuesta por parte de los estudiantes [*uptake*]. Estas fueron: (i) “repetición”: repetir la forma correcta; (ii) “incorporación”: incluir la forma correcta en un enunciado más largo; (iii) “auto-reparación”: auto-corrección siguiendo a un *feedback* implícito; (iv) “reparación asistida”: un compañero ayuda en la corrección; (v) “nuevo error”: producir un enunciado que contenga un nuevo error, que debía ser, a su vez, reparado. Solamente el 27% de las respuestas condujeron a la reparación del error. Asimismo, los instructores prefirieron los *recasts* por sobre los otros tipos de *feedback*, en un 55%. No obstante ello, los *recasts* fueron los menos efectivos para inducir algún tipo de respuesta por parte de los alumnos. Por otra parte, las *elicitaciones* resultaron ser los más efectivos. Pero si se tomaban en cuenta únicamente las categorías de respuesta “auto-reparación” y “reparación asistida” (dejando de lado las demás), se observó que los *recasts* y las *correcciones explícitas* no generaban ningún tipo de reparación del error; mientras que

la efectividad para la inducción de respuestas de los otros tipos se produjo en el orden: *elicitaciones* (26 %) = *pistas metalingüísticas* (26 %) > *pedidos de aclaración* (20 %) > *repeticiones* (11 %). En suma, los *prompts* constituyeron los tipos de *feedback* más efectivos. Los autores interpretaron los resultados aduciendo que los *recasts* son ambiguos; es decir que los alumnos no los perciben como modificaciones de sus propias producciones (*feedback*). En tal caso, una respuesta del tipo “repetición” que sigue a un *recast* no tendría valor adquisicional, sería únicamente un “eco”.

Lyster & Saito (2010) llevaron a cabo un estudio de meta-análisis, que arrojó los siguientes resultados con respecto a la efectividad de los tipos de *feedback*. En la condición “intra-grupo” los *prompts* obtuvieron el tamaño de efecto más grande (según el *d* de Cohen) y fueron significativamente más efectivos que los *recasts*. Por otra parte, las correcciones explícitas no produjeron diferencias significativas respecto de los otros tipos. Los autores interpretaron los resultados como sigue. En primer lugar, adujeron que los aprendientes se beneficiaban de la evidencia positiva ofrecida por los *recasts* y por la oportunidad ofrecida por este tipo de *feedback* de inferir evidencia negativa. Sin embargo, los estudiantes sacaban aún más provecho de los *prompts*, los cuales proveen evidencia negativa y demandan más esfuerzo, ya que exigen que el alumno produzca output modificado. Respecto de las *correcciones explícitas*, explicaron que la falta de diferencia significativa se debía a que el accionar de dicho tipo se sobrepone con el de los otros. Al igual que los *recasts*, la *corrección explícita* conlleva evidencia positiva al proveer una forma correcta. Al mismo tiempo, conlleva evidencia negativa, al indicar al estudiante que su enunciado fue incorrecto.

Della Putta & Visigalli (2010) estudiaron aprendientes ingleses de italiano L2 para evaluar los efectos de tres métodos de instrucción para promover la adquisición del sintagma nominal. Para cada método tuvieron en cuenta cuatro variables. En primer lugar, el “modo de instrucción”: *Focus on form*, *Focus on meaning* (basado únicamente en tareas comunicativas); *Focus on forms* (apuntando a formas lingüísticas aisladas, presentadas de forma explícita y por un periodo restringido de tiempo). En segundo lugar, la “presentación de la gramática”: implícita (sin apuntar al “darse cuenta”) o explícita (inductiva o deductiva, “dándose cuenta”). En tercer lugar, los tipos de *feedback*. Por último, se incluyó el tiempo de práctica del sujeto sobre un total de 90 horas de duración del curso. El primer grupo (G1) recibió un método con *Focus on forms* promoviendo el “darse cuenta”, gramática explícita, diferentes tipos de *feedback* (pero ninguno de ellos orientado exclusivamente a los errores en el sintagma nominal) y práctica por un periodo corto de tiempo. Al segundo grupo (G2) se le administró una mezcla de instrucción basada en *Focus on meaning* (30 horas) más *Focus on form* (60 horas), con gramática explícita, usando *recasts* orientados al sintagma nominal (60 horas) y práctica durante toda la duración del curso. El tercer grupo (G3) recibió el método que resultó ser el más efectivo, con las siguientes características. En primera instancia se les dio instrucción basada en *Focus on form* durante 80 horas; resaltando el input visual de sintagmas nominales (es decir, apuntando a la forma correcta esperada). Los estudiantes también trabajaron con ejercicios orientados a la fonología de las vocales finales (que son morfemas de género y número en italiano) por 50 horas. La intención era siempre la de focalizar la atención hacia la forma. También recibieron 10 horas de instrucción basada en *Focus on meaning* con gramática implícita utilizando *chunks* inanalizados para presentar los datos de los sintagmas nominales. Por ejemplo, se les presentaba el *chunk* “determinante + sustantivo singular” [*un libro*] o bien “numerales / cuantificadores + sustantivos plurales” [*un chilo di banane* ‘un kilo de bananas’]. Con respecto a la concordancia con el adjetivo, también se la introdujo mediante *chunks*. Al principio se utilizaron concordancias con esquemas “armónicos” del tipo [e - e - e]: *le cas-e bianch-e* (esp. *las casas blancas*). Y luego se presentaban esquemas no armónicos del tipo [e - á - e]: *le universit-á famos-e* (esp. *las universidades famosas*). Los alumnos trabajaron con tareas comunicativas, *roll-plays* y ejercicios con colocaciones. A continuación siguieron 80 horas de instrucción basada en *Focus on form* con gramática explícita (explicando reglas morfo-sintácticas). Se usaron *recasts* en producción oral orientados a los errores en los sintagmas nominales (80 horas). Los alumnos hicieron práctica durante todo el curso. En lo que atañe a la concordancia de número se obtuvo que los errores de concordancia plural disminuyeron en la dirección G1 > G2 > G3; mientras que para la concordancia singular se verificó G2 > G1 > G3.

En resumen, la investigación reseñada aquí mostró que los *prompts* constituyen el tipo más efectivo de *feedback* porque demandan modificación del output; sin embargo, los *recasts* también resultan efectivos porque brindan la oportunidad de inferir evidencia negativa. Respecto de la concordancia dentro del sintagma nominal en italiano, el método que incluyó instrucción basada en *Focus on form*, input visual resaltado,

chunks y tareas comunicativas redundó en menor cantidad de errores.

12.1.2. Feedback para la concordancia.

¿Es necesario proveer *feedback* de inmediato? Recuérdese que en los SDC los sistemas son iterativos: lo que resulta de una iteración se usa como punto de partida para la siguiente. Lo que promueve la adquisición es la repetición de los mismos eventos de uso del lenguaje a lo largo de tareas comunicativas diferentes. La repetición es mecanismo de generación de variabilidad. Debe dejarse que el sistema itere y no apresurarse a intervenir con *feedback*. Obsérvese el siguiente ejemplo de NATI [sesión 7]. La estudiante no sabe como decir *jueces* [it. *giudici*]. El investigador interviene inmediatamente ofreciendo un *feedback* explícito. Seguidamente la alumna repite la concordancia como si fuera un “eco”, ya que en el siguiente enunciado comete un error de plural en el artículo.

*STU: e@s:ita cuando tu no &cu +//.

STU: si no tienes una ley es posible [//] eh@fp tu debe [] eh@fp preguntar a [/] a la &magis [//] a [/] a los eh@fp [//] preguntar a [/] a los **judice@n** [*4] qué hacer .

%err: debes; los jueces

*STU: eh@fp i@s:ita **judici** (.) los +//?

*INV: **los jueces** .

*STU: **los jueces** eh@fp en [/] en diferentes [*] eh@fp +//.

%err: en diferentes

*STU: **lo jueces** [*3] han considerado muchos eh@fp materiales científicos [*0] .

Más adelante, se habla de los jueces otra vez. En los cuatro intentos proferidos por la alumna el plural del artículo no constituye un problema (hay un solo error). En cambio para “jueces” se observa la trayectoria: *iusos* > *jiures* > *jusun* > *jugos* > *juce*. Al principio “los jueces” fue percibido como un *chunk* inanalizado. Luego NATI se concentra en lo más fácil: la concordancia del determinante; para después pasar a “construir” el plural de *jueces*. En el último intento se acerca bastante: *juce* / *jueces*. Al menos logra insertar una -e-epentética, aunque le falte la -s final. Se debe dejar que el sistema *itere*, corregir de inmediato no tiene ningún efecto de larga duración.

*STU: ha preguntado al estado [/] al estado [//] a [//] a los **iusos@n** [*0] que +//.

%err: los jueces

...

STU: cuando tu tiene [] un vacío **lo jiures@n** [*3] eh@fp deben eh@fp decidir [*] eh@fp secundo@n [*] los principios [*0] [/] los eh@fp [/] los principios de la constitución italiana e@s:ita de la leys .

%err: los jueces deben decidir según los principios de la constitución italiana y de la ley

*STU: y eh@fp de el derecho internacional .

*STU: **los jusun@n** [*4] han hecho la [//] han [/] han hecho eh@fp las cosa justos [*3] [*4] .

%err: los jueces han hecho lo correcto

*STU: eh@fp es siempre una decisión muy difícil porque +//.

STU: yo [/] yo creo [] no [/] no [/] no [/] no fue una decisión fácil por [*] **los jugos@n** [*0] .

%err: para los jueces

*STU: consideraron todos eh@fp los elementos [*0][*0] .

...

*STU: y después el [//] eh@fp la decisión de **los juce** [*3] fue la decisión final de el [//] del último [//] eh@fp la decisión final .

%err: los jueces

A través del lenguaje creativo, como el visto en el ejemplo, el alumno crea variación (errores). Al principio el aprendiente puede no estar consciente de encontrarse atrapado en un atractor del error. Si el profesor provee *feedback* indirecto en modo de *promts*, puede ayudar a elevar el nivel de atención respecto del atractor del error. El primer paso es darse cuenta de que se está en el “basin of attraction” de un atractor del error. Si al principio, el estudiante no se auto-corrige; es preciso dejarlo continuar con la creación de lenguaje y ayudarlo con más *feedback* indirecto. Eventualmente los ciclos de comisión del error van durar cada vez menos. Es decir, el atractor se debilitará. Véase otro ejemplo de SONIA [sesión 8]. En este caso el alumno se auto-corrige inmediatamente. De hecho, las instancias en las que se usa “profesores” a lo largo de las sesión son: *los profesores [línea 170] > los profesores [línea 287] > los profesores [línea 294] > los profesor [línea 386] > los profesores [línea 390]*. O sea que la instancia incorrecta es simplemente una pequeña *perturbación*: el sistema vuelve de inmediato al atractor estable. Introducir *feedback* aquí no serviría de nada.

*STU: creo que no está soltando el problema de **los profesor** [*2] en la escuela superiora [*].

*STU: están [//] está un problema [//] hay un problema de motivación. también en la universidad de **los profesores**.

Pero obsérvese en cambio la trayectoria de “alemanes”: *los alemanos* [sesión 3, línea 178] > *alemanes ingenuos* [sesión 3, línea 197] > *los alemanos* [sesión 4, línea 186] > *los estudiantes alemanes* [sesión 4, línea 193] > *los alemanos* [sesión 8, línea 40] > *los alemanios* [sesión 9, línea 67] > *los alemanos* [sesión 9, línea 72] > *los alemanos* [sesión 9, línea 127]. En este caso parece haber inclusive bimodalidad: se salta del atractor correcto al del error con cuasi-periodicidad. Aquí el atractor del error es fuerte y se requiere un *Feedback* implícito: ¿alemanOS? ¿seguro?; un catalán - dos catalanES; un alemán, dos ...

