

**Análisis de tiempos hasta que se produce un error de concordancia en cuatro estudiantes italianos de ELE**

Marafioti, Pablo Ezequiel\*, a

**Artículo Original**

**Resumen**

**Abstract**

**Tabla de Contenido**

Se consideran diferentes factores intervinientes en la producción de concordancia plural en cuatro aprendices italianos de ELE, en un estudio de caso longitudinal, utilizando un análisis de tiempo hasta el evento "error de concordancia". Se aplicó un modelo de eventos de errores múltiples y otro de riesgos competitivos. Se categorizaron cuatro tipos de errores: género, -e- epentética, plural, mixto. Se hallaron efectos significativos para los siguientes factores: (i) 'determinantes' (adjetivos posesivos, indefinidos, demostrativos, interrogativos, exclamativos) y 'adjetivos' (calificativos, numerales, ordinales); (ii) sustantivos animados; (iii) concordancias en donde se podía aplicar la estrategia de poner plural en "-os" en plurales italianos terminados en "-i"; (iv) en aquellas donde se podía aplicar la estrategia de poner plural en "-es" con palabras italianas singulares terminados en "-e"; (v) con palabras cuyas desinencias tenían similitud media o baja con las del italiano; (vi) con sustantivos familiares y/o frecuentes; (vii) instancias con TYPES más frecuentes.

**A time-to-event analysis of agreement errors in four Italian learners of Spanish.**  
Different factors involved in the production of plural agreement in four Italian ELE learners are considered in a longitudinal case study, using an analysis of time until the event "agreement error" (a multiple error event model and a competitive risk model). The errors were classified as: gender, epenthetic -e-, plural, mixed. Significant effects were found for the following factors: (i) 'determinants' (indefinite, possessive, demonstrative, interrogative, exclamatory) and 'adjectives' (proper, ordinal, cardinal); (ii) animate nouns; (iii) instances of agreement in which the student could form a plural in "-os" from an Italian plural form in "-i"; (iv) in which the learner could apply a strategy by creating a plural in "-es" from an Italian singular form in "-e"; (v) agreement with words whose endings had a low or medium similarity to those of the Italian words; (vi) instances including familiar and / or frequent nouns; (vii) with more frequent TYPES.

Introducción	116
Método	118
Participantes y diseño del corpus	118
Variables creadas	119
Análisis de Sobrevivencia	120
Modelo de Eventos Múltiples	122
Modelo de Riesgos Competitivos	122
Efectos Esperados para ambos Modelos	122
Resultados	123
Discusión	124
Referencias	127

*Palabras clave:* español lengua extranjera, adquisición, concordancia, número.

*Keywords:* Spanish foreign language, acquisition, agreement, number.

Recibido el: 23 de enero de 2021; Aceptado el 4 de agosto de 2021

Editaron este artículo: Jazmín Cevalco, Paula Abate, Sebastián Miranda y Belén Cañadas.

**Introducción**

En el presente trabajo se identifican algunos factores influyentes en la producción de concordancia plural en cuatro aprendices italianos de español como lengua extranjera (ELE). Se trata de un estudio de caso longitudinal y observacional. También se busca 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 aplican modelos estadísticos del ámbito del análisis de datos de tiempos hasta un evento, en este caso, la ocurrencia del evento "error de concordancia". La intención es aplicar dicha metodología a ELE, no explorada hasta el momento.

Definimos 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',

<sup>a</sup> Facultad de Lenguas, Universidad Nacional de Córdoba, Córdoba, Argentina  
\*Enviar correspondencia a: Marafioti, P. E. E-mail: kenavo38@yahoo.com

Citar este artículo como: Marafioti, P. E. (2022). Análisis de tiempos hasta que se produce un error de concordancia en cuatro estudiantes italianos de ELE. *Revista Argentina de Ciencias del Comportamiento*, 14(3), 116-128.

