

# Factores influyentes en tipos de errores de concordancia nominal en un corpus de cuatro aprendientes italianos de español LE

Pablo Ezequiel Marafioti   
Universidad Nacional de Córdoba  
Córdoba, Argentina

## Resumen

Se consideran diferentes factores que intervienen en la producción de concordancia nominal de número en un corpus de cuatro aprendientes italianos de español LE, en un estudio de caso longitudinal. Se utilizó un modelo lineal generalizado mixto multinomial bayesiano. Los errores se categorizaron como: (i) de género [*las barrios*], (ii) de no tener en cuenta la inserción de *-e-* epentética [*los trenos*], (iii) de plural [*las casa*], (iv) mixtos [*les joven*]. Se hallaron efectos para los siguientes factores: (i) tipo de modificador; (ii) sustantivos animados; (iii) tres tipos de estrategias de aprendizaje; (iv) concordancias con palabras cuyas desinencias tenían similitud media o baja con las del italiano; (v) con sustantivos familiares y/o frecuentes; (vi) instancias cuyos TYPES fueran más frecuentes; (vii) concordancias que contenían ambos términos con *-e-* epentética; (viii) errores acumulados hasta la instancia corriente.

**Palabras clave:** segunda lengua; adquisición; concordancia nominal; análisis bayesiano.

## Abstract

**Factors that influence grammatical agreement errors in a corpus of four Italian learners of Spanish as a Foreign Language**

This study explored different factors that are thought to influence grammatical agreement in a corpus of four Italian learners of Spanish as a Foreign Language (SFL). The study adopted a longitudinal case study design. A generalized linear mixed multinomial Bayesian model was used. The errors were classified as: (i) errors in

agreement based on gender [*las barrios*], (ii) errors derived from not taking into account an insertion of -e- in words ending in consonant [*los trenos*], (iii) errors in agreement based on number [*las casa*] , (iv) mixed errors [*les joven*]. The following factors were found statistically significant: (i) type of modifier; (ii) animate nouns; (iii) three types of learning strategies; (iv) words endings similarity with those with those of the Italian words; (v) noun frequency and/or familiarity; (vi) instances whose TYPES were more frequent; (vii) instances in which both terms had plurals in “-es”; (vii) cumulated errors until the current instance.

**Key words:** second language; acquisition; nominal agreement; Bayesian analysis.

### Résumé

#### **Facteurs influents sur les types d'erreurs d'accord nominal dans un corpus de quatre étudiants italiens apprenant l'espagnol langue étrangère**

Différents facteurs impliqués dans la production de la concordance nominal plurielle pour un corpus de quatre étudiants italiens apprenant l'espagnol (LE) sont considérés dans une étude de cas longitudinale. Un modèle bayésien multinomial linéaire généralisé et mixte a été utilisé. Les erreurs ont été classées comme suit : (i) genre [*las barrios*], (ii) sans tenir compte de l'insertion d'un -e- épenthétique [*los trenos*], (iii) pluriel [*las casa*], (iv) mélange [*les joven*]. Des effets ont été constatés pour les facteurs suivants : (i) type de modificateur ; (ii) noms animés ; (iii) trois types de stratégies d'apprentissage ; (iv) concordances avec des mots dont les désinences avaient une similarité moyenne ou faible avec celles de l'italien ; (v) avec des noms familiers et/ou fréquents ; (vi) cas dont les TYPES étaient plus fréquents ; (vii) concordances contenant les deux termes avec -e- épenthétique ; (vii) erreurs accumulées jusqu'au cas présent.

**Mots-clés:** langue seconde ; acquisition ; accord nominal ; analyse bayésienne.

## **SOBRE EL AUTOR**

### **Pablo Ezequiel Marafioti**

Doctor en Ciencias del Lenguaje por la Universidad Nacional de Córdoba (Argentina), Especialista en Lingüística Teórica y Aplicada por la Universidad de Pavía (Italia), Licenciado en Letras (orientación en Lingüística) por la Universidad de Buenos Aires (Argentina). Realizó varios posgrados en Estadística. Áreas de especialización: sistemas dinámicos, modelos estadísticos y de minería de datos aplicados al cambio lingüístico y a la adquisición de lenguas extranjeras.

Correo electrónico: kenavo38@yahoo.com

## **CÓMO CITAR ESTE ARTÍCULO**

Marafioti, P. (2021). Factores influyentes en tipos de errores de concordancia nominal en un corpus de cuatro aprendientes italianos de español LE. *Lenguaje*, 49(2), 365-393.  
<https://doi.org/10.25100/lenguaje.v49i2.10915>.

## INTRODUCCIÓN

El objetivo general de este trabajo consiste en identificar algunos factores influyentes en la producción de concordancia plural en cuatro aprendientes italianos de español LE. Se trata de un estudio de caso longitudinal y observacional. Se buscará relacionar dichos factores con cuatro tipos de errores de concordancia plural, a saber: (i) de género [*las barrios*], (ii) de no tener en cuenta la inserción de *-e-* epentética [*los trenos*], (iii) de plural [*las casa*], (iv) mixtos [*les joven*]. Se empleará un modelo lineal generalizado multinomial mixto bayesiano, aplicado por primera vez a ELE. Además, dicha técnica permitirá establecer trayectorias de aprendizaje para cada aprendiente, con el objetivo de explorar la adquisición de la concordancia plural de modo dinámico.

La concordancia se define 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 aquél. Se denomina *dominio* al entorno sintáctico en el cual ocurre la concordancia. La concordancia se establece por covarianza sistemática de rasgos. En el presente trabajo el controlador será nominal. Por otra parte, los objetivos consistirán en artículos (definidos e indefinidos) y adjetivos (demostrativos, posesivos, indefinidos). Concuerdan con un solo controlador obligatoriamente. Los dominios relevantes serán el sintagma nominal, el sintagma verbal (predicativo) y la oración subordinada. La concordancia se considerará asimétrica (el género y número de los objetivos dependen del controlador nominal).

En general, la investigación sobre concordancia en el entorno nominal se ha enfocado más en el género que en el número; y mayormente en procesamiento que en producción. Se han estudiado efectos de género, animicidad y morfología del controlador; congruencia en animicidad y género entre controlador y objetivo; dificultad del artículo respecto del adjetivo; distancia entre controlador y objetivo. En este trabajo se amplía la evidencia para el área menos estudiada de la producción de número plural. Además, se incluyen factores no considerados hasta el momento, entre ellos: influencia de diferentes tipos de modificadores; frecuencia de tipos de concordancia y de errores cometidos; otras características del controlador (además de la animicidad); posibles estrategias de aprendizaje; dificultad con *-e-* epentética; concordancias con más de un objetivo.

La literatura de adquisición ha arrojado los siguientes resultados. En primera instancia, 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 parece ser 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 (Franceschina, 2001; Muñoz *et al.*, 2000; White *et al.*, 2004).

La concordancia de género y número del artículo resulta más fácil de adquirir que la del adjetivo. Esto parece ser así para cualquier nivel de competencia; en bilingües tempranos y tardíos; y tanto en producción como en procesamiento (Alarcón, 2011; Dowens *et al.*, 2010; Fernández-García, 1999; Montrul *et al.*, 2008; White *et al.*, 2004). Por otra parte, la concordancia plural del cuantificador *mucho* parece más fácil de adquirir que la de *bastante* y *demasiado* (Español-Echevarría y Prévost, 2004).

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* (Alarcón, 2011; Bruhn de Garavito y White, 2002; Fernández-García, 1999; McCarthy, 2008; Montrul *et al.*, 2008; White *et al.*, 2004); (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*), estos últimos en el orden de facilidad *consonante* > -e > *opuestos* (Alarcón, 2011; Fernández-García, 1999; Montrul *et al.*, 2008); (c) con controladores inanimados (como en *hospital*) respecto de los animados en correspondencia con el sexo biológico, como en *doctor/a* –Sagarra y 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).