12.2. Discusión.

En esta sección se discuten los resultados de la tesis. Es preciso aclarar que se trata de estudios de casos y además son observacionales, sin control experimental de sesgos. Por lo tanto, los resultados de los modelos estadísticos constituyen material de hipótesis a ser verificadas con diseños experimentales aleatorizados.

Un hallazgo bastante establecido en la literatura sobre adquisición de la concordancia en L2 es que la concordancia de género y número del artículo es más fácil de adquirir que la del adjetivo. En el modelo logístico mixto se observó que la chance de cometer un error de concordancia ($Y = 1$) respecto de no cometerlo ($Y = 0$) aumenta cuando se pasa de concordancias con artículo definido a aquellas con determinantes pero no hubo efecto respecto de los adjetivos. Si se discrimina por tipo de error, dicho factor de riesgo en determinantes se volvió a observar con los errores mixtos, tanto en el modelo multinomial como con el de riesgos competitivos. Sin embargo se verificó un factor de protección (baja chance / riesgo de error) cuando se pasa de artículos definidos a adjetivos con errores de género en el modelo multinomial y con errores de plural en el modelo de riesgos competitivos. Asimismo la presencia de adjetivos en las concordancias hizo subir el riesgo de error en el modelo de riesgos competitivos para errores mixtos. En lo que atañe al modelo de agentes estocástico temporal de redes [*Actor-Based Stochastic Model*] y para el caso de NATI, los adjetivos incrementan la probabilidad de que se establezca un enlace de error en la red en la próxima etapa. El LASSO mixto generalizado seleccionó la variable MOD tanto para el grupo de error alto (media 34%) como para el de error medio/bajo (media 17%). Sin embargo en el segundo grupo MOD no resultó significativo. En el primero, de error medio más alto, se observó un aumento de la chance de error para los adjetivos respecto de los artículos definidos. En lo que respecta a las variables más influyentes en la predicción de estatus de error, los artículos definidos se seleccionaron en el caso de SONIA y los adjetivos en el caso de JAKO. Además, para NATI influyó el artículo “los” para predecir el error.

Respecto del género, en la literatura se halló más facilidad de adquisición de la concordancia para los nombres de morfología transparente (terminados en *-o* , *-a*) respecto a los de morfología no transparente (terminados en *consonante*, *-e* y opuestos tipo *la mano*). En cuanto al plural, el morfema en *-es* se adquiere después del morfema en *-s*. La variable ES fue codificada como “0” si no hay necesidad de insertar una *-e* epentética (controladores y objetivos de morfología transparente para el género, terminados en *-e* y géneros

opuestos); y como “1” y “2”, para aquellos terminados en consonante que requieran plural en *-es*, en un único término de la concordancia o en ambos. Ni el modelo mixto logístico ni en los de tiempo hasta un evento se halló significancia estadística para ES. El modelo mixto multinomial arrojó un efecto de *protección* para *el error de plural* cuando se pasa de concordancias sin *-e-* epentética en ningún término de la concordancia a aquellas con *-e-* epentética en ambos términos (ES = “2”). Por tanto, las palabras terminadas en consonante *no* aumentan la chance de cometer error de plural. No hubo asociación a errores de género. Respecto del modelo de agentes estocástico temporal de redes, tienden a establecer enlaces de error las palabras con valores similares en los atributos de *-e-* epentética como en *camion-es azul-es* (SONIA y MIRKA) y las terminadas en *-e*, como en *cantante-s increíble-s* (SONIA, NATI y MIRKA). El LASSO mixto generalizado seleccionó la variable ES para el grupo de error medio / bajo pero no resultó significativa. En lo que atañe al primer cluster (error alto), resultaron significativas las palabras del *primer término* (ESP1) terminadas en *-os* y *-*es* (*-e-* epentética) respecto de la referencia *-us*, *-is* [chance de error disminuye] y también hubo diferencia significativa entre las palabras terminadas en *-as* con respecto a las terminadas en *-os* [chance de error aumenta]. O sea que, en el grupo de error alto, los plurales masculinos y los de palabras terminadas en consonante generan *menos* error que las terminadas en *-is/-us*, que en su mayoría eran posesivos (confirmando la dificultad de éstos). Pero también los femeninos plurales son más complicados que los masculinos (en los atributos, que tienden a estar en el primer término ESP1). Esto coincide con la literatura: el femenino es más difícil que adquirir que el masculino. En lo que respecta al segundo cluster (error medio/bajo) se observaron las siguientes diferencias significativas: (i) ESP1: las palabras del primer término terminadas en *-os* respecto de la referencia en *-is/-us* [chance de error disminuye]; y aquellas terminadas en *-as* respecto de las terminadas en *-os* [chance de error aumenta] (estos resultados coinciden con el primer grupo); (ii) ESP2: las palabras del segundo término de la concordancia terminadas en *-as* respecto de aquellas terminadas en *-os* (referencia) [chance de error disminuye] y aquellas terminadas en *-es* respecto de aquellas terminadas en *-as* [chance de error aumenta]. Aquí las femeninas parecen ser más fáciles que las masculinas en el segundo término de la concordancia, que tiende a coincidir con el controlador. Además hay un efecto de morfología «no transparente» para el plural: las palabras del segundo término terminadas en *-e* son más difíciles que las terminadas en *-a*. En cuanto a las variables seleccionadas para la predicción de estatus de error, el atributo ES = “1” se eligió para SONIA y JAKO. En suma, las concordancias que llevan *-e-* epentética en algún término ayudaron a no cometer errores de plural. Sin embargo, se observaron efectos inhibidores para los términos individuales de la concordancia terminados en *-e* y en *consonante* cuando se analizaron los datos con redes complejas y LASSO.

En la literatura también se reporta para la concordancia de género un efecto facilitador de los controladores inanimados respecto de los animados. ANIM=“1” incrementó la chance de error de *-e-* epentética (modelo multinomial); y error de plural (modelo de riesgos competitivos). También resultó factor de riesgo en el modelo de efectos múltiples. Fue una variable seleccionada por LASSO en el grupo de error alto; la cual arrojó aumento de chance de error. Además, en SONIA resultó ser influyente para predecir el estatus de error. Sin embargo, no se observó relación de ANIM con los errores de género, como en la literatura.

Otro factor crucial para la concordancia parece ser la distancia (lineal o estructural) entre controlador y objetivo. No obstante ello, la variable LDA (larga distancia) no resultó asociada al error en ninguno de los modelos aplicados. Tampoco tuvo efecto alguno el hecho de que la concordancia estuviera conformada por más de dos términos (GRAMS).

Se crearon siete estrategias que se supuso que podrían estar poniendo en acto los alumnos cuando creaban las concordancias. Las siguientes se asociaron al aumento o disminución de la chance / riesgo de error.

- EST1: si la palabra plural del italiano termina en *-i* poner en español plural en *-os*: *libr-i* > *libr-os*. Arrojó un efecto significativo en el modelo logístico mixto; en modelo multinomial para errores de género y mixtos; en el modelo de eventos múltiples y en el de riesgos competitivos para errores de género y mixto. Siempre bajó la chance / riesgo de error cuando se podía aplicar la estrategia a alguno de los términos de la concordancia, respecto de cuando no era posible su aplicación. El modelo LASSO seleccionó el atributo para el grupo de error medio/bajo pero no resultó significativo bajo los coeficientes promediados. Claramente la estrategia 1 facilita las concordancias correctas.

- EST2: si la palabra plural del italiano termina en -e poner en español plural en -as: *cas-e* > *cas-as*. Dicha estrategia bajaba la chance de error únicamente para los errores de -e- epentética en el modelo multinomial. Por lo tanto, fue de alcance muy limitado.
- EST5: si la palabra singular del italiano termina en -e, poner en español el plural en -es: *camion-e* > *camion-es* [casos en los cuales el español coincide con la aplicación del plural con -e- epentética]. Siempre disminuyó la chance / riesgo de error cuando se podía aplicar la estrategia a alguno de los términos de la concordancia, respecto de cuando no era posible su aplicación. Fue estadísticamente significativa en el modelo logístico mixto; en modelo multinomial para errores de -e- epentética y mixtos; en el modelo de eventos múltiples y en el de riesgos competitivos para errores mixtos. También fue seleccionada por LASSO para el grupo de error alto resultando en una disminución de chance de error. La interpretación de este resultado es que los hablantes sacan provecho del parecido de las palabras singulares en español e italiano y forman el plural agregando una -s a la palabra singular italiana. Como en español éstas coinciden con palabras terminadas en consonante que requieren plural en -es; sobrepasan de esta forma la dificultad de insertar plural con -e- epentética. Este factor quizás debe haber influido en el efecto facilitador de ES hallado para los errores de plural.

Además de la animicidad del controlador, se tuvieron en cuenta otras características como la concretud, familiaridad, frecuencia léxica e imaginabilidad. Debido a problemas de colinealidad se crearon dos variables mediante PCA: (i) IMA.CONC: un índice de imaginabilidad más concretud; (ii) FAM.LEX: un índice de familiaridad más frecuencia del controlador. Se las discretizó como FAM.LEX.f e IMA.CONC.f.

En lo que atañe a FAM.LEX.f, esta variable registró siempre un efecto facilitador para la concordancia. Arrojó efectos significativos en el modelo logístico mixto, en el multinomial para los errores de género, plural y mixto; en el de riesgos competitivos para los errores de -e- epentética, plural y mixtos. En el modelo estadístico de evolución de redes, las palabras con controlador de familiaridad / frecuencia léxica alta tienden tanto a generar como a recibir errores. El modelo LASSO seleccionó FAM.LEX.f para el grupo de error medio/bajo, que redujo significativamente la chance de error. Por lo tanto la familiaridad y frecuencia léxica del controlador protegen contra la comisión de errores.

En lo concerniente a IMA.CONC.f, los únicos efectos encontrados estuvieron relacionados con los modelos con datos de redes complejas. En el *Actor-Based Stochastic Model* las palabras de controlador con alta imaginabilidad y concretud fueron origen de error para JAKO. Para MIRKA se halló un efecto de acumulación de error en las palabras respecto de dicho atributo. Respecto de la predicción de error, IMA.CONC.f resultó seleccionado como atributo influyente en el caso de SONIA.

Asimismo se contó la cantidad de errores hechos hasta la instancia anterior a la actual dentro de una misma sesión. Dicho atributo discretizado se denominó CUMRES.f. No hubo efecto de errores acumulados para el modelo logístico mixto ni para los de tiempo hasta un evento pero sí para el multinomial en los errores mixtos, con acumulación media (entre dos y siete errores aproximadamente).

Fabs.SC.f se denominó al índice PCA discretizado de (el logaritmo de) las frecuencias absolutas de TYPES en el corpus propio más las frecuencias absolutas de TYPES en el corpus *EsTenTen*. Esta variable redundó siempre en una baja de la chance / riesgo de error para frecuencias altas. El efecto facilitador se verificó en el modelo logístico mixto; en el modelo multinomial para errores género, plural y mixto; y en el modelo de riesgos competitivos para errores de género y plural.

Se crearon dos variables cuantitativas, STEM y MORF, basadas en la distancia de *Levenstein*, para medir la distancia entre las raíces y las desinencias de cada palabra en español y en italiano. Además se las combinó en un índice para representar la distancia total del par de términos de la concordancia. Luego se discretizaron en los atributos MORF.f (distancia entre terminaciones con tres niveles: alto, medio, bajo) y STEM.f (distancia entre raíces con dos niveles: alto y bajo). En lo que atañe a STEM y MORF se halló un efecto el *Actor-Based Stochastic Model*: las palabras que tienen alta similaridad de raíz / desinencia con el italiano tienden a tener valores similares con el otro término de la concordancia errónea (en la raíz para SONIA y en la desinencia para JAKO). STEM.f únicamente fue seleccionada como atributo influyente para predecir el error en JAKO.