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. En el presente trabajo, el controlador será nominal. Los objetivos consistirán en artículos (definidos e indefinidos), adjetivos y pronombres (demostrativos, posesivos, indefinidos). 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 italiano, los nombres que terminan en -o suelen ser masculinos (*l'albero* ['el árbol']) y en -a, femeninos (*la stella* ['la estrella']), en similitud con el español. A su vez, al igual que en español, los terminados en -ma y -ta pertenecen al masculino (*il programma* ['el programa'], *il pirata* ['el pirata']) y, algunos son de género común (*il / la artista* ['el artista' / 'la artista']). Las formas terminadas en -e se asignan a ambas clases (*il leone* ['el león'], *la luce* ['la luz']), del mismo modo que en el español. 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 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 alomorfo 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 alomorfo 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']. A su vez, 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 primera a tercera persona (singular / plural) tienen diferentes alomorfos, 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 invariantes en -a también: *il cinema / i cinema* ['el / los cine(s)']; *il delta / i delta* ['el / los delta(s)'] (Carrera Díaz, 1989).

La adquisición del plural en español tiende a seguir las etapas: plural nulo > plural en /-s/ > plural en /-es/ (Bruhn de Garavito, 2008). Los errores de concordancia de número tienden a ser menos que los de género. A su vez, estos últimos persisten incluso luego de muchos años de producción oral (Franceschina, 2001; Muñoz Licerias, Díaz Rodríguez, & Mongeon, 2000; White, Valenzuela, Kozłowska-Macgregor, & Leung, 2004). La forma de *default* tiende a ser no marcada. Es decir, el singular para el número y el masculino para el género (Alarcón, 2011; Bruhn de Garavito & White, 2002; Fernández-García, 1999; Finnemann, 1992; McCarthy, 2008; Montrul, Foote, & Perpiñan, 2008; 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. Es decir, en bilingües tempranos y tardíos, tanto en producción como en procesamiento (Alarcón, 2011; Fernández-García, 1999; Gillon Dowens, Vergara, Barber, & Carreiras, 2010; Montrul et al., 2008; White et al., 2004). En cuanto al género, resulta más fácil producir y procesar la concordancia: (a) de masculino respecto al femenino (Alarcón, 2011; Bruhn de Garavito & 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*; (Alarcón, 2011; Fernández-García, 1999; Montrul et al., 2008); (c) con controladores inanimados (como en *hospital*) respecto de los animados (como en *doctor/a*; Sagarra & Herschensohn, 2013; aunque Alarcón, 2009 encontró el efecto contrario en núcleos de SN complejos); (d) En sintagmas nominales complejos del tipo *N1 de N2* [*el doctor del hospital*], cuando el género de N1 coincide con el N2 (Foote, 2015).

Por otra parte, González, Mayans, y Van der Bergh (2022) constataron los siguientes efectos para errores de concordancia de género y número en el ámbito nominal en composiciones escritas: (i) del plural respecto del singular; (ii) del femenino respecto del masculino; (iii) en los artículos femeninos (sin importar el rasgo de número). Por su parte, en estudios de procesamiento, se ha verificado que el aumento de distancia estructural (cantidad de nodos sintácticos entre controlador y objetivo) produce que la concordancia se procese con más lentitud en el dominio no local respecto al local (Foote, 2011; Gillon Dowens et al., 2010; Keating, 2009; 2010; Lichtman, 2009; Sagarra, 2007).

En este trabajo, se examinan posibles factores intervinientes en el error de concordancia, que amplían los ya estudiados en la literatura. En primer lugar, se tienen en cuenta los efectos de diferentes tipos de modificadores respecto del artículo. En coincidencia con estudios previos, se examina el posible efecto de la concordancia a larga distancia. A su vez, se consideran características del controlador, más allá de la animicidad, y efectos de frecuencia de esquemas

de concordancia y de errores cometidos en el curso de una sesión. Por su parte, se incluyen posibles estrategias de aprendizaje. A diferencia de los métodos aplicados en la literatura previa, el análisis de sobrevivencia es especialmente adecuado para analizar la adquisición en el tiempo; es decir, adoptando un enfoque dinámico.