Asimismo, en un estudio de González *et al.* (2019) se analizaron errores de concordancia de género y número en el ámbito nominal en composiciones escritas de 23 estudiantes holandeses de español LE. Encontraron efectos significativos de aumento de errores: (i) del plural respecto del singular; (ii) del femenino respecto del masculino; (iii) en los artículos femeninos (sin importar el rasgo de número).

El aumento de distancia estructural (cantidad de nodos sintácticos entre controlador y objetivo) causa que disminuya la sensibilidad a las violaciones de género y que la concordancia se procese más lentamente en el dominio no local (Dowens *et al.*, 2010; Foote, 2011; Keating, 2009, 2010; Lichtman, 2009; Sagarra, 2007). 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; aunque en Foote (2011) no hubo evidencia de ello.

El material suplementario y el código de R empleado se encuentra en: [https://github.com/pablomarafioti/PabloMarafioti/tree/master/analisis\\_bayesiano](https://github.com/pablomarafioti/PabloMarafioti/tree/master/analisis_bayesiano).

## METODOLOGÍA

La Tabla 1 muestra los pasos que se siguieron para la recolección, modelización y el análisis de los datos. Luego de la creación del corpus de producción oral, se pusieron marcadores según cuatro tipos de errores de concordancia (o falta de error). Estos constituyeron la variable respuesta, correspondiente a cada instancia de concordancia. Las concordancias se hallaban agrupadas por sujeto / alumno y por sesión. A continuación, se estableció una serie de covariables predictoras que describían diferentes características de las instancias. Despues se realizaron operaciones de pre-procesamiento, lo cual derivó en la base de datos de análisis. El primer paso de modelización consistió en la selección de variables predictoras más influyentes. Las variables elegidas se usaron en un modelo multinomial bayesiano, decidiendo también la estructura aleatoria más apropiada. En la etapa de análisis de datos se examinaron: (1) los efectos fijos para determinar los factores significativos en el aumento o disminución de la chance de cometer un determinado tipo de error; (2) la trayectoria de los efectos aleatorios para cada tipo de error con el objetivo de evaluar la dinámica del aprendizaje. También se examinaron las instancias del corpus para los efectos fijos hallados y los TYPES de concordancia.

**Tabla 1.** Metodología

370

ETAPA	PASOS
<b>(I) Base de datos</b>	CORPUS ORAL → TAGGING → INSTANCIAS (RESPUESTA) → VARIABLES PREDICTORAS → PRE-PROCESAMIENTO
<b>(II) Modelo bayesiano</b>	SELECCIÓN DE VARIABLES → MODELO BAYESIANO INICIAL → SELECCIÓN DE ESTRUCTURA ALEATORIA → MODELO FINAL
<b>(III) Análisis de datos</b>	EFFECTOS FIJOS, TRAYECTORIA DE EFFECTOS ALEATORIOS, INSTANCIAS DEL CORPUS PARA EFFECTOS FIJOS Y TYPES

### Corpus y creación de variables

Se analizan datos de cuatro casos de estudiantes de español como lengua extranjera. Se trató de cuatro adultos, de lengua nativa italiana, estudiantes del Instituto *Cervantes* de Milán en el año académico 2008/09. Cada alumno poseía un nivel distinto de competencia lingüística (según el Marco Común Europeo de Referencia). Se hicieron entrevistas de 30 minutos entre el alumno y el investigador (autor de este trabajo). 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. Hubo entre doce y catorce entrevistas por alumno. 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 Tabla 2 muestra el perfil de cada alumno.

**Tabla 2.** Perfil de los sujetos

Alumno	Nivel	Profesión	L1	Horas de español previas	Otras L2 estudiadas
SONIA	A1 / A2	Investigadora	Italiano	0 horas	alemán –inglés
NATI	B1	Empleada	Italiano	120 horas	Francés – inglés
JAKO	B2	Estudiante	Italiano	240 horas	Inglés
MIRKA	C1	Programadora	Italiano	360 horas	Japonés – inglés

Siguiendo a MacWhinney (2000), la codificación / transcripción de los datos se hizo mediante el formato CHAT, y el conteo mediante el programa CLAN. Cada concordancia se codificó con dos términos pero pudiendo haber más términos “objetivo”: por ejemplo, en *los libros azules* se codificaron dos instancias: *los libros* y *libros azules*. Se anotaron a continuación marcadores [‘tags’] en el *corpus* para realizar el conteo posterior. Son los siguientes: (i) [\*0] = ausencia de error; (ii) [\*1] = errores en el género; (iii) [\*2] = errores debidos al uso de la terminación “(-e)-s”: (a) por no tomar en cuenta la última consonante de la raíz léxica, que exige un plural con “e” epentética en -(e)-s; (b) por uso en contexto incorrecto u omisión en correcto; (iv) [\*3] = errores de plural, o sea ausencia de -s; (v) [\*4] = errores mixtos por acumulación de los anteriores. Se creó la variable respuesta “RES\_CAT” con niveles idénticos (0,1,2,3,4) a los “tags” descriptos, estableciendo “0” como referencia. Por ejemplo:

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

148 \*STU: después les@s:ita ehh@fp después las doce hay muchos trenos [\*2]  
→ error de -e- epentética: muchos trenes. [SONIA, sesión 7]

293 \*STU: si pero los veneciano [\*3] conocen donde ir por comprar mejor.  
→ error de plural: los venecianos. [SONIA, sesión 5]

144 \*STU: por les joven [\*4]. → error por acumulación: para los jóvenes. [SONIA, sesión 2]

Se crearon variables que caracterizaban cada instancia producida de concordancia. Se las describe a continuación (el primer nivel se considera el de referencia).

- *ESP.* Concordancia en español (sin error).

- *MOD*. Tipo de modificador del controlador. Niveles: 0 = artículo definido; 1 = artículo indefinido; 2 = determinante (adjetivos posesivos, indefinidos, demostrativos, interrogativos, exclamativos); 3 = adjetivos (calificativos, numerales, ordinales).
- *GRAM*. Si se trataba de una instancia de concordancia de más de dos términos. Niveles: 0 = dos términos; 1 = más de dos términos.
- *LDA*. Si la concordancia era o no a larga distancia: 0 = no, 1 = sí.
- *POS*. En el caso de *LDA* = 1, si la concordancia se da o no en el contexto de una relativa (distancia estructural), con los niveles: 0 = sin larga distancia; 1 = con larga distancia, sin subordinación [*los animales son muy sensibles*]; 2 = con larga distancia, con subordinación [*hay asignaturas que se consideran inútiles*].
- *DIS*. En el caso de *LDA* = 1, la distancia lineal entre controlador y objetivo; con los niveles: 0 = sin larga distancia; 1 = hasta 3 palabras (corta distancia); 2 = desde 4 hasta 9 palabras (larga distancia); 3 = distancia al controlador en enunciado fuera de donde se encuentra el objetivo (distancia a enunciados).
- *ES*. Se especificó si en el controlador, en el objetivo, o en ambos, había una desinencia que requería 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 los niveles: 0 = sin “e” epentética; 1 = con “e” epentética en un término; 2 = con “e” epentética en ambos términos.
- *ANIM*. Si el controlador era o no animado. Es decir, si la entidad a la que se refiere el nombre puede moverse o no por propia voluntad, según: 0 = inanimado, 1 = animado.
- *Fabs\_C y Fabs\_C*. La frecuencia del TYPE de concordancia. Cada TYPE especificaba el contexto de la concordancia. 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 el tipo de modificador, en el orden en que aparecían en la instancia. Después se indicó la terminación de cada término. En el caso de que se tratara de larga distancia, se especificó el lema del verbo o el pronombre relativo; también alguna estructura que implicara interferencia para computar la concordancia. En total se crearon 104 TYPES. Por ejemplo, la instancia *romanos alegres* en el contexto [*los romanos son muy alegres*] se codificó como: [L-n-<SER>-j-os-es]. Se trata de una concordancia a larga distancia marcada por “L”. 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]. Dichas frecuencias fueron calculadas a partir del corpus de datos propio [Fabs\_C] y de un corpus del español electrónico *online* [Fabs\_S]. Para esto último, se apeló al corpus del español *EsTenTen* de *Sketch Engine* (Kilgarriff *et al.*, 2014).