Ahora bien, MORF.f arrojó significatividad estadística pero el tamaño del efecto resultó *inverso* al esperado. Las categorías de distancias medias y altas disminuyeron la chance / riesgo de error respecto de la

distancia baja, que impuso mayor dificultad a los aprendientes. Esto se verificó para la distancia media en el modelo logístico mixto; para las distancias media y alta en errores de plural en el modelo multinomial; para la distancia media en el modelo de eventos múltiples; y, en el modelo de riesgos competitivos, para la distancia media en los errores de *-e-* epentética y plural, y la distancia alta en los errores de plural. Por otra parte, la distancia media y alta se seleccionó como variable influyente para predecir el error en JAKO. ¿Por qué no hubo efecto facilitador de la distancia baja entre las terminaciones de los términos de la concordancia, del español y el italiano? En el capítulo seis se propusieron las siguientes explicaciones. En primera instancia, la presencia de numerosos posesivos, los cuales, en términos de la distancia de *Levenstein* requieren de operaciones de “borrado” y “sustitución” lo cual podría ser “cognitivamente” más demandante. En segundo lugar, la aparición de plurales invariantes en italiano (it. *foto, moto, video*). Luego, se vieron usos de plurales con “diferente-s” [it. *divers-e*] y controladores femeninos, donde se puede confundir la terminación del singular del español con el femenino plural del italiano. También hubo dos usos de un numeral en la centena que en italiano no lleva concordancia. No parecen influir las estrategias EST1, EST2 y EST5 porque no aparecen asociadas a errores de plural, que son la mayoría para MORF.f. El Cuadro 1 resume los resultados hallados para las variables aquí consideradas.

En el capítulo 9 se agruparon las variables que bajaban la chance / riesgo de error para cada tipo de error de concordancia; según los resultados de los modelos multinomial y de tiempo hasta un evento. También se tuvo en cuenta, para cada instancia, la proporción de errores cometidos hasta la instancia anterior, contando desde el inicio de cada sesión. Se crearon tres simulaciones basadas en modelos de sistemas dinámicos: *Lokta-volterra*; *juego de coordinación* y *Gradient Symbolic Computation* [GSC]. Las simulaciones tenían siempre dos atractores estables: «correcto» y «error». Para cada instancia, se intentó recrear un sendero de activación en el «paisaje de eventos mentales» hasta que el sistema se acercara a uno de los atractores. Las variables especificadas en cada tipo de error se tomaron como un contexto que sesgaba el flujo de activación hacia uno de los atractores. Para lograr esto, se plantearon diferentes mecanismos de similitud entre las variables de la instancia que se procesaba y cada uno de los contextos que disminuían el error. En general los modelos aproximaron bien los datos globales de error reales. En cuanto a la aproximación a las series de tiempo reales, SONIA alcanzó el mejor desempeño con el modelo GSC; NATI con el de *Lokta-Volterra*; y JAKO con el del *juego de coordinación*. MIRKA resultó la de peor desempeño en todos los modelos. Todos los modelos dinámicos analizados consideraban al error como atractor que se alcanzaba (o no) en el último paso temporal. Es decir que el error se trató como una representación continua y solamente al final, en el atractor (o cerca de éste), como una representación discreta. Por lo tanto, «Error» o «Correcto» son propiedades emergentes de los sistemas dinámicos; al igual que los patrones globales de error agregados y por series temporales.

El capítulo 10 se avocó a la dinámica de error macroscópica. Se respondieron las siguientes preguntas.

1. ¿Cuáles sesiones tienen dinámica similar para cada aprendiente? Para cada sesión de cada aprendiente se encontraron los “motivos”. Se consideró que cada “motivo” representaba un estado diferente de dinámica, con dos atractores $a = \text{“correcto”}$ y $b = \text{“error”}$ según: (i) atractor a fuerte: “aaa”; (ii) atractor a débil: “aab”, “aba”, “baa”; (iii) atractor b fuerte: “bbb”; (iv) atractor b débil: “bba”, “bab”, “abb”. Se interpretó “fuerte” como permanencia en el atractor y “débil” como un solo apartamiento del atractor hacia el otro. El objetivo consistió en clasificar las sesiones en grupos de dinámica similar, según estados similares. Para ello, se consideró a cada sesión como un «texto» cuyas palabras eran los “motivos”. Se aplicó *latent Dirichlet allocation* (LDA), para encontrar los “tópicos” de los que hablaban los «textos». Se hallaron tres tópicos: (i) el tópico “2”, mayormente asociado al atractor “correcto”; (ii) el tópico “3”, mayormente asociado al atractor “error”; (iii) tópico “1”, de dinámica intermedia. JAKO y SONIA evidenciaban más sesiones en el tópico “2” del atractor correcto. En cambio, para NATI y MIRKA cerca de la mitad de las sesiones pertenecían al atractor fuerte del error (tópico “3”).

Predictor	Efecto [dirección del efecto esperada]	L	M (tipo de error)	EM	RC (tipo de error)	ABSM	LASSO-C-L	PRED
MOD	MOD(1) - MOD(0) [↑]					mod: "adjetivo"		mod=0 (S)
	MOD(2) - MOD(0) [↑]	↑	↑ (mixto)		↓ (plural), ↑ (mixto)	I (N)	↑ (G1)	✓
	MOD(3) - MOD(0) [↑]		↓ (género)		↓ (plural), ↑ (mixto)		↑ (G1)	mod=3 (J)
Fabs.SC.f	Fabs.SC.f(1) - Fabs.SC.f(0) [↓]	↓	↓ (-e- ep, plural, mixto)		↓ (género, plural)			
MORF.f	MORF.f(1) - MORF.f(0) [↑]	↓	↓↓ (plural)	↓	↓ (-e- ep., plural)			✓ (J)
	MORF.f(2) - MORF.f(0) [↑↑]		↓ (plural)		↓↓ (plural)			✓ (J)
CUMRES.f	CUMRES(1) - CUMRES(0) [↑]		↑ (mixto)					
	CUMRES(2) - CUMRES(0) [↑↑]							
FAM.LEX.f	FAM.LEX.f(1) - FAM.LEX.f(0) [↓]	↓			↓ (-e- ep., mixto, plural)	I, II (S)	↓ (G2)	
EST1	EST1(1) - EST1(0) [↓]	↓	↓ (género, mixto)	↓	↓ (género, mixto)			
EST2	EST2(1) - EST2(0) [↓]		↓ (-e- ep.)					
EST3	EST3(1) - EST3(0) [↑]							
EST4	EST4(1) - EST4(0) [↓]							
EST5	EST5(1) - EST5(0) [↓]	↓	↓↓ (-e- ep., mixto)	↓	↓ (mixto)		↓ (G1)	
EST6	EST6(1) - EST6(0) [↓]							
EST7	EST7(1) - EST7(0) [↑]							
ANIM	ANIM(1) - ANIM(0) [↑]		↑ (-e- ep.)	↑	↑ (plural)		↑ (G1)	✓ (S)
ES	ES(1) - ES(0) [↑]					III (S, M)		✓ (S, J)
	ES(2) - ES(0) [↑↑]		↓↓ (plural)			I (N)		
LDA	LDA(1)-LDA(0) [↑]							
GRAMS	GRAMS(1)-GRAMS(0) [↑]							
STEM.f	STEM.f(1)-STEM.f(0) [↑]							✓ (J)
IMA.CONC.f	IMA.CONC.f(1) - IMA.CONC.f(0) [↓]					I (J), V (M)		✓ (S)
ESP1	"os","as","es","es*" vs "is"/"us" (referencia) (primer término de concordancia)					"es", III (S, N, M)	↓ "os","es*" vs ref. (G1) ↑ "as" vs "os" (G1) ↑ "as" vs "os" (G2) ↓ "os" vs ref. (G2)	"as" (J) ✓
ESP2	"as","es","es*" vs "os" (referencia) (segundo término de concordancia)						↓ "as" vs "os" (G2) ↑ "es" vs "as" (G2)	
MORF	cuantitativa					III (J)		
STEM	cuantitativa					III (S)		

Cuadro 12.1: Resumen de resultados. Leyenda: (↑) = riesgo aumenta; (↓) = riesgo disminuye; (↑↑/↓↓) = mayor tamaño de efecto. Hipótesis (ABSM): I, II, III, IV, V; S = Sonia; N = Nati; J = Jako; M = Mirka; G1 = grupo 1; G2 = grupo 2; ✓ = predictor seleccionado. L = modelo logístico mixto; M = modelo multinomial mixto; EM = modelo de eventos múltiples; RC = modelo de riesgos competitivos; ABSM = modelo estocástico basado en actores; LASSO-C-L = LASSO mixto, clustering y modelo logístico mixto sobre grupos; PRED = atributos influyentes para predicción de estatus de error; ref. = nivel de referencia; -e- ep. = error de -e- epentética.

2. ¿Qué regímenes de dinámica diferente es posible identificar? Se utilizó el plano Entropía (H_s) - Complejidad (C_s). La medida de entropía mide la incertidumbre sobre el estatus de error de la palabra que aparecerá a continuación en una serie. Una situación de máximo / mínimo conocimiento-incertidumbre no es compleja; o sea, sencillamente existe completo orden o desorden. La medida de complejidad capta los estados intermedios entre ambos extremos. A través de un plano de incertidumbre / complejidad, se creó un espacio de fase del sistema de aprendizaje para cada aprendiente, en el que cada estado (sesión) es un punto en dicho plano; y se investigó el movimiento / trayectoria hacia el punto atractor que representa al hablante de nivel superior (baja entropía, baja complejidad) o una región de atracción que represente a un hablante principiante (alta entropía, baja complejidad), una vez cruzado un punto crítico. Se ubicó el atractor "sin error" en el punto ($H_s = 0$; $C_s = 0$) y el del "error" en el punto ($H_s = 1$; $C_s = 0$). Se estableció el punto crítico (arbitrario) $f_{error} = 0,21$ para

designar aquel punto más allá del cual se ingresa en el *basin of attraction* del atractor NO_ERROR; y el punto crítico (arbitrario) $f_{error} = 0,35$ para designar aquel punto más allá del cual se ingresa en el *basin of attraction* del atractor ERROR. La región intermedia se designó de TRANSICIÓN entre ambos regímenes. Se observó que tanto JAKO como SONIA estaban pasando del régimen de TRANSICIÓN al de NO_ERROR, con JAKO más avanzado. En contraste, para NATI y MIRKA se hallaron más sesiones en el tránsito del régimen ERROR al de TRANSICIÓN. Es preciso agregar que la medida de complejidad C_s fue seleccionada por LASSO en el grupo de error alto y resultó ser factor de protección contra el error.