## Método

### Participantes y diseño del corpus

Se analizan datos de cuatro casos de estudiantes de ELE. Se trató de cuatro alumnos 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* estaba 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 (2020). 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, ej.: *leer o hablar con muchos personas* [SONIA, sesión 1]; (iii) [\*2] = errores debidos al uso de la terminación "(-e)-s" (inserción de -e-epentética), ej.: *después las doce hay muchos trenos* [SONIA, sesión 7]; (iv) [\*3] = errores de plural, o sea ausencia de -s, ej.: *los veneciano conocen donde ir por comprar mejor* [SONIA, sesión 5]; (v) [\*4] = errores mixtos por acumulación de los anteriores, ej.: *por les joven* [SONIA, sesión 2]. El conteo se hizo con el programa CLAN (Mac Whinney, 2020).

## Variables Creadas

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.

**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 [ES]: 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 refiere el nombre puede moverse o no por propia voluntad. Ej.: Animados: “niño”, “perro”; Inanimados: “planta”, “auto”, “edificio”, según: 0 = *inanimado*, 1 = *animado*.

**Fabs.SC.f.** 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 y/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 [variable  $Fabs_c$ ] y de un corpus del español electrónico *online* [variable  $Fabs_s$ ]. Para esto último, se apeló al corpus del español *EsTenTen* de *Sketch Engine* (Kilgarriff et al., 2014). Primero, se las transformó según:

$$Fabs_c = \log(Fabs_c + 1) \quad y$$

$Fabs_s = \log(Fabs_s + 1)$ . Dado que ambas se hallaban correlacionadas se obtuvo un índice a partir de ambas utilizando Análisis de Componentes Principales [PCA] (Peña, 2002), llamándolo “Fabs.SC.f”. Luego, dicho índice se discretizó usando clustering por mezcla de gaussianas (Scrucca, Fop, Murphy, & Raftery, 2016), en los niveles: 1 = frecuencia alta, 0 = frecuencia baja.

**IMA.CONC.f.** Índice a partir de PCA combinando los siguientes rasgos del controlador extraídos de la base de datos “BuscaPalabras” (Davis & Perea, 2005): (i) Concretud (CONC): índice subjetivo en escala de 1 a 7 que indica cuán concreta es una palabra de menos (+ abstracta) a más (+ concreta); (ii) Imaginabilidad (IMA): índice subjetivo en escala de 1 a 7 que indica la intensidad con la que una palabra evoca imágenes. Se discretizó el índice PCA mediante clustering por mezcla de gaussianas, en los niveles: 1 = alto, 0 = bajo.

**FAM.LEX.f.** Índice a partir de PCA combinando los siguientes rasgos del controlador extraídos de la base de datos “BuscaPalabras”: (i) Familiaridad (FAM): índice subjetivo en escala de 1 a 7, que indica cuán frecuentemente una palabra es oída, leída o producida diariamente; (ii) Frecuencia (LEXESP): frecuencia de la palabra en el corpus “BuscaPalabras”, en escala por mil; y transformada como:

$LEXESP = \log(LEXESP + 1)$ . Se discretizó el índice PCA mediante clustering por mezcla de gaussianas, en los niveles: 1 = alto, 0 = bajo.

**STEM.f y MORF.f.** Se crearon dos variables basadas en la distancia de *Levenstein* (Nerbonne, Van Ommen, Gooskens, & Wieling, 2013; Oakes, 1998), con el objetivo de medir la similitud entre las raíces léxicas entre el español y el italiano (STEM); y entre los morfemas de género y número *plural* (MORF). 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. Se las discretizó mediante clustering por mezcla de gaussianas, según: (i) MORF.f con niveles: 0 = distancia baja, 1 = distancia media, 2 = distancia alta; (ii) STEM.f con niveles: 0 = distancia baja; 1 = distancia alta.