- *CUMRES*. Cantidad de errores cometidos 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.
- *TIME*. Índice de tiempo para cada instancia dentro de la sesión, reiniciándose (en *uno*) al inicio de cada sesión.

Los siguientes atributos, sobre rasgos del controlador, se extrajeron de la base de datos “BuscaPalabras” (Davis y 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.

373

A modo de ilustración de los atributos descriptos hasta ahora, considérese el siguiente fragmento de transcripción (MIRKA, sesión 6):

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 sé 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 tenemo [\*] 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] .

Se registraron las variables como se muestra en la Tabla 3. Solamente hay un error de plural en la primera instancia [*los derechos*]. Salvo la última instancia [*animales sensibles*], no hay concordancias a larga distancia. En dicha instancia POS = 1, DIS = 1.

**Tabla 3.** Ejemplo ilustrativo de registro de atributos

INSTANCIA	ESP	MOD	LDA	ES	GRAM	ANIM	FAM	IMA	CONC	LEX
										ESP
lo derechos (1)	los derechos	0	0	0	1	0	6,17	3,71	3,62	130
mismos derechos (2)	mismos derechos	2	0	0	1	0	6,17	3,71	3,62	130
las personas	las personas	0	0	0	0	1	7	6,22	5,49	171,79
los seres (1)	los seres	0	0	1	1	1	5,29	4,23	2,37	82,5
seres humanos (2)	seres humanos	3	0	1	1	1	5,29	4,23	2,37	82,5
los animales	los animales	0	0	1	0	1	6,63	6,31	3,54	73,04
[animales] <muy sensibles	animales sensibles	3	1	1	0	1	6,63	6,31	3,54	73,04

Además, se crearon dos variables basadas en la distancia de *Levenstein* (Nerbonne *et al.*, 2013; Oakes, 1998), con el objetivo de medir la similitud entre las raíces léxicas entre el español y el italiano, y entre los alomorfos 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). Se asignaron los siguientes pesos para las operaciones: (i) DELETE = 0.3; (ii) SUBSTITUTE = 0.6; INSERT = 1.

374

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). Dichas estrategias buscaron identificar casos que facilitaran o dificultaran la producción de concordancias. Fueron:

- *Estrategia 1 (EST1)*: si la palabra plural del italiano termina en *-i* poner en español plural en *-os*.
- *Estrategia 2 (EST2)*: si la palabra plural del italiano termina en *-e* poner en español plural en *-as*.
- *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.
- *Estrategia 4 (EST4)*: si la palabra plural del italiano termina en *-e*, poner en español el plural en *-es*. Por ejemplo: *vacanze* > *vacaciones*; *strade* > *calles*; *volte* > *veces*.
- *Estrategia 5 (EST5)*: si la palabra singular del italiano termina en *-e*, poner en español el plural en *-es*. Por ejemplo, la palabra *sole* ['sol'] podría ser la base para formar el

plural español agregando “s”: *sole* > *soles*; y el singular también, sacando “s”: *sole* > *sol*; *istituzione* > *instituciones*. Es decir, casos en los cuales el español coincide con la aplicación del plural con -e- epentética.

- *Estrategia 6 (EST6)*: si la palabra *singular* del italiano termina en -e, poner en español el plural en -es. Por ejemplo, la palabra *grande* ['grande'] o *studente* ['estudiante'] podrían formar plural (y singular) a partir de una base singular en italiano: *grandes*, *estudiantes*. Otros casos: *fonte* > *fuentes*; *abitudine* > *costumbres*; *dolce* > *dulces*. Son casos que no coinciden con -e- epentética.
- *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.

La Tabla 4 ejemplifica los casos en italiano, español y la instancia efectivamente producida por el alumno.

**Tabla 4. Ejemplos de Estrategias**

ITALIANO PLURAL	ITALIANO SINGULAR	ESPAÑOL	INSTANCIA	1	2	3	4	5	6	7
<b>molte volte</b>	(molta) volta	muchas veces	muchas vesas	0	1	0	1	0	0	0
<b>vacanze</b>	vacanza	vacaciones	vacacione	0	0	0	1	0	0	0
<b>gradevoli</b>	gradevole	agradables	agreables							
<b>uniche moto</b>	unica moto	únicas motos	unicas moto	0	1	1	0	0	0	0
<b>molti cinema</b>	(molto) cinena	muchos cines	muchos cines	1	0	1	0	0	0	0
<b>responsabilitá sociale</b>	responsabilitá sociali	responsabilidad es sociales	responsabilidad es sociales	0	0	0	0	1	0	1
<b>le abitudini</b>	l'abitudine	las costumbres	los costumbre	0	0	0	0	0	1	0

375

Se realizaron las siguientes operaciones de pre-procesamiento. En primer lugar, 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: (i) Frecuencia de controlador (Corpus “BuscaPalabras”):  $LEXESP = \log(LEXESP + 1)$ ; (ii) Frecuencia absoluta de TYPE en corpus propio:  $Fabs_C = \log(Fabs_C + 1)$ ; (iii) Frecuencia absoluta de TYPE en corpus *EsTenTen*:  $Fabs_S = \log(Fabs_S + 1)$ ; (iv)  $CUMRES = \log(CUMRES + 1)$ .

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. Se utilizó el paquete *mice* [*Multivariate Imputation by Chained Equations*] de R (Van Buuren y Groothuis-Oudshoorn, 2011), que hace imputación múltiple.

Se aplicó un Análisis de Componentes Principales (PCA, por sus siglas en inglés) para solucionar la correlación hallada entre las variables Imaginabilidad (IMA) y Concretud (CONC), por un lado, y entre Familiaridad (FAM) y Frecuencia (LEXESP), por otro. También resultó alta la correlación entre las frecuencias de TYPE (pero sin estar correlacionadas con las anteriores). La técnica permite obtener nuevas variables ortogonales llamadas componentes principales, que se calculan como combinación lineal de las variables cuantitativas originales (Peña, 2002). En el caso de las variables del controlador, en la primera componente (se llamará “IMA.CONC”) cargaban las variables Imaginabilidad y Concretud; y en la segunda (se llamará “FAM.LEX”), Familiaridad y LEXESP. En lo que respecta a las variables de frecuencia de TYPE, en la única componente cargaban con fuerza Fabs\_S y Fabs\_C.

Se decidió discretizar los atributos cuantitativos utilizando *clustering* por mezcla de gausianas. Se utilizó el paquete *mclust* de R (Scrucca *et al.*, 2016) para hacer el agrupamiento. Se discretizaron los atributos: MORF, STEM, IMA.CONC, FAM.LEX, Fabs.CS, CUMRES. El atributo FAM.LEX se discretizó en dos categorías poniendo como punto de corte a la mediana; ya que el *clustering* no resultó efectivo.

Por último, se llevó a cabo un 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. Dicha variable definió seis grupos. Los grupos se caracterizan como sigue: (1) distancias MORF.f medias y altas; y un 25 % de instancias con -e- epentética; (2) distancias MORF.f altas y solamente casos de -e- epentética en un solo término. Por otro lado, consiste exclusivamente de distancias altas de raíz STEM.f; (3) en su mayoría más de dos términos de concordancia y por tener distancias MORF.f medias, STEM.f baja y sin -e- epentética; (4) distancias medias MORF.f en su mayoría; instancias sin -e- epentética; (5) 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]; (6) distancias altas MORF.f en su mayoría; instancias con -e- epentética en un término. A modo de resumen, las Tablas 5 y 6 muestran las variables creadas.