3. ¿Cuáles sesiones evidencian cambio de dinámica?. Se empleó *Recurrence Quantification Analysis* (RQA), una técnica que permite medir propiedades de atractores en el espacio de fase [phase space] reconstruido. Los atractores se definen en términos de regiones en el espacio (multi-)dimensional que son visitadas una y otra vez por el sistema dinámico. RQA mide el grado de nuevas visitas del sistema en dichas regiones y computa ciertas medidas basadas en dichas recurrencias. Para cada alumno, se creó una serie temporal bivariada con categorías por tipo de error en el alfabeto usual de tipos de error $A = \{0; 1; 2; 3; 4\}$, discriminado por término de concordancia. El interés radicó en evaluar la sincronización entre instancias correctas en ambos términos. SONIA cambiaba la dinámica en las sesiones 3,7,8,9,10,12; NATI, en las sesiones 4,14; JAKO en las sesiones 3,6,12,14; y MIRKA en las sesiones 1,2,6,8,12. Las medidas tuvieron valores más bajos para NATI y MIRKA, como era esperable. RQA reveló más sesiones con cambio de dinámica (por ejemplo, más períodos de *laminaridad*) en SONIA y MIRKA. O sea que MIRKA, aunque cometía más errores, estos se encontraban más lejos en el tiempo durante más sesiones que en NATI.
4. ¿Dónde están las transiciones de dinámica?. Se calculó la *entropía de permutación* por medio de una ventana móvil de largo $w = 80$ para cada aprendiente. Se aplicó *Change Point Analysis* para descubrir el punto de cambio de locación. Para SONIA el cambio se produce entre finales de la sesión 5 y mediados de la sesión 8; para NATI, durante las sesiones 5 y mediados de la 7; en JAKO, durante el periodo entre las sesiones 9 y principios de la 11, para MIRKA; entre las sesiones 2 y principios de la 3. Sin embargo para MIRKA no hubo mucha certeza de que el cambio estuviera allí. Se observó que SONIA y JAKO cambiaban a una mejor dinámica; pero NATI y MIRKA a una peor. También se aplicó una serie de indicadores que anticipan bifurcaciones. La evidencia de transiciones arrojó coincidencia entre los indicadores y el análisis de *change point* hacia la sesión 8 en SONIA, y hacia la 11 en JAKO. En NATI parece haber un cambio entre las sesiones 7-8, según ambas técnicas. En cuanto a MIRKA, los indicadores mostraron una transición hacia la sesión 11. Por último, el análisis de potenciales apuntó hacia dos atractores, justificando la posibilidad de transiciones entre ellos. Entre los indicadores, aquel más consistente resultó el modelo de *drift-difusion-jump*. Con todo, muchas veces los indicadores o bien no fueron concluyentes o bien demostraron discrepancia entre ellos.

En lo que atañe a los resultados sobre la predicción del estatus de error, en general no se logró superar el 80% de precisión. SVM resultó ser el clasificador de mejor desempeño. SONIA y JAKO alcanzaron mejores niveles de precisión que NATI y MIRKA. Los atributos derivados del grafo fueron seleccionados como influyentes para todos los alumnos. Es preciso notar que ni NATI ni MIRKA tuvieron atributos influyentes de concordancia seleccionados y con ellos los clasificadores obtuvieron peor resultado.

12.3. A modo de cierre: algunos consejos para la enseñanza de la concordancia.

A modo de cierre, se presentan algunos consejos para enseñar la concordancia en español L2, a la luz de los resultados de la tesis.

En lo que respecta al *feedback*, sería deseable esperar a que los alumnos cometan suficientes errores (creen variabilidad) y logren notar diferencias en su producción con sus compañeros. En todo caso se

debería optar por un *feedback* implícito (*Promts*) y *online*. Este sería más efectivo cerca de una bifurcación o cambio de régimen, donde típicamente aumenta la variabilidad.

Otra propuesta sería realzar la pronunciación de *-s / -es* en el input y exhortar a los alumnos a que los pronuncien claramente, especialmente para que los morfemas se mantengan activos en la memoria a corto plazo. El principal desafío es que los alumnos italianos deben crear una posición a ser llenada con los morfemas de plural, que no existe en su propia lengua.

También sería prudente incluir instrucción basada en *Focus on Form* y presentar las concordancias primero como *chunks* inanalizados para aumentar las posibilidades de que los alumnos noten patrones frecuentes del tipo [os - os]; etc. O sea, decir *las ciudades* [as - es] o *grandes ciudades* [es - *es] y no únicamente *ciudades* cuando se presente el léxico o bien utilizando ejercicios con colocaciones. Recuérdese que se halló una relación entre frecuencia de los patrones de concordancia (variable TYPE) y probabilidad de cometer errores. Por otra parte, se recomendaría enseñar primero concordancias que siguen esquemas armónicos (Por ej.: [as - as: *mesas coloradas*]) y luego pasar a los no armónicos (Por ej.: [es - as: *costumbres argentinas*]). Aunque la cantidad de los términos de la concordancia no influyó en el error en este trabajo, también sería prudente comenzar con concordancias de dos términos y luego pasar a las de tres, cuatro, etc.

Tampoco resultó significativa la concordancia a larga distancia. Pero existe amplia evidencia en la literatura de que constituye un factor fundamental a considerar. Por lo tanto se recomienda el orden: concordancia nominal > concordancia predicativa / clíticos > concordancia en subordinada.

Las estrategia EST5 resultó estadísticamente significativa en varios modelos, lo cual estaría ayudando a los alumnos a remontar las concordancias con controladores terminados en consonante. Aunque EST6 no influyó en el error, podría ser positivo enseñar ambas a los hablantes italianos.

Asimismo, la variable MORF.f llevó a un efecto inverso al esperado. El análisis de los ejemplos reveló dificultades con los posesivos y los plurales invariantes del italiano.

Aunque las palabras de número nocional son pocas en español (Ej.: *dinero* [singular]), es preciso identificar las diferencias con el italiano (it. *soldi* [plural]).

Otro efecto notoriamente presente en la literatura y en la presente tesis fue que la concordancia con el artículo definido resulta más fácil que aquella establecida con los determinantes y los adjetivos. Por lo tanto el instructor debería concentrarse en estos últimos. En especial, notando los casos donde el singular del español coincide con el plural del italiano (Ej.: esp. *este* [masc. sing], it. *queste* [fem. sing.]; esp. *diferente* [masc. / fem. sg.], it. *diverse* [fem. sg.]).

La chance / riesgo de error aumentó con controladores animados y disminuyó con controladores familiares y frecuentes. En consecuencia, es aconsejable prestar más atención a sustantivos animados, poco familiares o infrecuentes porque inducirían a error.

La teoría de los sistemas dinámicos nos invita a considerar al lenguaje como fenómeno complejo y emergente. Cada actividad en la clase debería mantener dicha complejidad (Larsen Freeman, 2003). Cuando hablan, los alumnos deben controlar varios niveles al mismo tiempo (fonología, morfo-sintaxis, semántica, etc.). El instructor debe tener en cuenta esta complejidad y no tratar de reducirla por razones pedagógicas; por ejemplo haciendo ejercicios tipo “drill” para la concordancia (que serían más manejables para los alumnos pero anularía la complejidad inherente). El alumno debe manejar la concordancia mientras trata de controlar gran cantidad de otras variables. Consecuentemente, ser “más didáctico” no significa “cortar la lengua en partes” para facilitarle la vida al estudiante.

Esta tesis ha sido un intento de operacionalizar la noción de “cambio en el desarrollo de L2” usando herramientas cuantitativas no lineales y estadísticas que logren captar la temporalidad. La mayoría de los formalismos de la teoría lingüística (minimalismo, gramáticas a unificación, etc.) están hechos para la sincronía¹, donde las distribuciones de probabilidad de los patrones se han estabilizado. No parece útil aplicarlos a la adquisición, donde la lengua se encuentra en flujo constante. Las ciencias de la complejidad iluminan otro camino: pasar de teorías que manipulan símbolos discretos a otra donde lo importante es un flujo de estados “mezcla”; con patrones / configuraciones que modifican sus distribuciones de probabilidad hasta alcanzar atractores discretos estables.

Las conclusiones de esta tesis tienen la limitación de haber sido derivadas de cuatro estudios de caso.

¹Con la notable excepción de la *Gramática Fluida de las Construcciones* de Luc Steels y la *Gramática Armónica* de Paul Smolensky.

Quedaría pendiente extender el corpus de datos a más alumnos de diferentes L1. Sin embargo ni la predicción ni la generalización (inferencia poblacional) pueden constituir objetivos de la agenda de investigación de este enfoque. Esto es así porque el proceso no lineal que se aduce responsable del desarrollo lingüístico causa alta variabilidad. Por lo tanto, no se debería estudiar la adquisición buscando un estado discreto o etapas de orden de adquisición universales; o bien usando estudios croslingüísticos o modelos estadísticos que no tengan en cuenta la temporalidad o más fuentes de varianza. Si se considera al lenguaje como un flujo no lineal, continuo, fractal y adaptativo entonces las técnicas deben pasar por las series de tiempo (no lineales), el análisis de la variabilidad y cambios de régimen, estudios longitudinales y simulaciones; como se trató de hacer en esta tesis.

La teoría de sistemas dinámicos complejos trata sobre transiciones. Lo importante es identificar los factores que influyen en las trayectorias individuales y usarlos para guiar el proceso de aprendizaje; sin preocuparse por realizar predicciones para la adquisición (imposibles dada la naturaleza no lineal del fenómeno) o la generalización a la población de aprendientes (imposible por usar estudios longitudinales de caso) [Caspi, 2010].

Bibliografía

- Acuña-Fariña, Juan Carlos (2009). “The linguistics and psycholinguistics of agreement: a tutorial overview”. En: *Lingua* 119, págs. 389-424.
- (2018). “Aspects of the constructional nature of agreement”. En: *Constructions* 1, págs. 1-24.
- Adger, David (2003). *Core Syntax*. Oxford University Press.
- Alarcón, Irma (2009). “The processing of gender agreement in L1 and L2 Spanish: Evidence from reaction time data”. En: *Hispania* 92.4, págs. 814-828.
- (2011). “Spanish grammatical gender under complete and incomplete acquisition: Early and late bilinguals’ linguistic behavior within the noun phrase”. En: *Bilingualism: Language and Cognition* 14.3, págs. 332-350.
- Alemán Bañón, José, Robert Fiorentino y Alison Gabriele (2014). “Morphosyntactic processing in advanced second language (L2) learners: An event-related potential investigation of the effects of L1-L2 similarity and structural distance”. En: *Second Language Research* 30.3, págs. 275-306.
- Ambadiang, Théophile (1999). “La flexión nominal: Género y número.” En: *Gramática descriptiva de la lengua española: Vol. 3. Entre la oración y el discurso. Morfología*. Ed. por Ignacio Bosque y Violeta Demonte. Madrid: Espasa Calpe. Cap. 74, págs. 4843-4914.
- Amigó, José María, Samuel Zambrano y Miguel A. F. Sanjuán (2008). “Combinatorial detection of determinism in noisy time series”. En: *Europhysics Letters* 83.6, pág. 60005.
- Andersen, Roger W. (1990). “Models, processes, principles and strategies: Second language acquisition inside and outside the classroom”. En: *Second language acquisition-foreign language learning*. Ed. por Bill VanPatten y James F. Lee. Clevedon, UK: Multilingual Matters, págs. 45-78.
- Andersen, Roger y Yasuhiro Shirai (1994). “Discourse Motivations for Some Cognitive Acquisition Principles”. En: *Studies in Second Language Acquisition* 16, págs. 133-156.
- Anderson, John R. y col. (2004). “An integrated theory of the mind”. En: *Psychological Review* 111.4, págs. 1036-1060.
- Aslin, Richard N. y Elissa L. Newport (2012). “Statistical learning: From learning items to generalizing rules”. En: *Current Directions in Psychological Science* 21, págs. 170-176.
- Atkinson, Richard C. y Richard M. Shiffrin (1968). “Human memory: A proposed system and its control processes”. En: *The Psychology of Learning and Motivation*. Ed. por K. W. Spence y J. T. Spence. Vol. 2. London: Academic Press, págs. 89-195.
- Baars, Bernard J. (1997). “In the theatre of consciousness: Global workspace theory, a rigorous scientific theory of consciousness”. En: *Journal of Consciousness Studies* 4, págs. 292-309.
- (2002). “The conscious access hypothesis: Origins and recent evidence”. En: *Trends in Cognitive Sciences*. 6.1, págs. 47-52.
- Baba, Kyoko y Ryo Nitta (2013). “Phase Transitions in Development of Writing Fluency From a Complex Dynamic Systems Perspective”. En: *Language Learning* 64.1, págs. 1-35.
- Bandt, Christoph y Bernd Pompe (2002). “Permutation Entropy: A Natural Complexity Measure for Time Series”. En: *Physical review letters* 88.17, pág. 174102. DOI: 10.1103/PhysRevLett.88.174102.
- Barabási, Albert-László y Márton Pósfai (2016). *Network science*. Cambridge University Press.
- Bardovi-Harlig, Kathleen (2015). “One Functional Approach to SLA: The Concept-Oriented Approach”. En: *Theories in second language acquisition: an introduction*. Ed. por Bill Van Patten y Jessica Williams. second. Routledge. Cap. 4.
- Bischi, Bernd y col. (2016). “mlr: Machine Learning in R”. En: *Journal of Machine Learning Research* 17.170, págs. 1-5. URL: <http://jmlr.org/papers/v17/15-066.html>.