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). Dichas estrategias buscaron identificar casos que facilitarían o dificultarían la producción de concordancias

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* u 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*.

Respecto de la noción de “transfer” (la aplicación de rasgos lingüísticos de una lengua a otra), vale la pena notar que las estrategias EST1,

EST2, EST4, EST5, EST6 implicarían un “transfer” positivo, ayudando a que el alumno no cometa errores de concordancia.

Se recolectaron 1857 casos de concordancia en total. Los casos faltantes representaron el 8.6 % de la base de datos. Para solucionarlo, se utilizó el paquete *mice* [*Multivariate Imputation by Chained Equations*] de *R* (Van Buuren & Groothuis-Oudshoorn, 2011), que realiza imputación múltiple.

El material complementario al presente artículo y el código de *R* utilizado se encuentra en: [https://github.com/pablomarafioti/PabloMarafioti/tree/master/analisi\\_de\\_sobrevivencia](https://github.com/pablomarafioti/PabloMarafioti/tree/master/analisi_de_sobrevivencia)

### Análisis de Supervivencia

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 supervivencia (Klein & Moeschberger, 2005; Machin, Cheung, & Parmar, 2006; Moore, 2016; Tableman & Kim, 2005). En los datos de supervivencia, 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  $\tau$  como 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  $\delta = I(\tau < v)$ , la cual 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-establecidos. 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 “T = 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.

Tabla 1.  
*Variables Creadas*

Variable	Descripción	Clase	Niveles
ALUMNO	Alumno	Cualitativa	-
SESIÓN	Sesión transcrita	Cualitativa	-
LINEA	Línea en la transcripción .CHA	Cuantitativa	[6, 515]
INSTANCIA	Instancia de concordancia observada	Caracteres	-
MOD	Tipo de modificador del controlador	Cualitativa	0 = artículo definido; 1 = artículo indefinido; 2 = determinante; 3 = adjetivo
LDA	Concordancia a larga distancia	Cualitativa	0 = sin larga distancia; 1 = con larga distancia
ES	Presencia de -e- epentética	Cualitativa	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	0 = dos términos; 1 = más de dos términos
IMA.CONC.f	Índice discretizado PCA	Cualitativa	0 = [-3.48, 0.58] = bajo; 1 = [0.58, 2.35] = alto
FAM.LEX	Índice discretizado PCA	Cualitativa	0 = [-4.24, 0.17] = bajo; 1 = [0.17, 1.98] = alto
ANIM	Animicidad del controlador	Cuantitativa	0 = inanimado; 1 = animado
ESP	Instancia en Español	Caracteres	-
MORF.f	Similitud entre terminaciones	Cualitativa	0 = [2.8, 2.2, 2.6, 2.4, 3] = baja 1 = [3, 3.2, 3.4] = media 2 = [3.4, 3.6] = alta
STEM.f	Similitud entre raíces léxicas	Cualitativa	0 = [1.8, 4] = baja; 1 = [4, 10.2] = alta
EST1	Estrategia 1	Cualitativa	0 = no aplica; 1 = aplica
EST2	Estrategia 2	Cualitativa	0 = no aplica; 1 = aplica
EST3	Estrategia 3	Cualitativa	0 = no aplica; 1 = aplica
EST4	Estrategia 4	Cualitativa	0 = no aplica; 1 = aplica
EST5	Estrategia 5	Cualitativa	0 = no aplica; 1 = aplica
EST6	Estrategia 6	Cualitativa	0 = no aplica; 1 = aplica
EST7	Estrategia 7	Cualitativa	0 = no aplica; 1 = aplica
Fabs.SC.f	Índice discretizado PCA	Cualitativa	0 = [-5.56, 0.45] = baja; 1 = [0.45, 1.61] = alta