**Tabla 5.** Discretización de atributos utilizando clustering por mixtura de gausianas

Atributo	Descripción	Discretización	Casos	Ejemplos del Corpus
MORF.f	Similitud 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	Similitud 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)
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)
		2 = [1.94; 3.33]	639	“los estudiantes” (2.63); “las ciudades” (2.89)
CUMRES.f	Errores acumulados	0 = [0; 0.69)	509	“los idiomas” (0); “los profesores” (0.69)
		1 = [0.69; 1.94)	709	“muchos lugares” (1.38); “las bromas” (1.6)
		2 = [1.94; 3.33]	639	“los estudiantes” (2.63); “las ciudades” (2.89)

**Tabla 6.** Variables creadas

Variable	Descripción	Clase	Niveles
ALUMNO	Alumno	Cualitativa	-
SESIÓN	Sesión transcripta	Cualitativa	-
LINEA	Línea en la transcripción .CHA	Cuantitativa	[6, 515]
INSTANCIA	Concordancia observada	Caracteres	-
MOD	Tipo de modificador del controlador	Cualitativa	0 = art. definido; 1 = art. indefinido; 2 = determinante; 3 = adjetivo
LDA	Concordancia a larga distancia	Cualitativa	0 = sin larga distancia; 1 = con larga distancia
POS	LDA: distancia estructural	Cualitativa	0 = sin larga distancia; 1 = sin subordinada; 2 = con subordinada
DIS	LDA: distancia lineal	Cualitativa	0 = sin; 1 = corta; 2 = larga; 3 = distancia a enunciados.
ES	Presencia de -e- epentética	Cualitativa	0 = sin; 1 = en un término; 2 = en ambos términos
GRAM	Concordancia de 2 términos o más	Cualitativa	0 = dos términos; 1 = más de dos términos
CONC	Concretud del controlador	Cuantitativa	Escala de 1 a 7
FAM	Familiaridad del controlador	Cuantitativa	Escala de 1 a 7
IMA	Imaginabilidad del controlador	Cuantitativa	Escala de 1 a 7
LEXESP	Log(Frecuencia del controlador + 1)	Cuantitativa	[2.32, 744.6]
378	ANIM	Cuantitativa	0 = inanimado; 1 = animado
	ESP	Caracteres	-
	MORF	Cuantitativa	[2.2, 3.6]
	STEM	Cuantitativa	[1.8; 10.2]
	EST1	Cualitativa	0 = no aplica; 1 = aplica
	EST2	Cualitativa	0 = no aplica; 1 = aplica
	EST3	Cualitativa	0 = no aplica; 1 = aplica
	EST4	Cualitativa	0 = no aplica; 1 = aplica
	EST5	Cualitativa	0 = no aplica; 1 = aplica
	EST6	Cualitativa	0 = no aplica; 1 = aplica
	EST7	Cualitativa	0 = no aplica; 1 = aplica
	CUMRES	Cuantitativa	[1, 3.33]
	Fabs_C	Cuantitativa	[0.69, 5.54]
	Fabs_S	Cuantitativa	[0.69, 16.05]
	EsTenTen		
	GRUPO6	Cualitativa	Grupos 1 a 6
	TIME	Cuantitativa	1, 2, 3, ..., $N_{sesión}$

## Expectativa de 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 espera que exista más chance de error con *tipos* de errores con concordancias que contengan: (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; (iv) concordancias de más de dos términos; (v) controlador animado; (vi) TYPE infrecuente; (vii) similitud baja (alta distancia) entre las raíces léxicas del español y el italiano; (viii) similitud media (media distancia) entre las terminaciones del español y el italiano; (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 la dirección del efecto se considera indeterminado. Además, se espera hallar un efecto inhibitorio 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-.

## Selección de variables

Se llevó a cabo selección de modelos (multinomiales generalizados) basado en medidas de información (Burnham y Anderson, 2010), optimizando la función de log-verosimilitud por medio de una red neuronal, utilizando el paquete *nnet* de R (Venables y Ripley, 2002). 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 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* (medida de información de Akaike) como medida de información para la selección. En lo que atañe a la importancia relativa de las predictoras sobre 2048 modelos del grupo 1, y 1024 del grupo 2, sobre el total de modelos; se observó que LDA, STEM.f, GRAMS, DIS, POS, EST4 poseían 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  $\frac{W(i)}{W(1)} > \frac{1}{8}$ ). Las predictoras promediadas que resultaron

significativas fueron: Fabs.SC.f, MOD, ANIM, FAM.LEX.f, ES, MORF.f, CUMRES.f, EST1, EST2, EST5, GRUPO6.

## Modelo bayesiano

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, 2010). La Tabla 7 muestra el criterio de información de devianza [DIC] con el objetivo de elegir la estructura aleatoria adecuada de los modelos (debe minimizarse). 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 efecto de aleatorio de ordenada al origen y tiempo. Se observa en la Tabla 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. Nótese que las correlaciones intra clase [ICC] cada  $c$  (se usó la moda para la estimación puntual) resultan bajas.

**Tabla 7.** Modelo Bayesiano: selección de estructura aleatoria

380

Medida	Modelo (i)	Modelo (ii)	Modelo (iii)
DIC	2873.743	<b>2857.807</b>	2873.343
ICC(1)	0.343	0.302	0.31
[ESS; CI(lo, up)]	[576.943; (0.165, 0.507)]	[814.635; (0.132, 0.441)]	[705.153; (0.163, 0.496)]
ICC(2)	0.002	0.002	0.0012
[ESS; CI(lo, up)]	[290.597; (3e-08, 0.336)]	[278.429; (1e-05, 0.315)]	[419.289; (7e-07, 0.337)]
ICC(3)	0.063	0.0005	0.0004
[ESS; CI(lo, up)]	[1380.019; (0.018, 0.151)]	[1103.633; (3e-08, 0.091)]	[1020.867; (5e-09, 0.119)]
ICC(4)	0.181	0.001	0.001
[ESS; CI(lo, up)]	[634.272; (0.011, 0.307)]	[518.895; (5e-09, 0.18)]	[515.107; (2e-07, 0.221)]

ICC( $c$ ) [ $c=1,2,3,4$ ]: Coeficiente de correlación intra clase; ESS: tamaño muestral efectivo; CI(lo , up):

Intervalo de credibilidad (inferior, superior).

En la ecuación (1) se escribe el modelo mixto multinomial bayesiano general (teórico) 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  $c = 1, \dots, 4$ .

$$\begin{aligned}
g(E[y_{ij} = c|v_{0ij}, v_{1ij}]) = \log \left[ \frac{\pi_{ijc}}{1 - \pi_{ijc}} \right] = & (\beta_{0c} + v_{0ic}) + (\beta_{1c} + v_{1ic})FAM.LEX.f \\
& \beta_{2c}MOD_1 + \beta_{3c}MOD_2 + \beta_{4c}MOD_3 + \beta_{5c}CUMRES.f_1 + \beta_{6c}CUMRES.f_2 \\
& \beta_{7c}Fabs.SC.f + \beta_{8c}MORF.f_1 + \beta_{9c}MORF.f_2 + \beta_{10c}EST1 + \beta_{11c}EST5 + \beta_{12c}EST2 \\
& \beta_{13c}GRUPO6_2 + \beta_{14c}GRUPO6_3 + \beta_{15c}GRUPO6_4 + \beta_{16c}GRUPO6_5 + \beta_{17c}GRUPO6_6 \\
& \beta_{18c}ES_1 + \beta_{19c}ES_2 + \beta_{20c}ANIM + \beta_{21c}TIME
\end{aligned} \tag{1}$$

La Tabla 8 describe los parámetros del modelo a estimar, para *cada* modelo  $c = 1,2,3,4$ . Hubo 22 efectos fijos a estimar por modelo.