- Blangiardo, Marta y Michela Cameletti (2016). *Spatial and Spatio-Temporal Bayesian Models with R-INLA*. John Wiley y Sons.
- Blei, David, Andrew Ng y Michael Jordan (2003). “Latent Dirichlet Allocation”. En: *Journal of Machine Learning Research* 3, págs. 993-1022.
- Blythe, Richard y Alan McKane (2007). “Stochastic Models of Evolution in Genetics, Ecology and Linguistics”. En: *Journal of Statistical Mechanics*.
- Bojanowski, Michal y Bartosz Chrol (2019). *Proximity-based Methods for Link Prediction in Graphs with R package 'linkprediction'*. URL: <http://recon.icm.edu.pl/wp-content/uploads/2019/05/linkprediction.pdf>.
- Bond, Kristi y col. (2011). “Individual Differences and the Role of the L1 in L2 Processing: An ERP Investigation”. En: *Proceedings of the 11th GASLA Conference*. Ed. por Julia Herschensohn y Darren Tanner. Somerville, MA: Cascadilla, págs. 17-29.
- Bonilla, Carrie (2014). “From number agreement to the subjunctive: Evidence for Processability Theory in L2 Spanish”. En: *Second Language Research* 31, págs. 53-74.
- Bonnini, Stefano y col. (2014). *Nonparametric Hypothesis Testing: Rank and Permutation Methods with Applications in R*. John Wiley & Sons, Ltd.
- Börjars, Kersti, Rachel Nordlinger y Louisa Sadler (2019). *Lexical-Functional Grammar: An Introduction*. Cambridge University Press.
- Bot, Kees De, Wander Lowie, Steven Thorne y col. (2013). “Dynamic Systems Theory as a comprehensive theory of second language development”. En: *Contemporary Approaches to Second Language Acquisition*. Ed. por María del Pilar García Mayo, María Junkal Gutierrez Mangado y María Martínez Adrián. Benjamins. Cap. 10, págs. 199-220.
- Bot, Kees De, Wander Lowie y Marjolijn Verspoor (2007). “A Dynamic Systems Theory approach to second language acquisition”. En: *Bilingualism: Language and Cognition* 10.1, págs. 7-21.
- Bresnan, Joan y col. (2016). *Lexical-functional syntax*. Wiley-Blackwell.
- Bruhn de Garavito, Joyce (2008). “Acquisition of the Spanish plural by French L1 speakers: The role of transfer”. En: *The role of features in second language acquisition*. Ed. por Juana Liceras, Helmut Zobl y Helen Goodluck. Mahwah, NJ: Lawrence Erlbaum, págs. 270-298.
- Bruhn de Garavito, Joyce y Lydia White (2002). “The Second Language Acquisition of Spanish DPs: The Status of Grammatical Features”. En: *The Acquisition of Spanish Morphosyntax: The L1/L2 Connection*. Ed. por Ana Teresa Pérez-Leroux y Juana Muñoz Liceras. Dordrecht: Kluwer, págs. 153-178.
- Burnham, Kenneth P. y David R. Anderson (2010). *Model selection and multimodel inference: a practical information-theoretic approach*. Springer.
- Byrne, David (2002). *Interpreting quantitative data*. London: Sage.
- Carpenter, R. Stephen y W. A. Brock. (2011). “Early warnings of unknown nonlinear shifts: a nonparametric approach”. En: *Ecology* 92.12, págs. 2196-2201.
- Carrera Díaz, Manuel (1989). *Manual de Gramática Italiana*. Barcelona: Ariel.
- Caspi, Tal (2010). “A Dynamic Perspective on Second Language Development”. Tesis doct. University of Groningen.
- Chawla, Nitesh V. y col. (2002). “Smote: Synthetic minority over-sampling technique”. En: *Journal of Artificial Intelligence Research* 16, págs. 321-357.
- Cho, Pyeong Whan, Matthew A Goldrick y Paul Smolensky (2017). “Incremental parsing in a continuous dynamical system: Sentence processing in gradient symbolic computation”. En: *Linguistics Vanguard* 3.1.
- Cho, Pyeong Whan y col. (2018). “Dynamic encoding of structural uncertainty in gradient symbols”. En: *Proceedings of the 8th Workshop on Cognitive Modeling and Computational Linguistics, CMCL 2018, Salt Lake City, Utah, USA, January 7, 2018*, págs. 19-28.
- Chomsky, Noam (2000). “Minimalist Inquiries: The Framework”. En: *Step by Step: Essays on Minimalist Syntax in honor of Howard Lasnik*. Ed. por Roger Martin, David Michaels y Juan Uriagereka. Cambridge, MA.: The MIT Press.
- (2001). “Derivation by Phase”. En: *Ken Hale: A Life in Language*. Ed. por Michael Kenstovicz. Cambridge, MA: The MIT Press, págs. 1-54.

- Clahsen, Harald y Peter Muysken (1986). "The availability of Universal Grammar to adult and child learners: A study of the acquisition of German word order". En: *Second Language Research* 2, págs. 93-109.
- Cohen, Jacob (1988). *Statistical power analysis for the behavioral sciences*. Hillsdale, NJ: Lawrence Erlbaum Associates.
- Corbett, Greville (1991). *Gender*. Cambridge: Cambridge University Press.
- (2006). *Agreement*. Cambridge: Cambridge University Press.
- Cover, Thomas M. y Joy A. Thomas (1991). *Elements of Information Theory*. New York: Wiley.
- Croft, William (2000). *Explaining Language Change: An Evolutionary Approach*. Longman Linguistics Library. Harlow, UK: Pearson Education.
- (2001). *Radical Construction Grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press.
- Crutchfield, James P. y David P. Feldman (2003). "Regularities unseen, randomness observed: Levels of entropy convergence". En: *Chaos: An Interdisciplinary Journal of Nonlinear Science* 13.1, págs. 25-54.
- Dakos, Vasilis, Stephen R. Carpenter y col. (2012). "Methods for Detecting Early Warnings of Critical Transitions in Time Series Illustrated Using Simulated Ecological Data". En: *Plos One* 7.7, e41010.
- Dakos, Vasilis, Stephen R. Carpenter y col. (2015). "Resilience indicators: prospects and limitations for early warnings of regime shifts". En: *Philosophical Transactions B* 370.1659, págs. 1-10.
- Dalrymple, Mary (2001). *Syntax and Semantics 34: Lexical Functional Grammar*. Vol. 34.
- Daniel, Sorensen y Daniel Gianola (2002). *Likelihood, Bayesian and MCMC Methods in Quantitative Genetics*. Springer.
- Davis, Colin J. y Manuel Perea (2005). "BuscaPalabras: A program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish". En: *Behavior Research Methods* 37.4, págs. 665-671.
- Dehaene, Stanislas y Lionel Naccache (2001). "Towards a cognitive neuroscience of consciousness: Basic evidence and a workspace framework". En: *Cognition* 79, págs. 1-37.
- Dehling, Herold, Roland Fried y Martin Wendler (2018). *A Robust Method for Shift Detection in Time Series*. URL: <http://martin-wendler.de/publications/>.
- DeKeyser, Robert M. (2015). "Skill acquisition theory". En: *Theories in second language acquisition: an introduction*. Ed. por B. Van Patten y J. Williams. second. Routledge. Cap. 6.
- Deo, Ashwini (2015). "The semantic and pragmatic underpinnings of grammaticalization paths: The progressive to imperfective shift". En: *Semantics and Pragmatics* 8, págs. 1-52.
- Diessel, Holger (2015). "Usage-based construction grammar". En: *Handbook of Cognitive Linguistics*. Ed. por Ewa Dabrowska y Dagmar Divjak. Berlin: Mouton de Gruyter, págs. 296-322.
- Dörnyei, Zoltan (2014). "Researching complex dynamic systems: Retrospective qualitative modelling in the language classroom". En: *Language Teaching* 47.1, págs. 80-91.
- Doughty, Catherine (2003). "Instructed SLA: constraints, compensation, and enhancement". En: *The handbook of second language acquisition*. Ed. por Doughty Catherine J. y Michael H. Long. Oxford: Blackwell.
- Dussias, Paola E. y col. (2013). "When Gender and Looking go Hand in Hand: grammatical gender Processing in L2 Spanish". En: *Studies in Second Language Acquisition* 35, págs. 353-387.
- Eckmann, J.P., S.O. Kamphorst y D. Ruelle. (1987). "Recurrence plots of dynamical systems". En: *Europhys. Lett. (EPL)* 4.9, págs. 973-977.
- Ellis, Nick C. (2001). "Memory for language". En: *Cognition and Second Language Instruction*. Ed. por P. Robinson. Nueva York: Cambridge University Press, págs. 33-68.
- (2003). "Constructions, chunking, and connectionism: The emergence of second language structure". En: *Handbook of Second Language Acquisition*. Ed. por Catherine Doughty y Michael H. Long. Oxford: Blackwell, págs. 63-103.
- (2005). "At the interface: Dynamic interpretations of explicit and implicit language knowledge." En: *Studies in Second Language Acquisition* 27, págs. 305-362.
- (2015). "Usage based approaches to SLA". En: *Theories in second language acquisition: an introduction*. Ed. por Bill Van Patten y Jessica Williams. second. Routledge. Cap. 5.