### Modelo de Eventos Múltiples

La función de riesgo (o tasa de riesgo instantánea) indica la probabilidad de sobrevivir un intervalo corto de tiempo adicional sabiendo que el individuo sobrevivió hasta el tiempo  $t$ . 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):

$$HR = \frac{h(t)}{h_0(t)} = \exp(\beta_1 x_1 + \beta_2 x_2 + \dots + \beta_p x_p) = \psi.$$

Obsérvese que se ha escrito  $\psi$  como una constante que no depende del tiempo. Es decir que, no obstante  $h(t)$  y  $h_0(t)$  sí lo hagan, su ratio se mantiene igual a lo largo del tiempo. Este es el supuesto de HR proporcional. Si  $HR > 1$ , el riesgo aumenta; si  $HR < 1$ , disminuye. El modelo proporcional de Cox permite estudiar el efecto de las variables explicativas sobre la función de riesgo. En el modelo de eventos múltiples, se extiende dicho modelo a 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. En cuanto a la relación entre TYPE y TOKEN, 247 TYPES de 1152 tenían entre 2 y 32 TOKENS, y 127 solamente 2. Todas las covariables se asumieron 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. Se ajustó un modelo de Andersen y Gill estratificado por alumno (Hosmer, Lemeshow, & May, 2008). Se estima una hazard basal común para los eventos dentro de un mismo estrato  $V$  (un estrato por alumno:  $V = 1,2,3,4$ ). El modelo supone independencia entre los eventos dentro de cada individuo, lo cual 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, que ajusta por datos agrupados (varios eventos en individuos; Lin & Wei, 1989).

### 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. El modelo admite que el HR de las covariables pueda ser específico para la  $k$ -ésima causa de “muerte” (error) [( $k = 1, \dots, 4$ ; según: 1 = error de género, 2 = error de -e-epentética, 3 = error de plural, 4 = error mixto)]. 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 dicha correlación, se ajustó un modelo de “fragilidad compartida” con efectos aleatorios de ID:SESIÓN. Dicho 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». Los efectos aleatorios dan cuenta de la fragilidad del TYPE  $i$ -ésimo, que modela estos efectos de covariables no observadas. Se asume que se distribuyen como Gamma.

### Efectos Esperados para ambos Modelos

Se desea encontrar las variables que influyen en el riesgo de que una concordancia tenga un error ( $\delta = 1$ ) de modo recurrente (modelo de eventos múltiples); o bien en el riesgo de que una concordancia tenga un error de un tipo determinado (modelo de riesgos competitivos). 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. Se espera que el riesgo aumente con: (i) artículos indefinidos, determinantes y adjetivos; (ii) concordancias a larga distancia; (iii) -e-epentética en un término o en ambos; (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) 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.

## Resultados

### Modelo de Eventos Múltiples

Se ajustó un modelo de Andersen y Gill con varianza “sándwich” por TYPE y estratos por

Tabla 2.

Modelo de eventos múltiples

	Co.	exp.co.	se.co.	rob.se	Z	Pr(z)	Lo95	Up95
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

Nota: Co.: betas estimados; exp(co.): hazard ratios; se(co.): error típico de betas estimados; rob.se: error típico de beta con estimador sandwich, z: co. / rob.se; Pr(z): p-valor; lo95: extremo izquierdo de intervalo de confianza de 95 por ciento para hazard ratio; up95: 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 [tt(ANIM1)] fue significativo, confirmando la dependencia temporal del coeficiente de ANIM. Con respecto al riesgo de recurrencia de error (columna “exp.co”), 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 [similitud 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 [similitud baja] entre las terminaciones de español e italiano (MORF.f2), dicha baja no resultó significativa.