**Tabla 8.** Parámetros del modelo logístico multinomial bayesiano

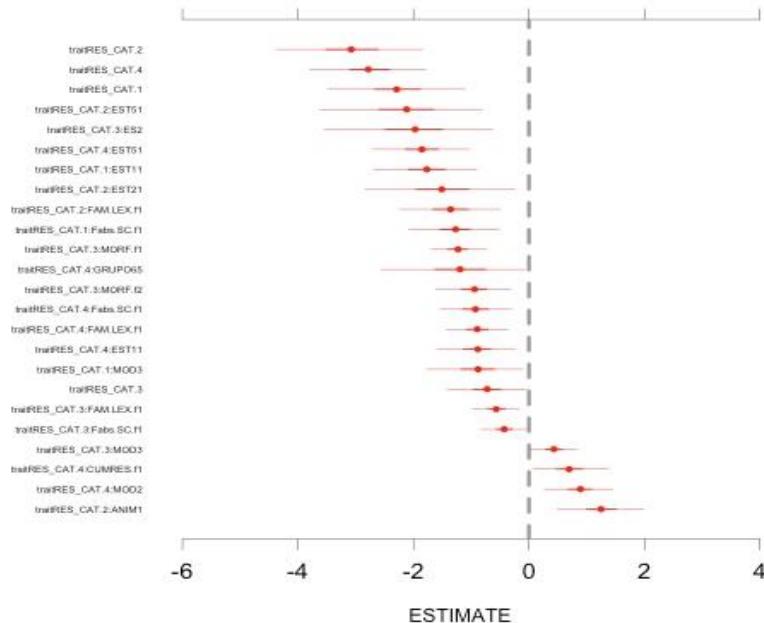
Parámetro	Descripción	381
$v_{0ic}$	la desviación del grupo $i_c$ de la ordenada al origen	
$v_{1ic}$	la desviación del grupo $i_c$ de la media marginal de FAM.LEX.f.	
$\beta_{0c}$	la media basal marginal	
$\beta_{1c}$	el efecto de FAM.LEX.f, nivel 1 (referencia: FAM.LEX.f = 0)	
$\beta_{2c}$	el efecto de MOD, nivel 1 (referencia: MOD = 0)	
$\beta_{3c}$	el efecto de MOD, nivel 2 (referencia: MOD = 0)	
$\beta_{4c}$	el efecto de MOD, nivel 3 (referencia: MOD = 0)	
$\beta_{5c}$	el efecto de CUMRES.f, nivel 1 (referencia: CUMRES.f = 0)	
$\beta_{6c}$	el efecto de CUMRES, nivel 2 (referencia: CUMRES.f = 0)	
$\beta_{7c}$	el efecto de Fabs.SC.f, nivel 1 (referencia: Fabs.SC.f = 0)	
$\beta_{8c}$	el efecto de MORF.f, nivel 1 (referencia: MORF.f = 0)	
$\beta_{9c}$	el efecto de MORF.f nivel 2 (referencia: MOD = 0)	
$\beta_{10c}$	el efecto de EST1, nivel 1 (referencia: EST1 = 0)	
$\beta_{11c}$	el efecto de EST5, nivel 1 (referencia: EST1 = 0)	
$\beta_{12c}$	el efecto de EST2, nivel 1 (referencia: EST1 = 0)	
$\beta_{13c}$	el efecto de GRUPO6, nivel 2 (referencia: GRUPO6 = 1)	
$\beta_{14c}$	el efecto de GRUPO6, nivel 3 (referencia: GRUPO6 = 1)	
$\beta_{15c}$	el efecto de GRUPO6, nivel 4 (referencia: GRUPO6 = 1)	
$\beta_{16c}$	el efecto de GRUPO6, nivel 5 (referencia: GRUPO6 = 1)	
$\beta_{17c}$	el efecto de GRUPO6, nivel 6 (referencia: GRUPO6 = 1)	
$\beta_{18c}$	el efecto de ES, nivel 1 (referencia: ES = 0)	
$\beta_{19c}$	el efecto de ES, nivel 2 (referencia: ES = 0)	
$\beta_{20c}$	el efecto de ANIM, nivel 1 (referencia: ES = 0)	
$\beta_{21c}$	el efecto de TIME	

## RESULTADOS

### Factores influyentes para cada tipo de error (efectos fijos)

Se tomó la media de la posterior de los coeficientes  $\beta$  como medida puntual<sup>1</sup>. La Figura 1 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], 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) fue 0.036, el intervalo de credibilidad resultó (-2.366; 0.050), como incluye al cero no se consideró dicho predictor.

382

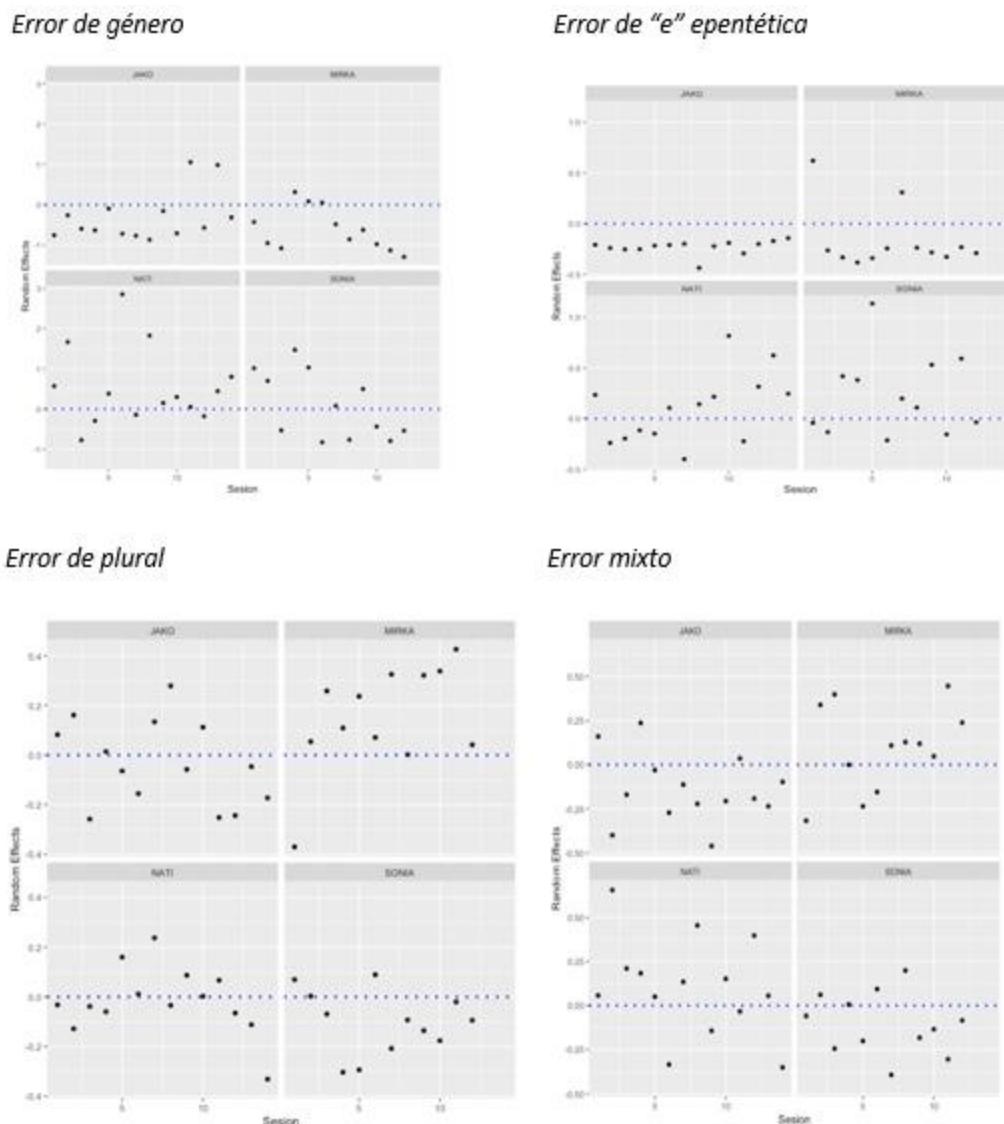


**Figura 1.** Forest Plot: estimaciones puntuales de efectos simples e intervalos de credibilidad (escala del logit)

<sup>1</sup> La correlación de Pearson entre medias y modas de los coeficientes fue  $r = 0.988$ .