- Ellis, Nick C. (2019). "Essentials of a Theory of Language Cognition". En: *The Modern Language Journal* 103, págs. 39-60.
- Ellis, Nick y Fernando Ferreira-Jr. (2009). "Construction Learning as a Function of Frequency, Frequency Distribution, and Function". En: *The Modern Language Journal* 93.3, págs. 370-385.
- Enders, Craig K. (2010). *Applied Missing Data Analysis*. New York: Guilford Press.
- Eskildsen, Soren (2012). "L2 negation constructions at work". En: *Language Learning* 62, págs. 335-372.
- Español-Echevarría, Manuel y Philippe Prévost (2004). "Acquiring Number Specification on L2 Spanish Quantifiers: Evidence against the Rich Agreement Hypothesis." En: *Proceedings of the 2003 Conference on Generative Approaches to Language Acquisition*. Ed. por J. Van Kampen y S. Baauw. Utrecht: LOT, págs. 151-162.
- Fernández-García, Marisol (1999). "Patterns of gender agreement in the speech of second language learners". En: *Advances in Hispanic linguistics: Papers from the 2nd Hispanic linguistics symposium*. Ed. por Javier Gutiérrez-Rexach y Fernando Martínez-Gil. Somerville, MA: Cascadilla Press, págs. 3-15.
- Finnemann, Michael D. (1992). "Learning agreement in the noun phrase: The strategies of three first-year Spanish students". En: *IRAL* 30.2, págs. 121-136.
- Foote, Rebecca (2011). "Integrated Knowledge of Agreement in Early and Late English-Spanish Bilinguals". En: *Applied Linguistics* 32, págs. 187-220.
- (2015). "The production of gender agreement in native and L2 Spanish: The role of morphophonological form". En: *Second Language Research* 31, págs. 343-373.
- Foucart, Alice y Cheryl Frenck-Mestre (2011). "Grammatical gender processing in L2: Electrophysiological evidence of the effect of L1-L2 syntactic similarity". En: *Bilingualism: Language and Cognition* 14, págs. 379-399.
- Franceschina, Florencia (2001). "Morphological or syntactic deficit in near-native speakers? An assessment of some current proposals". En: *Second Language Research* 17, págs. 213-247.
- Friedman, Jerome, Trevor Hastie y Rob Tibshirani (2010). "Regularization Paths for Generalized Linear Models via Coordinate Descent". En: *Journal of Statistical Software* 33.1, págs. 1-22.
- García, Adolfo M. (2012). "La organización anatómica del cerebro bilingüe: datos existentes y nueva evidencia a favor del modelo declarativo / procedural". En: *Revista Argentina de Neuropsicología* 20, págs. 1-23.
- Gass, Susan M. y Alison Mackey (2015). "Input, Interaction, and Output in Second Language Acquisition". En: *Theories in second language acquisition: an introduction*. Ed. por Bill Van Patten y Jessica Williams. second. Routledge. Cap. 10.
- Gelman, Andrew (2006). "Prior Distributions for Variance Parameters in Hierarchical Models". En: *Bayesian Analysis* 1.3, págs. 515-533.
- Gelman, Andrew y col. (2004). *Bayesian Data Analysis*. Chapman y Hall/CRC.
- Gillon Dowens, Margaret, Taomei Guo y col. (2011). "Gender and Number Processing in Chinese Learners of Spanish: Evidence from Event Related Potentials". En: *Neuropsychologia* 49.7, págs. 1651-1659.
- Gillon Dowens, Margaret, Marta Vergara y col. (2010). "Morphosyntactic Processing in Late Second Language learners". En: *Journal of Cognitive Neuroscience* 22.8, págs. 1870-1887.
- Goldberg, Adele E. (2013). "Constructionist Approaches". En: *The Oxford Handbook of Construction Grammar*. Ed. por Thomas Hoffmann y Graeme Trousdale. Oxford University Press.
- Goldschneider, Jennifer M. y Robert M. DeKeyser (2001). "Explaining the "natural order of L2 morpheme acquisition in English: A meta-analysis of multiple determinants". En: *Language Learning* 51, págs. 1-50.
- Gómez, Álvaro (2014). "Modelos de mixturas finitas para la caracterización y mejora de las redes de monitorización de la calidad del aire." Tesis de maestría. Universidad de Granada.
- Gómez, Rebecca L. (2002). "Variability and detection of invariant structure". En: *Psychological Science* 13, págs. 431-436.
- Grambsch, Patricia M. y Terry M. Therneau (1994). "Proportional hazards tests and diagnostics based on weighted residuals." En: *Biometrika* 81.3, págs. 515-526.
- Greenacre, Michael (2008). *La práctica del análisis de correspondencias*. Fundacion BBVA.

- Groll, Andreas y Gerhard Tutz (2014). “Variable Selection for Generalized Linear Mixed Models by L1-Penalized Estimation”. En: *Statistics and Computing* 24.2, págs. 137-154.
- Grün, Bettina y Friedrich Leisch (2007). “Flexmix: An R package for finite mixture modelling”. En: *R News* 7.1, págs. 8-13.
- Grüter, Theres, Casey Lew-Williams y Anne Fernald (2012). “Grammatical gender in L2: A production or a real-time processing problem?”. En: *Second Language Research* 28, págs. 191-215.
- Gu, Fei y Kristopher J. Preacher (2014). “A State Space Modeling Approach to Mediation Analysis”. En: *Journal of Educational and Behavioral Statistics* 39.2, págs. 117-143.
- Guyon, Isabelle y col. (2002). “Gene Selection for Cancer Classification Using Support Vector Machines”. En: *Machine Learning* 46.1, págs. 389-422.
- Hadfield, Jarrod D. (2010a). “MCMC methods for Multi-response Generalised Linear Mixed Models: The MCMCglmm R Package”. En: *Journal of Statistical Software* 33.2, págs. 1-22.
- (2010b). *MCMCglmm: Markov chain Monte Carlo methods for Generalised Linear Mixed Models*. URL: <http://cran.r-project.org/web/packages/MCMCglmm/index.html>.
- Halkidi, Maria, Yannis Batistakis y Michalis Vazirgiannis (2001). “On Clustering Validation Techniques”. En: *Journal of Intelligent Information Systems* 17, págs. 107-145.
- Harris, James W. (1991). “The exponence of gender in Spanish”. En: *Linguistic Inquiry* 22, págs. 27-62.
- Hastie, Trevor, Jerome Friedman y Robert Tibshirani (2017). *The Elements of statistical learning: data mining, inference, and prediction*. Springer.
- Hawkins, Roger y C. Chan (1997). “The partial accessibility of Universal Grammar in second language acquisition: The failed functional features hypothesis”. En: *Second Language Research* 13, págs. 187-226.
- Hedeker, Donald y Robert D. Gibbons (2006). *Longitudinal Data Analysis*. John Wiley & Sons.
- Heidelberger, Philip y Peter D. Welch (1981). “A spectral method for confidence interval generation and run length control in simulations”. En: *Communications of the ACM*. 24.233-245.
- Hell, Janet G. van y Natasha Tokowicz (2010). “Event-Related Brain Potentials and Second Language Learning: Syntactic Processing in Late L2 Learners at Different L2 Proficiency Levels”. En: *Second Language Research* 26.1, págs. 43-74.
- Hennig, Christian (2007). “Cluster-wise assessment of cluster stability”. En: *Computational Statistics and Data Analysis* 52, págs. 258-271.
- (2013). “How many bee species? A case study in determining the number of clusters”. En: *Data Analysis, Machine Learning and Knowledge Discovery*. Ed. por Myra Spiliopoulou, Barbara Schmidt-Thieme y Ruth Janning. Berlin: Springer, págs. 41-49.
- (2019). “Cluster Validation by Measurement of Clustering Characteristics Relevant to the User”. En: *Data Analysis and Applications 1*. Ed. por Christos H. Skiadas y James R. Bozeman. John Wiley & Sons, Ltd. Cap. 1, págs. 1-24.
- Hilbe, Joseph M. (2009). *Logistic Regression Models*. Boca Raton, FL.: Chapman & Hall/CRC.
- Hosmer, David W., Stanley Lemeshow y Susanne May (2008). *Applied Survival Analysis: Regression Modeling of Time to Event Data*. Wiley.
- Hosmer, David W., Stanley Lemeshow y Rodney X. Sturdivant (2013). *Applied logistic regression*. New Jersey: Wiley.
- Huffaker, Ray, Marco Bittelli y Rodolfo Rosa (2017). *Nonlinear Time Series Analysis With R*. Oxford University Press.
- Jakubíček, Milos y col. (2010). “Fast Syntactic Searching in Very Large Corpora for Many Languages”. En: *PACLIC*, págs. 741-47.
- Jescheniak, Jörg D. y Willem J. M. Levelt (1994). “Word frequency effects in speech production: Retrieval of syntactic information and of phonological form”. En: *Journal of Experimental Psychology: Learning, Memory and Cognition* 20, págs. 824-843.
- Jurafsky, Dan y James H. Martin (ago. de 2017). “An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition.” Third Edition draft.
- Kaplan, Daniel y Leon Glass (1995). *Understanding Nonlinear Dynamics*. New York: Springer.
- Kateri, María (2014). *Contingency Table Analysis. Methods and Implementation Using R*. Springer.

- Keating, Gregory D. (2009). "Sensitivity to Violations of Gender Agreement in Native and Nonnative Spanish: An Eye-Movement Investigation". En: *Language Learning* 59.3, págs. 503-535.
- (2010). "The Effects of Linear Distance and Working Memory on the Processing of Gender Agreement in Spanish". En: *Research in Second Language Processing and Parsing*. Ed. por Bill VanPatten y Jill Jegerski. Philadelphia: John Benjamins, págs. 113-134.
- Kilgarriff, Adam y col. (2014). "The Sketch Engine: ten years on". En: *Lexicography* 1, págs. 7-36.
- Kim, Jon-Bok y Peter Sells (2008). *English Syntax: an introduction*. The University of Chicago Press.
- Kira, Kenji y Larry A. Rendell (1992). "The Feature Selection Problem: Traditional Methods and a New Algorithm". En: *Proceedings of the Tenth National Conference on Artificial Intelligence*. AAAI'92. San Jose, California: AAAI Press, págs. 129-134. ISBN: 0-262-51063-4.
- Klein, John P. y Melvin L. Moeschberger (2005). *Survival Analysis: Techniques for Censored and Truncated Data*. New York: Springer.
- Kormos, Judit (2006). *Speech Production and Second Language Acquisition*. Mahwah, New Jersey.: Lawrence Erlbaum Associates.
- (2011). "Speech production and the cognition hypothesis." En: *Second Language Task Complexity: Researching the Cognition Hypothesis of Language Learning and Performance*. Ed. por P. Robinson. John Benjamins., págs. 39-60.
- Kot, Mark (2001). *Elements of mathematical ecology*. Cambridge University Press.
- Kotz, Sonja A. (2009). "A critical review of ERP and fMRI evidence on L2 syntactic processing". En: *Brain and Language* 109, págs. 68-74.
- Kuhn, Max y Kjell Johnson (2013). *Applied Predictive Modeling*. Springer.
- Lamberti, Pedro W. y col. (2004). "Intensive entropic non-triviality measure". En: *Physica A: Statistical Mechanics and its Applications* 334.1, págs. 119-131.
- Landauer, Thomas, Peter Foltz y Darrell Laham (1998). "Introduction to Latent Semantic Analysis". En: *Discourse Processes* 25, págs. 259-284.
- Lany, Jill A., Rebecca Gómez y Louann A. Gerken (2007). "The role of prior experience in language acquisition". En: *Cognitive Science* 31, págs. 481-508.
- Larsen Freeman, Diane (1997). "Chaos/complexity science and second language acquisition". En: *Applied Linguistics* 18.2, págs. 140-165.
- (2003). *Teaching language: From grammar to grammaring*. Thomson/Heinle.
- (2006). "The Emergence of Complexity, Fluency, and Accuracy in the Oral and Written Production of Five Chinese Learners of English". En: *Applied Linguistics* 27.4, págs. 590-619.
- (2011). "A Complexity Theory Approach to Second Language Development/Acquisition". En: *Alternative approaches to second language acquisition*. Ed. por Dwight Atkinson. New York: Routledge.
- (2012). "On the roles of repetition in language teaching and learning". En: *Applied Linguistics Review* 3.
- (2015). "Complexity Theory". En: *Theories in second language acquisition: an introduction*. Ed. por Bill Van Patten y Jessica Williams. second. Routledge. Cap. 12.
- Larsen freeman, Diane y Lynne Cameron (2008). *Complex Systems and Applied linguistics*. Oxford: Oxford University Press.
- Larsen Freeman, Diane y Nick Ellis (2009). "Constructing a second language: Analyses and computational simulations of the emergence of linguistic constructions from usage". En: *Language Learning* 59.1, págs. 90-125.
- Larson, J. Walter, Peter R. Briggs y Michael Tobis (2011). "Block-entropy analysis of climate data". En: *Procedia Computer Science* 4, págs. 1592-1601.
- Lenton, Timothy M. y Valerie N. Livina. (2016). "Detecting and anticipating climate tipping points." En: *Extreme Events: Observations, Modeling, and Economics*. AGU monograph 214. Ed. por M. Ghil & J. Urrutia-Fucugauchi M. Chavez. Wiley-Blackwell, págs. 51-62.
- Levelt, Willem J. M. (1989). *Speaking: From Intention to Articulation*. Cambridge, Massachusetts.: MIT Press.
- Levelt, Willem, Ardi Roelofs y Antje Meyer (1999). "A theory of lexical access in speech production". En: *Psycholinguistics: critical concepts in psychology*, 278-377 (2002) 22, págs. 1-75.