### Modelo de Eventos Competitivos

Se ajustó un modelo de “fragilidad compartida”. Siguiendo el paradigma de la información fueron seleccionadas las variables: ANIM, EST1, EST5, MORF.f, Fabs.SC.f, FAM.LEX.f, MOD. Hubo coeficientes específicos para cada tipo de error. Sin embargo, se observó

alumno; con las predictoras ANIM, EST1, EST5, MORF.f. Dichas predictoras surgieron de una selección de modelos siguiendo el paradigma de la información (Burnham & Anderson, 2010). Sin embargo, no se cumplió el supuesto de *hazards* proporcionales para la variable ANIM. 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. La Tabla 2 muestra los resultados del ajuste.

que para las variables Fabs.SC.f, FAM.LEX.f (para errores de plural) y EST1 (para error de género), no se cumplía el supuesto de *hazards* proporcionales. Para acomodar esto, se ajustó el modelo estratificando para estas variables según  $V = \{1: t \leq 300; 2: t > 300\}$ . Es decir, para dichas variables hubo un coeficiente específico para cada intervalo de datos.

La Tabla 3 expresa los *hazard* ratios ( $HR = \exp(\beta)$ ) para cada coeficiente, 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). Las variables aparecen codificadas como: VARIABLE (NIVEL).TIPO DE ERROR:GRUPO={1,2}, donde GRUPO indicaba la estratificación por intervalo de datos, cuando se la aplicaba. Se reportan solamente las variables que resultaron significativas.

El riesgo de error sube para los adjetivos MOD(3) [error plural y mixto], los determinantes MOD(2) [error plural y mixto] y los controladores animados ANIM(1) [error de -e- epentética] (pero obsérvese el largo intervalo de confianza, que hace que la estimación sea incierta), que son



Tabla 3.  
Modelo de riesgos competitivos

	exp.coef.	lower95	upper95	Size	Dir
Fabs.SC.f(1).1	0.508	0.293	0.882	49.181	DOWN
MORF.f(1).2	0.361	0.178	0.730	63.911	DOWN
MORF.f(1).3	0.433	0.310	0.606	56.682	DOWN
MORF.f(2).3	0.384	0.258	0.572	61.624	DOWN
MOD(2).3	1.498	1.104	2.032	49.773	UP
MOD(2).4	3.043	1.804	5.134	204.283	UP
MOD(3).3	1.652	1.193	2.287	65.179	UP
MOD(3).4	2.697	1.588	4.580	169.674	UP
ANIM(1).2	4.657	2.244	9.665	365.698	UP
FAM.LEX.f(1).2	0.311	0.142	0.682	68.859	DOWN
FAM.LEX.f(1).4	0.603	0.391	0.931	39.704	DOWN
EST1(1).4	0.321	0.176	0.584	67.919	DOWN
EST5(1).4	0.413	0.196	0.870	58.692	DOWN
Fabs.SC.f(1).3:group=2	0.491	0.314	0.767	50.909	DOWN
FAM.LEX.f(1).3:group=2	0.621	0.427	0.904	37.867	DOWN
EST1(1).1:group=1	0.049	0.007	0.358	95.069	DOWN

Nota. 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.

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 en las concordancias con terminaciones de similitud 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 similitud media [MORF.f(1)] y baja [MORF.f(2)] con el italiano, 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]. El riesgo de error de plural sube con los determinantes [MOD(2)] y los adjetivos [MOD(3)]. En lo referente al error mixto, sube el riesgo en presencia de determinantes [MOD(2)] y adjetivos [MOD(3)]; y baja con controladores familiares y/o frecuentes [FAM.LEX.f(1)], y los contextos donde es aplicable primera [EST1(1)] y la quinta [EST5(1)] estrategia.