### **Trayectorias para cada tipo de error según sujeto (Efectos aleatorios)**

La Figura 2 ilustra 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 (línea punteada) 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. Al 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 2.** Trayectorias de los factores aleatorios de ordenada al origen

### Análisis de errores

#### Error de género vs. no error

La chance de cometer un error de concordancia ( $Y = 1$ ) respecto de no cometerlo ( $Y = 0$ ) *disminuye* como se describe a continuación. En primera instancia, cuando se pasa de concordancias con artículo definido a aquellas con adjetivos; por ejemplo: *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 vuelve análoga “-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 cuanto a 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.) > *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 asemejar *le* > *las*); (iii) géneros diferentes en ambas lenguas: *i film* (masc. pl.); otros casos irregulares: *gli analfabeti* (masc. pl.) [masc. sing.: *l'analfabeta*] > *los analfabetas*. En suma, el caso del error de género parece estar más asociado al controlador que al tipo de modificador.

En segundo lugar, si se pasa de concordancias con TYPE infrecuente a aquellas con TYPE frecuente; por ejemplo: *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]. En tercer lugar, cuando se pasa de concordancias en las que no se puede aplicar la estrategia EST1 (poner plural en -os si el plural italiano termina en -i) a aquellas en donde se puede aplicar; por ejemplo: *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].

385

#### *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 se pasa de concordancias con controlador no familiar o infrecuente a aquellas con controlador familiar o frecuente. Constituyen ejemplos: *calles grandes* [SONIA, sesión 4, línea 60]; *las mujeres* [NATI, sesión 13, línea 129]; *pocos trenos* [MIRKA, sesión 1, línea 158]. En segunda instancia, cuando se pasa de concordancias en las que no se puede aplicar la estrategia EST2 (poner plural en -as si el plural italiano termina en -e) a aquellas en las que se puede aplicar. Por último, cuando se pasa de concordancias en las que no se puede aplicar la estrategia EST5 (si palabra singular del italiano termina en -e, agregar -s a la desinencia en italiano) a aquellas en las que se puede aplicar.

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; por ejemplo: *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 “alemán” en español y forman el plural a partir de éste: “alemanos”, en lugar de insertar

-e<sup>2</sup>. 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 italiana 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: [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*.

#### *Error de plural vs. no error*

La chance de cometer un error de concordancia ( $Y = 1$ ) respecto de no cometerlo ( $Y = 0$ ) *disminuye* en los casos siguientes. En primer lugar, cuando se pasa de concordancias con TYPE infrecuente a aquellas con TYPE frecuente; por ejemplo: *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]. En segundo lugar, cuando se pasa de concordancias sin -e- epentética en ningún término de la concordancia a aquellas con -e- epentética en ambos términos. 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. En tercer lugar, cuando se pasa de concordancias con controlador no familiar o infrecuente a aquellas con controlador familiar o frecuente; por ejemplo: *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]. En cuarto lugar, cuando se pasa de concordancias con similitud alta (distancia baja) de terminaciones entre español e italiano a aquellas con similitud media (distancia media); por ejemplo: *la herbas* [SONIA, sesión 6, línea 78]; *muchas lluvias* [NATI, sesión 5, línea 71]; *cartas escrita* [JAKO, sesión 2, línea 172]; *los pollo* [MIRKA, sesión 3, línea 34]. Por último, cuando se pasa de concordancias con similitud alta (distancia baja) de terminaciones entre español e italiano a aquellas con similitud alta (distancia baja). Constituyen 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. Son ejemplos: *días ocupado* [SONIA, sesión 8, línea 369]; *solucionse secretas* [NATI, sesión 7, línea 109], *diferentes fuente* [JAKO, sesión 7, línea 186]; *platos típico* [MIRKA, sesión 4, línea 51].

<sup>2</sup> Idéntico error se produce con *il treno* > los \*trenos (trenes).

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 (similitud baja) tuvieron efecto negativo, aunque menor que en el caso de la distancia media. Examinando los ejemplos del corpus de distancia baja (similitud alta) con errores de concordancia, se observa que muchos de los determinantes eran posesivos de tercera persona plural ("sus") [18/33] y la mayoría contenían errores. En italiano la construcción de posesivo cambia según la referencia sea: (i) a la tercera persona singular: *le sue case* ('sus casas' [de el / ella]); (ii) a la tercera persona plural: *le loro case* ('sus casas' [de ellos/ellas]). Otra posibilidad es el hecho de que veinte de los 33 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 de *amici* > *amigos* requiere de una sustitución más una inserción. Los pesos de las operaciones en el algoritmo fueron: DELETE = 0.3 < SUBSTITUTE = 0.6 < INSERT = 1. Quizás "borrado" + "inserción" requiere en realidad más "costo" del estipulado; 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 dos instancias del determinante "diferente": *diferente personas* [JAKO, sesión 2, línea 59], *diferente postaciones (ubicaciones)* [JAKO, sesión 8, línea 48]. En italiano "diverso" tiene cuatro formas: *divers-o masc. sing. / -a fem. sing. / -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 *cuattrocento*, que en italiano no lleva concordancia: *cuatrociento kilómetros* [JAKO, sesión 10, línea 71] y *cuatrociento habitantes* [JAKO, sesión 12, línea 15]. En suma, el efecto encontrado pudo deberse a la presencia de numerosos posesivos, los cuales, en términos de la distancia de Levenshtein requieren de operaciones de "borrado" y "sustitución", lo que podría ser más demandante. En segundo lugar, la aparición de plurales invariantes en italiano. Luego, el uso de plurales con "diferente-s" [*divers-e*] y controladores femeninos, donde se puede confundir la terminación del singular del español con el femenino plural del italiano. Por último, la aparición de numerales que carecen de género en italiano.

#### *Error mixto vs. no error*

La chance de cometer un error de concordancia ( $Y = 1$ ) respecto de no cometerlo ( $Y = 0$ ) *disminuye* como sigue. En primer lugar, cuando se pasa de concordancias con TYPE infrecuente a aquellas con TYPE frecuente; por ejemplo: *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]. En segundo lugar, cuando se pasa de

concordancias con controlador no familiar o infrecuente a aquellas con controlador familiar o frecuente; por ejemplo: *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]. En tercer lugar, cuando se pasa de concordancias en las que no se puede aplicar la estrategia EST1 a aquellas en donde se puede aplicar; por ejemplo: *este daños* [NATI, sesión 12, línea 104]; *procesos mental* [JAKO, sesión 5, línea 199]; *todo lo gener* [MIRKA, sesión 2, línea 246]. En cuarto lugar, cuando se pasa de concordancias en las que no se puede aplicar la estrategia EST5 a aquellas en las que se puede aplicar. Son 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. sing. *istituzione* > esp. pl. *instituciones* (instancia: “mucho *instituciones*”); it. sing. *professore* > esp. pl. *profesores* (instancia: “los *profesor*”); it. sing. *inondazione* > esp. pl. *inundación* (instancia: “*inundaciones peligroso*”); it. sing. *animale* > esp. pl. *animales* (instancia: “*animales común*”); it. sing. *funzione* > esp. pl. *funciones* (instancia: “*estas función*”); it. sing. *essere* > esp. pl. *seres*; it. sing. *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.