- Lew-Williams, Casey y Anne Fernald (2010). "Real-Time Processing of Gender-Marked Articles by Native and Non-native Spanish Speakers". En: *Journal of Memory and Language* 63, págs. 447-464.
- Liceras, Juana M., Lourdes Díaz y Caroline Mongeon (2000). "N-drop and Determiners in Native and Non-native Spanish: More on the Role of Morphology in the Acquisition of Syntactic Knowledge". En: *Current Research on the Acquisition of Spanish*. Ed. por Ronald P. Leow y Cristina Sanz. Somerville, MA: Cascadilla Press, págs. 67-96.
- Lichtman, Karen (2009). "Acquisition of Attributive and Predicative Adjective Agreement in L2 Spanish". En: *Proceedings of the 10th Generative Approaches to Second Language Acquisition Conference (GASLA 2009)*. Ed. por Melissa Bowles y col. Somerville, MA: Cascadilla Proceedings Project, págs. 231-247.
- Lin, D.Y. y L. J. Wei (1989). "The robust inference for the Cox proportional hazards model". En: *Journal of the American Statistical Association* 84, págs. 1074-1078.
- Littel, Ramon C., Jane Pendergast y Ranjini Natarajan (2000). "Modelling covariance structure in the analysis of repeated measures data". En: *Statistics in Medicine* 19, págs. 1793-1819.
- Livina, Valerie y col. (2011). "Changing climate states and stability: from Pliocene to present". En: *Climate Dynam.* 37, págs. 2437-2453.
- Lonardi, Stefano y col. (2002). "Finding motifs in time series". En: *Proc. of the 2nd Workshop on Temporal Data Mining*, págs. 53-68.
- Long, Michael y Peter Robinson (1998). "Focus on form: Theory, research and practice". En: *Focus on Form in Classroom Second Language Acquisition*. Ed. por Catherine Doughty y Jessica Williams. Cambridge: Cambridge University Press, págs. 15-41.
- Lorimor, Heidi y col. (2008). "Agreement and attraction in Russian". En: *Language and Cognitive Processes* 23, págs. 769-799.
- Lowie, Wander (2013). "Dynamic Systems Theory Approaches to Second Language Acquisition". En: *The Encyclopedia of Applied Linguistics*. Ed. por Carol A. Chapelle. Oxford: Wiley-Blackwell.
- Lowie, Wander, Rika Plat y Kees de Bot (2014). "Pink Noise in Language Production: A Nonlinear Approach to the Multilingual Lexicon". En: *Ecological Psychology* 26.3, págs. 216-228.
- Lowie, Wander y Marjolijn Verspoor (2015). "Variability and variation in second language acquisition orders: A dynamic re-evaluation". En: *Language Learning* 65, págs. 63-88.
- Lyster, R. y L. Ranta (1997). "Corrective feedback and learner uptake: Negotiation of form in communicative classrooms". En: *Studies in Second Language Acquisition* 19, págs. 37-66.
- Lyster, Roy y Kazuya Saito (2010). "Oral feedback in classroom SLA: A meta-analysis". En: *Studies in Second Language Acquisition* 32, págs. 265-302.
- MacWhinney, Brian (2010). *The CHILDES project: tools for analyzing talk. Part 1 & 2*. Electronic Edition.
- Machin, David, Yin Bun Cheung y Mahesh Parmar (2006). *Survival Analysis: A Practical Approach*. John Wiley & Sons.
- MacWhinney, Brian (2015). "Introduction". En: *The Handbook of Language Emergence*. Ed. por Brian MacWhinney y William O'Grady. John Wiley & Sons, Ltd, págs. 1-31.
- Manning, Christopher D., Prabhakar Raghavan e Hinrich Schütze (2008). *Introduction to information retrieval*. Cambridge university press.
- Manning, Christopher D., Mihai Surdeanu y col. (2014). "The Stanford CoreNLP Natural Language Processing Toolkit". En: *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics: System Demonstrations*, págs. 55-60.
- Marafioti, Pablo E. (2016). "Factores cognitivos en la producción oral de la L2: memoria, atención y complejidad en el enfoque por tareas." En: *Mente bilingüe. Abordajes psicolingüísticos y cognitivistas*. Ed. por Adolfo M. García y Sonia Suárez Cepeda. Comunicarte. Cap. 7, págs. 173-215.
- Marrero, Victoria y Carmen Aguirre (2003). "Plural acquisition and development in Spanish". En: *Linguistic theory and language development in Hispanic Languages*. Ed. por Silvia Montrul y Francisco Ordóñez. Cascadilla Press, págs. 275-296.
- Marwan, Norbert, M. Carmen Romano y col. (2007). "Recurrence plots for the analysis of complex systems". En: *Physics Reports* 438, págs. 237-329.

- Marwan, Norbert, Stefan Schinkel y Jürgen Kuhrt (2013). "Recurrence Plots 25 years later, gaining confidence in dynamical transitions". En: *Europhys. Lett.* 101.
- McCarthy, Corrine (2008). "Morphological Variability in the Comprehension of Agreement: An Argument for Representation over Computation". En: *Second Language Research* 24.4, págs. 459-486.
- McLaughlin, Judith y col. (2010). "Brain Potentials Reveal Discrete Stages of L2 Grammatical Learning". En: *Language Learning* 60.2, págs. 123-150.
- Menard, Scott (2011). "Standards for Standardized Logistic Regression Coefficients". En: *Social Forces* 89.4, págs. 1409-1428.
- Mitchell, Tom (1997). *Machine Learning*. McGraw Hill.
- Molenaar, Peter C. M. (2008). "On the implications of the classical ergodic theorems: Analysis of developmental processes has to focus on intra-individual variation". En: *Developmental Psychobiology* 50, págs. 60-69.
- Montrul, Silvina, Israel De la Fuente y col. (2013). "The role of experience in the acquisition and production of diminutives and gender in Spanish: Evidence from L2 learners and heritage speakers". En: *Second Language Research* 29, págs. 87-118.
- Montrul, Silvina, Rebecca Foote y Silvia Perpiñan (2008). "Gender Agreement in Adult Second Language Learners and Spanish Heritage Speakers: The Effects of Age and Context of Acquisition". En: *Language Learning* 58.3, págs. 503-553.
- Moore, Dirk F. (2016). *Applied Survival Analysis Using R*. Springer.
- Murray, James D. (2002). *Mathematical Biology I. An Introduction*. 3.^a ed. Interdisciplinary Applied Mathematics. New York: Springer.
- Nason, Guy Philip (2008). *Wavelet methods in statistics with R*. Springer.
- Nerbonne, John y col. (2013). "Measuring socially motivated pronunciation differences". En: *Approaches to Measuring Linguistic Differences*. Ed. por Lars Borin y Anju Saxena. De Gruyter Mouton, págs. 107-140.
- Newman, Mark E. J. (2010). *Networks: an introduction*. Oxford University Press.
- Newport, Elissa L. y Richard N. Aslin (2004). "Learning at a distance I. Statistical learning of non-adjacent dependencies". En: *Cognitive Psychology* 48, págs. 127-162.
- O'Grady, William (2005). *Syntactic Carpentry: an emergentist approach to syntax*. Mahwah, Nueva Jersey: Lawrence Erlbaum Associates.
- O'Grady, William, Miseon Lee y Miho Choo (2003). "A subject-object asymmetry in the acquisition of relative clauses in Korean as a second language". En: *Studies in Second Language Acquisition* 25, págs. 433-438.
- Oakes, Michael P. (1998). *Statistics for corpus linguistics*. Edinburgh UP.
- Ortega, Lourdes (2015). "Second Language Learning explained? SLA across 10 contemporary theories". En: *Theories in second language acquisition: an introduction*. Ed. por Bill Van Patten y Jessica Williams. second. Routledge. Cap. 13.
- Pacton, Sébastien y Pierre Perruchet (2008). "An attention-based associative account of adjacent and nonadjacent dependency learning". En: *Journal of Experimental Psychology: Learning, Memory, and Cognition* 34, págs. 80-96.
- Pallotti, Gabriele (2007). "An operational definition on the emergence criterion". En: *Applied Linguistics* 28.3, págs. 361-382.
- Paradis, Michel (2009). *Declarative and Procedural Determinants of Second Languages*. Amsterdam: John Benjamins.
- Pelucchi, Bruna, Jessica F. Hay y Jenny R. Saffran (2009). "Learning in reverse: Eight-month-old infants track backwards transitional probabilities". En: *Cognition* 113, págs. 244-247.
- Peña, Daniel (2002). *Análisis de datos multivariantes*. Mc Graw Hill.
- Percival, Donald B. y Andrew T. Walden (2008). *Wavelet methods for time series analysis*. Cambridge Univ. Press.
- Picallo, Carmen (2005). "Some notes on grammatical gender and I-pronouns". En: *Fachbereich Sprachwissenschaft, Arbeitspapier* 119, págs. 107-121.

- Picard, Franck (2007). *An introduction to mixture models. Technical report. Laboratoire Statistique et Genome. Univ. d'Evry.*
- Pienemann, Manfred (1998). *Language processing and second language development: Processability Theory.* Amsterdam: John Benjamins.
- (2005). *Cross-linguistic aspects of Processability Theory.* Amsterdam: John Benjamins.
- Pienemann, Manfred y Anke Lenzing (2015). “Processability Theory”. En: *Theories in second language acquisition: an introduction.* Ed. por Bill Van Patten y Jessica Williams. second. Routledge. Cap. 9.
- Pietsch, Lukas (2005). *Variable Grammars: Verbal Agreement in Northern Dialects of English.* Tübingen: Niemeyer.
- Pollard, Carl e Ivan Sag (1994). *Head-Driven Phrase Structure Grammar.* The University of Chicago Press.
- Prabasaj, Paul, Michael L. Pennell y Stanley Lemeshow (2013). “Standardizing the power of the Hosmer-Lemeshow goodness of fit test in large datasets”. En: *Statistics in Medicine* 32.1, págs. 67-80.
- Prévost, Philippe y Lydia White (2000). “Missing inflection or impairment in second language acquisition ? Evidence from tense and agreement”. En: *Second Language Research* 16.2, págs. 103-133.
- Putta, Paolo Della y Manuela Visigalli (2010). “Come aiutare l’acquisizione della morfosintassi del sintagma nominale italiano: confronto fra tre percorsi glottodidattici rivolti a studenti anglofoni”. En: *Italiano LinguaDue* 2.
- Putter, H., M. Fiocco y R. B. Geskus (2007). “Tutorial in biostatistics: competing risks and multi-state models.” En: *Statistics in Medicine* 26.11, págs. 2389-2430.
- R Development Core Team. (s.f.). *R: a language and environment for statistical computing.* Vienna, Austria. URL: <http://www.R-project.org>.
- Rendón, Eréndira y col. (2011). “Internal versus External cluster validation indexes”. En: *International Journal of Computers and Communications* 5.1, págs. 27-34.
- Reshef, David N. y col. (2011). “Detecting Novel Associations in Large Data Sets”. En: *Science* 334, págs. 1518-1524.
- Ribeiro, Haroldo V. y col. (2012). “Complexity-entropy causality plane: a useful approach for distinguishing songs”. En: *Physica A: Statistical Mechanics and its Applications* 391.7, págs. 2421-2428.
- Ripley, Ruth M. y col. (2018). *Manual for Siena version 4.0; R package version 1.2-12.* URL: <https://www.cran.r-project.org/web/packages/RSiena/>.
- Ritter, Elizabeth (1991). “Two Functional Categories in Noun Phrases: Evidence from Modern Hebrew”. En: *Syntax and semantics 25: Perspectives on Phrase Structure.* Ed. por Susan Rohstein. New York: Academic Press, págs. 37-62.
- (1992). “Cross-linguistic evidence for number phrase”. En: *Canadian Journal of Linguistics* 37, págs. 197-218.
- (1993). “Where’s Gender?” En: *Linguistic Inquiry* 24, págs. 795-803.
- Rokach, Lior y Oded Maimon (2008). *Data Mining with Decision Trees. Theory and Applications.* River Edge, NJ.: World Scientific Publishing Co.
- Romberg, Alexa R. y Jenny R. Saffran (2013). “All together now: concurrent learning of multiple structures in an artificial language”. En: *Cognitive Science* 37.7, págs. 1290-1320.
- Rothman, Jason y Diego Pascual y Cabo (2014). “Generative Approaches to Spanish Second Language Acquisition”. En: *The handbook of Spanish second language acquisition.* Ed. por Kimberly L. Geeslin. Malden, MA: Wiley-Blackwell, págs. 46-63.
- Rothman, Jason, Jorge Gonzalez Alonso y David Miller (2018). “The acquisition of second language Spanish morphosyntax”. En: *The Cambridge Handbook of Spanish Linguistics.* Ed. por Kimberly L. Geeslin. Cambridge: Cambridge University Press, págs. 689-715.
- Saffran, Jenny R., Richard N. Aslin y Elissa L. Newport (1996). “Statistical learning by 8-month-old infants”. En: *Science* 274, págs. 1926-1928.
- Sag, Ivan, Thomas Wasow y Emily Bender (2003). *Syntactic Theory: A Formal Introduction.* Second. University of Chicago Press.
- Sagarra, Nuria (2007). “Online processing of gender agreement in low proficient English-Spanish late bilinguals”. En: *Current issues in linguistic theory series.* Ed. por José Camacho y col. Amsterdam, The Netherlands: John Benjamins, págs. 240-253.