## Discusión

La [Tabla 4](#) compara los resultados de ambos modelos con las expectativas del análisis. Las casillas vacías indican efectos no significativos. 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 efecto, el riesgo de cometer error subió para los errores mixtos y de plural en determinantes y adjetivos (modelo de riesgos competitivos). En coincidencia con la literatura previa, se encontró también un efecto de animidad del controlador. Es decir, los controladores animados hicieron subir el riesgo de error tanto en el modelo de eventos múltiples como en el de riesgos competitivos. Asimismo, a mayor familiaridad y /o frecuencia del controlador disminuyó el riesgo para los errores de -e- epentética, mixtos y plural ( $t > 300$ ). No hubo efecto de concreción y/o imaginabilidad del controlador. Haber encontrado dichos efectos para el controlador muestra que es necesario incluir más características de este, además de la animidad. Otro factor crucial

Tabla 4.  
Expectativas del análisis y resultados

Predictor	Efecto	Dirección esperada	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)	- ↓		

Nota. (↑)= riesgo aumenta; (↓)= riesgo disminuye

reportado para la concordancia parece ser la distancia (lineal o estructural) entre controlador y objetivo. No obstante, 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. En primer lugar, EST1: si la palabra plural del italiano termina en *-i* poner en español plural en *-os* [*libr-i* > *libr-os*]. Dicha estrategia arrojó un efecto significativo en el modelo de eventos múltiples y en el de riesgos competitivos para errores de género ( $t < 300$ ) y mixto. Por otro lado, los resultados indicaron que siempre bajó el 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. Estos hallazgos sugieren que la estrategia 1 facilita las concordancias correctas. En segundo lugar, EST5: si la palabra singular del italiano termina en *-e*, poner en español el plural en *-es*; p. ej. *camion-e* > *camion-es* [casos en los cuales el español coincide con la aplicación del plural con *-e*-epentética]. En este caso siempre disminuyó el 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 de eventos múltiples y en el de riesgos competitivos para errores mixtos. Este resultado sugiere 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. Es decir, dado que 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. Es posible que este factor haya influido en el hecho de no haber encontrado efecto en la variable ES, tras haber especulado que las concordancias con *-es* en algún término conllevarían mayor dificultad. En términos del “transfer”, EST1 y EST5 redundaron en una transferencia “positiva”, ayudando a que no se cometieran errores de concordancia.

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 riesgo de error para frecuencias altas. El efecto facilitador se verificó en el modelo de riesgos competitivos para errores de género y plural ( $t > 300$ ).

Por último, se encontró efecto en la dirección *inversa* a la esperada en MORF.f (riesgo bajó en lugar de subir) para ambos modelos. 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 similitud media / alta (distancias media / alta). Sucede con los errores de plural y de *-e*-epentética. Examinando los ejemplos del corpus de distancia baja (similitud alta) con errores de concordancia, se observó que muchos de los determinantes eran posesivos de tercera persona plural (“sus”) [18 instancias sobre un total de 37] y la mayoría contenían errores. 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* (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]. En suma, el efecto encontrado pudo deberse a la presencia de numerosos posesivos en la categoría de referencia de la variable. Entonces, lo que se observó mayormente en MORF.f fue un efecto de dificultad de los posesivos para formar las concordancias.

Considerando estos resultados en conjunto, puede sugerirse que el modelo de riesgos competitivos detectó mayor cantidad de efectos: (i) arrojó la misma dirección de efecto que el modelo de eventos múltiples para las variables ANIM, EST1, EST5, MORF.f(1); (ii) detectó efectos que el otro modelo no logró ver en MOD, Fabs. SC.f, MORF(2), FAM.LEX.f. La mayor familiaridad / frecuencia del controlador tuvo efecto protector en la mayoría de los errores; y la animicidad redundó en mayor cantidad de errores de concordancia (pero solamente en el caso de errores de plural). Las estrategias EST1 y EST5 bajaron el riesgo de error de concordancia, en especial para el tipo de error mixto, el más difícil

("transfer" positivo). Además, la estrategia EST5 podría estar ayudando en los casos de plurales con -e- epentética. Por otra parte, se detectó un efecto de frecuencia de TYPES de concordancia para errores de género y de plural.