Por otra parte, la chance de cometer un error de concordancia ( $Y = 1$ ) respecto de no cometerlo ( $Y = 0$ ) *aumenta* en dos casos. Primero, cuando se pasa de concordancias con artículo definido a aquellas con determinante. Son 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]. Segundo, cuando se pasa de concordancias con hasta 2 errores acumulados (sin contar el error actual) a aquellas con hasta 7 errores acumulados. Constituyen ejemplos: *cuánta vez* [SONIA, sesión 6, línea 294]; *muchas personaje* [NATI, sesión 8, línea 32], *tan hombres* [JAKO, sesión 1, línea 68]; *grande directore* [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.

#### *Errores según TYPES más frecuentes*

Se ha visto que la alta frecuencia de TYPES de concordancia ayuda a evitar los tipos de errores de género, plural y mixto. La Tabla 9 ilustra dichos TYPES frecuentes, la cantidad de errores según el tipo (entre corchetes) y algunos ejemplos del corpus. En general, la mayoría de los casos involucran errores de plural. En éstos, una buena parte se produce con concordancias “armónicas” del tipo “as”-“as” / “os”-“os”. De las “inarmónicas” destacan aquellas en “os”- “\*es” y “as”-“\*es”, las cuales también tienen varios casos de errores mixtos.

**Tabla 9.** Tipos de errores según TYPES de concordancia de alta frecuencia

<b>TYPES</b>	<b>Género</b>	<b>Plural</b>	<b>Mixto</b>
<b>d-n-as-*es</b>	<i>pocos televisiones</i> (NATI, 14, 6) [3]	<i>muchas motivaciones</i> (MIRKA, 9, 185) [2]	<i>muchas posibilidad</i> (MIRKA, 12, 311)[7]
<b>d-n-as-as</b>	<i>muchos cartas</i> (SONIA, 4, 155 )[6]	<i>esta historias</i> (NATI, 8, 32)[15]	<i>este empresa</i> (NATI, 10, 206)[5]
<b>d-n-os-os</b>	<i>muchos tiendas</i> (NATI, 1, 253)[1]	<i>otros mundo</i> (MIRKA, 4, 8)[6]	<i>este chicos</i> (MIRKA, 4, 207)[3]
<b>l-n-as-*es</b>	<i>les ciudades</i> (SONIA, 5, 235) [2]	<i>la elecciones</i> (NATI, 13, 144)[7]	<i>las mujera</i> (NATI, 13, 124)[7]
<b>l-n-as-as</b>	<i>les cosas</i> (SONIA, 4, 89)[1]	<i>las escuela</i> (MIRKA, 3, 210)[22]	<i>los libretes</i> (MIRKA, 11, 70)[3]
<b>l-n-j-os-os-os</b>	[0]	<i>los niño solos</i> (SONIA, 6, 243)[6]	[0]
<b>l-n-os-*es</b>	[0]	<i>lo automóviles</i> (MIRKA, 7, 21)[11]	<i>los profesor</i> (SONIA, 8, 386)[4]
<b>l-n-os-es</b>	[0]	<i>los lenguaje</i> (MIRKA, 12, 279)[7]	<i>le estudiantes</i> (SONIA, 4, 123)[1]
<b>l-n-os-os</b>	<i>las huevos</i> (MIRKA, 6, 278)[5]	<i>los médico</i> (NATI, 7, 83)[21]	<i>la sitios</i> (MIRKA,2,102)[1]

Leyenda: l = artículo definido, d = determinante, n = nombre, j = adjetivo, terminaciones de los términos de la concordancia: "os", "as", "es", "\*es" (-e- epentética), referencia del ejemplo en el corpus: (Sujeto, sesión, línea), [número de casos].

## DISCUSIÓN

La Tabla 10 muestra las expectativas cumplidas del análisis. El tipo de modificador [MOD] fue factor de riesgo en el caso del error mixto con determinantes. Los adjetivos resultaron factor de protección para los errores de género, pero fueron factor de riesgo para los errores de plural. No obstante, el análisis de las instancias reveló que los errores no se debían al tipo de modificador para el error de género. Fabs.SC.f (alta frecuencia de TYPES) y FAM.LEX.f (alta familiaridad y frecuencia léxica del controlador) fueron factores de protección para tres de los cuatro tipos de errores, siguiendo lo esperado. El hecho de haber encontrado efectos de riesgo para la animicidad (en línea con la literatura) y la familiaridad / frecuencia del controlador aboga por la inclusión de más características del controlador en los análisis. CUMRES.f (errores acumulados) solamente resultó significativo para la respuesta categórica en errores mixtos. 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]. En ambos casos el efecto observado resultó contrario al esperado. Se especuló con la posible dificultad asociada a posesivos, numerales, formas invariantes del italiano e instancias en las que se

puede confundir la terminación del singular del español con el femenino plural del italiano. La estrategia EST1 verificó la expectativa de factor de protección en errores de género y mixtos. En cambio, el efecto de protección contra error la estrategia EST2 fue más débil (error de *-e-* epentética). Por último, en ES(2) [-*e-* epentética en ambos términos de la concordancia] no verificó ningún error de plural; por tanto, la presencia de *-e-* epentética no pone dificultades a los cuatro aprendientes. 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 [EST5 fue factor de protección]. 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. Contrariamente a la literatura, no se encontraron efectos de larga distancia.

Tabla 10. Resultados

Predictor	NIVELES		
	[efecto esperado]	[efecto hallado]	TIPO DE ERROR
MOD [↑]	MOD(2) - MOD(0) [↑ ]		Mixto
	MOD(3) - MOD(0) [↓ ] [↑ ]		Género, plural
Fabs.SC.f [↓]	Fabs.SC.f(1) - Fabs.SC.f(0) [↓ ]		género, plural, mixto
390 MORF.f [↑]	MORF.f(1) - MORF.f(0) [↓ ]		Plural
	MORF.f(2) - MORF.f(0) [↓ ]		Plural
FAM.LEX.f [↓]	FAM.LEX.f(1) - FAM.LEX.f(0) [↓ ]		- <i>e-</i> epentética, plural, mixto
CUMRES.f [↑]	CUMRES.f(1) - CUMRES.f(0) [↑ ]		Mixto
EST1 [↓]	EST1(1) - EST1(0) [↓ ]		género, mixto
EST5 [↓]	EST5(1) - EST5(0) [↓ ]		- <i>e-</i> epentética, mixto
EST2 [↓]	EST2(1) - EST2(0) [↓ ]		- <i>e-</i> epentética
ANIM [↑]	ANIM(1) - ANIM(0) [↑ ]		- <i>e-</i> epentética
ES [↑]	ES(2) - ES(0) [↓ ]		Plural

(↑) = efecto aumenta; (↓) = efecto disminuye

A partir de lo hallado, se podrían desprender algunos consejos para la enseñanza de la concordancia plural en ELE para alumnos de habla italiana. En primer lugar, se comenzaría enseñando las concordancias armónicas en “as-as”, “os-os” porque son las que obtuvieron más casos facilitadores de TYPES; para luego pasar a las “inarmónicas”, en particular a las que involucren “\*es” [-*e-* epentética]. Además, sería de gran ayuda para los estudiantes aprender la estrategia EST5, que parece evitar errores con *-e-* epentética y mixtos. Las estrategias EST1 y EST2 son ya bastante usuales en los alumnos. Otro efecto notoriamente presente en la literatura y en el presente trabajo 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 en los que 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 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. Por otra parte, el análisis de instancias de la variable MORF.f reveló mayor dificultad con los posesivos y plurales invariantes del italiano. El italiano posee una morfología “fusiva” (versus “aglomerativa” en español), con lo cual, los alumnos deben “crear” una posición por “llenar” con los alomorfos de plural. Sería provechoso enseñar las concordancias haciendo que los alumnos marquen la pronunciación de dichos alomorfos para facilitar la permanencia en la memoria a largo plazo. También sería una buena idea 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.