- Sagarra, Nuria y Julia Herschensohn (2010). "The Role of Proficiency and Working Memory in Gender and Number Agreement Processing in L1 and L2 Spanish". En: *Lingua* 120.8, págs. 2022-2039.
- (2013). "Processing of gender and number agreement in late Spanish bilinguals". En: *International Journal of Bilingualism* 17.5, págs. 607-627.
- Sandoval, Michelle y Rebecca L. Gómez (2013). "The development of nonadjacent dependency learning in natural and artificial languages". En: *Wiley Interdisciplinary Reviews: Cognitive Science* 4.5, págs. 511-522.
- Scheffer, Marten y col. (2009). "Early-warning signals for critical transitions". En: *Nature* 461.3, págs. 53-59.
- Schinkel, Stefan y col. (2009). "Confidence bounds of recurrence-based complexity measures". En: *Physics Letters A*. 373, págs. 2245-2250.
- Schmidt, Richard (2001). "Attention". En: *Cognition and Second Language Instruction*. Ed. por Peter Robinson. Nueva York: Cambridge University Press, págs. 3-32.
- Schmidt, Richard S. (2010). "Attention, awareness, and individual differences in language learning". En: *Proceedings of CLaSIC 2010*. Ed. por W. M. Chan y col. Centre for Language Studies. Singapur: National University of Singapore, págs. 721-737.
- Schoorlemmer, Erik (2009). "Agreement, Dominance and Doubling. The morphosyntax of DP". Universiteit Leiden.
- Scrucca, Lucca y col. (2016). "mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models". En: *the R Journal* 8.1, págs. 205-233.
- Seekell, David A., Stephen R. Carpenter y Michael L. Pace. (2011). "Conditional Heteroscedasticity as a Leading Indicator of Ecological Regime Shifts". En: *The American Naturalist* 178.4, págs. 442-451.
- Smith, Garrett, Julie Franck y Whitney Tabor (2018). "A Self-Organizing Approach to Subject-Verb Number Agreement". En: *Cognitive Science* 42 Suppl 4.
- Smolensky, Paul (2006). "Harmony in Linguistic Cognition". En: *Cognitive science* 30, págs. 779-801.
- Smolensky, Paul, Matthew Goldrick y Donald Mathis (2014). "Optimization and quantization in gradient symbol systems: a framework for integrating the continuous and the discrete in cognition". En: *Cognitive science* 38.6, págs. 1102-1138.
- Snijders, Tom A. B. (1996). "Stochastic actor-oriented models for network change". En: *Journal of Mathematical Sociology* 21, págs. 149-172.
- Snijders, Tom A. B., Gerhard G. van de Bunt y Christian E. G. Steglich (2010). "Introduction to stochastic actor-based models for network dynamics". En: *Social Networks* 32.1, págs. 44-60.
- Snijders, Tom A. B., Christian E. G. Steglich y Michael Schweinberger (2007). "Modeling the co-evolution of networks and behavior". En: *Longitudinal models in the behavioral and related sciences*. Ed. por Kees van Montfort, Johan Oud y Albert Satorra. Mahwah, NJ.: Lawrence Erlbaum, págs. 41-71.
- Sorace, Antonella (2011). "Pinning Down the Concept of 'Interface' in Bilingualism". En: *Linguistic Approaches to Bilingualism* 1, págs. 1-33.
- Spivey, Michael (2007). *The continuity of mind*. Oxford: Oxford University Press.
- Spoelman, Marianne y Marjolijn Verspoor (2010). "Dynamic patterns in development of accuracy and complexity: a longitudinal case study in the acquisition of Finnish". En: *Applied Linguistics* 31, págs. 532-553.
- Sprott, Julien C. (2003). *Chaos and Time series analysis*. Oxford: Oxford University Press.
- Strogatz, Steven H. (1994). *Nonlinear dynamics and chaos: with applications to physics, biology, chemistry, and engineering*. Perseus Books Publishing, L.L.C.
- Stroup, Walter W. (2013). *Generalized Linear Mixed Models: Modern Concepts, Methods and Applications*. CRC Press, Chapman Hall Book.
- Swain, Merrill (1985). "Communicative competence: Some roles of comprehensible input and comprehensible output in its development". En: *Input in second language acquisition*. Ed. por Susan M. Gass y Carolyn G. Madden. MA: Newbury: Rowley, págs. 235-253.
- (2005). "The Output Hypothesis: Theory and research". En: *Handbook on research in second language learning and teaching*. Ed. por E. Hinkel. Mahwah, NJ: Lawrence Erlbaum, págs. 471-483.
- (2008). "The output hypothesis: its history and its future". En: *Foreign Language Teaching And Research* 40.1, págs. 45-50.

- Taatgen, Niels A. y John R. Anderson (2008). "Constraints in Cognitive Architectures". En: *The Cambridge Handbook of Computational Psychology*. Ed. por Ron Sun. Cambridge: Cambridge University Press, págs. 170-186.
- Tableman, Mara y Jong Sung Kim (2005). *Survival analysis using S: analysis of time-to-event data*. Chapman & Hall/CRC.
- Tan, Pang-Ning, Michael Steinbach y Vipin Kumar (2005). *Introduction to Data Mining*. Pearson/Addison-Wesley.
- Therneau, Terry M. y Elizabeth J. Atkinson (2015). *An Introduction to Recursive Partitioning: Using the RPART Routines*. Inf. téc. Mayo Foundation.
- Tibshirani, Robert (1996). "Regression Shrinkage and Selection via the Lasso". En: *Journal of the Royal Statistical Society. Series B (Methodological)* 58.1, págs. 267-288.
- Tokowicz, Natasha y Brian MacWhinney (2005). "Implicit and Explicit Measures of Sensitivity to Violations in Second Language Grammar". En: *Studies in Second Language Acquisition* 27, págs. 173-204.
- Tordera Yllescas, Juan Carlos (2008). *Introducción a la Gramática Léxico-Funcional*. Publicacions de La Universitat de Valencia.
- Ullman, Michael T. (2005). "A cognitive neuroscience perspective on second language acquisition: The declarative/procedural model". En: *Mind and Context in Adult Second Language Acquisition: Methods, Theory, and Practice*. Ed. por Cristina Sanz. Washington, DC: Georgetown University Press, págs. 141-178.
- Van Buuren, Stef (2013). *Flexible Imputation of Missing Data*. Boca Raton: CRC Press.
- Van Buuren, Stef y Karin Groothuis-Oudshoorn (2011). "mice: Multivariate imputation by chained equations in R". En: *Journal of Statistical Software* 45.3, págs. 1-67.
- Van den Bos, Esther, Morten H. Christiansen y Jennifer B. Misyak (2012). "Statistical Learning of Probabilistic Nonadjacent Dependencies by Multiple-cue Integration". En: *Journal of Memory and Language* 67, págs. 507-520.
- Van Geert, Paul (1994). *Dynamic Systems of development. Change between complexity and chaos*. London: Harvester Wheatsheaf.
- (1995). "Growth dynamics in development". En: *Mind as motion: Explorations in the dynamics of cognition*. Ed. por Robert F. Port and Timothy J. Van Gelder. The MIT Press, págs. 313-338.
- (2003). "Dynamic systems approaches and modeling of developmental processes". En: *Handbook of developmental psychology*. Ed. por Jaan Valsiner y Kevin J. Conolly. Sage, págs. 640-672.
- (2014). "Group versus individual data in a dynamic systems approach to development". En: *Enfance* 3, págs. 283-312.
- Van Heugten, Marieke y Elizabeth K. Johnson (2010). "Linking infants' distributional learning abilities to natural language acquisition". En: *Journal of Memory and Language* 63, págs. 197-209.
- Van Patten, Bill (2015). "Input Processing in adult SLA". En: *Theories in second language acquisition: an introduction*. Ed. por Bill Van Patten y Jessica Williams. second. Routledge. Cap. 7.
- Van Patten, Bill y Jessica Williams (2015). *Theories in second language acquisition: an introduction*. second. Routledge.
- Venables, W. N. y B. D. Ripley (2002). *Modern Applied Statistics with S*. New York: Springer.
- Verspoor, Marjolijn, Wander Lowie y Marijn van Dijk (2008). "Variability in second language development from a dynamic systems perspective". En: *The Modern Language Journal* 92, págs. 214-231.
- Villemereuil, Pierre (2012). *Estimation of a biological trait heritability using the animal model. How to use the MCMCglmm R package*. URL: <http://devillemereuil.legtux.org/downloads/>.
- Villemereuil, Pierre, Olivier Gimenez y Blandine Doligez (2013). "Comparing parent-offspring regression with frequentist and Bayesian animal models to estimate heritability in wild populations: a simulation study for Gaussian and binary traits". En: *Methods in Ecology and Evolution* 4, págs. 260-275.
- Wang, Rong y col. (2012). "Flickering gives early warning signals of a critical transition to a eutrophic lake state". En: *Nature* 492, págs. 419-422.
- Webber, Charles L. (Jr.), Cornel Ioana y Norbert Marwan (Eds.) (2016). *Recurrence Plots and Their Quantifications: Expanding Horizons*. Springer.
- Webber, Charles L. (Jr.) y Norbert Marwan (2015). *Recurrence Quantification Analysis*. Springer.

- Webber, Charles L. (Jr.) y Joseph P. Zbilut (2005). "Recurrence quantification analysis of non linear dynamical systems". En: *Tutorials in contemporary non linear methods for the behavioral sciences*. Ed. por M. A. Riley y G. C. Van Orden. Cap. 2, págs. 26-94.
- Weibull, Jörgen W. (1997). *Evolutionary game theory*. MIT Press.
- White, Lydia (1989). *Universal Grammar and Second language acquisition*. Amsterdam: John Benjamins.
- (2015). "Linguistic theory, Universal Grammar and Second Language Acquisition". En: *Theories in second language acquisition: an introduction*. Ed. por Bill Van Patten y Jessica Williams. second. Routledge. Cap. 3.
- White, Lydia y col. (2004). "Gender and Number Agreement in Nonnative Spanish". En: *Applied Psycholinguistics* 25, págs. 105-133.
- Yanovich, Igor (2017). "Analyzing imperfective games". En: *Semantics and Pragmatics* 10.
- Zanin, Massimiliano y col. (2012). "Permutation Entropy and Its Main Biomedical and Econophysics Applications: A Review". En: *Entropy* 14.8, págs. 1553-1577.