Este trabajo es de naturaleza observacional. Sus resultados podrían constituir hipótesis a ser evaluadas en eventuales experimentos conductuales que controlen los sesgos.

391

## REFERENCIAS

Alarcón, I. (2009). The processing of gender agreement in L1 and L2 Spanish: Evidence from Reaction Time Data. *Hispania*, 92(4), 814-828. <https://www.jstor.org/stable/40648463>.

Alarcón, I. (2011). Spanish gender agreement under complete and incomplete acquisition: Early and late bilinguals' linguistic behavior within the noun phrase. *Bilingualism: Language and Cognition*, 14(3), 332-350. <https://doi.org/10.1017/S1366728910000222>.

Bruhn de Garavito, J. (2008). Acquisition of the spanish plural by french L1 speakers: the role of transfer. En J. Liceras, H. Zobl y H. Goodluck (Eds.), *The role of features in second language acquisition* (pp. 270-298). <https://doi.org/10.4324/9781315085340-9>.

Bruhn de Garavito, J., y White, L. (2002). The second language acquisition of Spanish DPs: The status of grammatical features. En A.T. Pérez-Leroux y J. Muñoz (Eds.), *The Acquisition of Spanish Morphosyntax: The L1/L2 Connection* (pp. 153-178). Kluwer Academic Publishers.

Burnham, K., y Anderson, D. (2010). *Model selection and multimodel inference: a practical information-theoretic approach* (2<sup>a</sup> ed.). Springer.

Corbett, G. (2006). *Agreement*. Cambridge University Press.

Davis, C., y Perea, M. (2005). BuscaPalabras: a program for deriving orthographic and phonological neighborhood statistics and other psycholinguistic indices in Spanish. *Behavior Research Methods*, 37(4), 665-671. <https://doi.org/10.3758/BF03192738>.

Dowens, M., Vergara, M., Barber, H., y Carreiras, M. (2010). Morphosyntactic processing in late second-language learners. *Journal of Cognitive Neuroscience*, 22(8), 1870-1887. <https://doi.org/10.1162/jocn.2009.21304>.

Español-Echevarría, M., y Prévost, P. (2004). Acquiring number specification on L2 Spanish quantifiers: Evidence against the rich agreement hypothesis. En J. Van Kampen y S. Baauw (Eds.), *Proceedings of the 2003 conference on generative approaches to language acquisition* (pp. 151-162). LOT.

Fernández-García, M. (1999). Patterns of gender agreement in the speech of second language learners. En J. Gutiérrez-Rexach y F. Martínez-Gil (Eds.), *Advances in hispanic linguistics: papers from the 2nd Hispanic Linguistics Symposium* (pp. 3-15). Cascadilla Press.

Foote, R. (2011). Integrated knowledge of agreement in early and late English-Spanish bilinguals. *Applied Psycholinguistics*, 32(1), 187-220. <https://doi.org/10.1017/S0142716410000342>.

Foote, R. (2015). The production of gender agreement in native and L2 Spanish: The role of morphophonological form. *Second Language Research*, 31(3), 343-373. <https://doi.org/10.1177/0267658314565691>.

Franceschina, F. (2001). Morphological or syntactic deficit in near-native speakers? An assessment of some current proposals. *Second Language Research*, 17(3), 213-247. <https://doi.org/10.1177/026765830101700301>.

González, P., Mayans, D., y Van der Bergh, H. (2019). Nominal agreement in the interlanguage of Dutch L2 learners of Spanish. *International Review of Applied Linguistics in Language Teaching*. <https://doi.org/10.1515/iral-2017-0174>.

Hadfield, J. (2010). MCMC Methods for Multi-Response Generalized Linear Mixed Models: The MCMCglmm R Package. *Journal of Statistical Software*, 33(2), 1-22. <https://doi.org/10.18637/jss.v033.i02>.

Keating, G.D. (2009). Sensitivity to violations of gender agreement in native and nonnative Spanish: An eye-movement investigation. *Language Learning*, 59(3), 503-535. <https://doi.org/10.1111/j.1467-9922.2009.00516.x>.

Keating, G.D. (2010). The effects of linear distance and working memory on the processing of gender agreement in Spanish. En B. VanPatten y J. Jegerski (Eds.), *Research in Second Language Processing and Parsing* (pp. 113-134). John Benjamins.

Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., y Suchomel, V. (2014). The Sketch Engine: Ten years on. *Lexicography*, 1(1), 7-36. <https://doi.org/10.1007/s40607-014-0009-9>.

Muñoz, J., Díaz, L., y Mongeon, C. (2000). N-drop and determiners in native and non-native Spanish: More on the role of morphology in the acquisition of syntactic

knowledge. En R.P. Leow y C. Sanz (Eds.), *Current research on the acquisition of Spanish* (pp. 67-96). Cascadilla Press.

Lichtman, K. (2009). Acquisition of Attributive and Predicative Adjective Agreement in L2 Spanish. En M. Bowles, T. Ionin, S. Montrul y A. Tremblay (Eds.), *Proceedings of the 10th Generative Approaches to Second Language Acquisition Conference* (pp. 231-247). <http://www.lingref.com/cpp/gasla/10/abstract2273.html>.

McCarthy, C. (2008). Morphological variability in the comprehension of agreement: An argument for representation over computation. *Second Language Research*, 24(4), 459-486. <https://doi.org/10.1177/0267658308095737>.

MacWhinney, B. (2000). *The CHILDES Project: Tools for Analyzing Talk* (3<sup>rd</sup> ed.). Lawrence Erlbaum Associates.

Montrul, S., Foote, R., y Perpiñan, S. (2008). Gender agreement in adult second language learners and Spanish heritage speakers: The effects of age and context of acquisition. *Language Learning*, 58(3), 503-553. <https://doi.org/10.1111/j.1467-9922.2008.00449.x>.

Nerbonne, J., Van Ommen, S., Gooskens, C., y Wieling, M. (2013). Measuring socially motivated pronunciation differences. En L. Borin y A. Saxena (Eds.), *Approaches to Measuring Linguistic Differences* (pp. 107-140). <https://doi.org/10.1515/9783110305258>.

Oakes, M. P. (1998). *Statistics for Corpus Linguistics*. Edinburgh University Press.

O'Grady, W. (2005). *Syntactic Carpentry: An Emergentist Approach to Syntax*. Lawrence Erlbaum Associates.

Peña, D. (2002). *Análisis de Datos Multivariantes*. Mc Graw Hill.

Sagarra, N. (2007). Online processing of gender agreement in low proficient English-Spanish late bilinguals. En J. Camacho, N. Flores-Ferrán, L. Sánchez, V. Déprez y M.J. Cabrera (Eds.), *Romance Linguistics 2006* (pp. 240-253). <https://doi.org/10.1075/cilt.287.18sag>.

Sagarra, N., y Herschensohn, J. (2013). Processing of gender and number agreement in late spanish bilinguals. *International Journal of Bilingualism*, 17(5), 607-627. <https://doi.org/10.1177/1367006912453810>.

Scrucca, L., Fop, M., Murphy, T.B., y Raftery, A.E. (2016). Mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal*, 8(1), 289-317. <https://doi.org/10.32614/RJ-2016-021>.

Van Buuren, S., y Groothuis-Oudshoorn, K. (2011). Mice: multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3), 1-67. <https://doi.org/10.18637/jss.v045.i03>.

Venables, W.N., y Ripley, B.D. (2002). *Modern Applied Statistics with S* (4<sup>a</sup> ed.). Springer.

White, L., Valenzuela, E., Kozlowska-Macgregor, M., y Leung, Y. (2004). Gender and number agreement in nonnative Spanish. *Applied Psycholinguistics*, 25(1), 105-133. <https://doi.org/10.1017/S0142716404001067>.