En este trabajo, se intentó llevar a cabo un análisis de los posibles factores que inciden en el riesgo de cometer errores de concordancia plural usando una técnica estadística poco usada en la literatura de adquisición de ELE. Se ampliaron los factores considerados en la literatura previa, y se adoptó un enfoque más dinámico, incorporando al tiempo (hasta el error de concordancia) como variable dependiente. Entre las limitaciones de este estudio, cabe destacar que se trata de un estudio de caso observacional. Sin embargo, los hallazgos obtenidos pueden constituir hipótesis a ser evaluadas en un futuro mediante estudios experimentales que controlen los sesgos y extiendan la población de referencia. Los resultados pueden ser de utilidad para el docente que quiera planificar actividades teniendo en cuenta los factores de facilidad y dificultad hallados.

### Referencias

- Alarcón, I. (2009). The processing of gender agreement in L1 and L2 Spanish: Evidence from Reaction Time Data. *Hispania*, 92(4), 814-828.
- Alarcón, I. (2011). Spanish grammatical gender under complete and incomplete acquisition: early and late Bilinguals' linguistic behavior within the noun phrase. *Bilingualism: Language and Cognition*, 14(3), 332-350. doi: 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 & H. Goodluck (Eds.), *The role of features in second language acquisition* (pp. 270–298). Mahwah, NJ: Lawrence Erlbaum.
- Bruhn de Garavito, J., & White, L. (2002). The second language acquisition of Spanish DPs: the status of grammatical features. En A. T. Pérez-Leroux & J. Muñoz Liceras (Eds.), *The Acquisition of Spanish Morphosyntax: The L1/L2 Connection* (pp. 153–178). Dordrecht: Kluwer.
- Burnham, K. P., & Anderson, D. R. (2010). *Model selection and multimodel inference: a practical information-theoretic approach*. New York: Springer.
- Carrera Díaz, M. (1989). *Manual de Gramática Italiana*. Barcelona: Ariel.
- Corbett, G. (2006). *Agreement*. Cambridge: Cambridge University Press.
- Davis, C. J., & 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. doi: 10.3758/BF03192738
- Fernández-García, M. (1999). Patterns of gender agreement in the speech of second language learners. En J. Gutiérrez-Rexach & F. Martínez-Gil (Eds.), *Advances in hispanic linguistics: papers from the 2nd Hispanic Linguistics Symposium* (pp. 3–15). Somerville, MA: Cascadilla Press.
- Finnemann, M. D. (1992). Learning agreement in the noun phrase: the strategies of three first-year Spanish students. *International Review of Applied Linguistics in Language Teaching*, 30(2), 121-136. doi: 10.1515/iral.1992.30.2.121
- Foot, R. (2011). Integrated knowledge of agreement in early and late English-Spanish bilinguals. *Applied Linguistics*, 32(1), 187-220. doi: 10.1017/S0142716410000342
- Foot, R. (2015). The production of gender agreement in native and L2 Spanish: the role of morphophonological form. *Second Language Research*, 31(3), 343-373. doi: 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. doi: 10.1177/026765830101700301
- Gillon Dowens, M., Vergara, M., Barber, H., & Carreiras, M. (2010). Morphosyntactic processing in late second language learners. *Journal of Cognitive Neuroscience*, 22(8), 1870-1887. doi: 10.1162/jocn.2009.21304.
- González, P., Mayans, D., & Van der Bergh, H. (2022). Nominal agreement in the interlanguage of Dutch L2 learners of Spanish. *International Review of Applied Linguistics in Language Teaching*, 60(2), 363-382. doi: 10.1515/iral-2017-0174
- Hosmer, D. W., Lemeshow, S., & May, S. (2008). *Applied survival analysis: regression modeling of time to event data*. New Jersey: Wiley.
- 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. doi: 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 & J. Jegerski (Eds.), *Research in Second Language Processing and Parsing* (pp. 113–134). Philadelphia: John Benjamins.
- Kilgariff, A., Baisa, V., Jan Bušta, Jakubiček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: ten years on.

- Lexicography*, 1(1), 7-36. doi: 10.1007/s40607-014-0009-9
- Klein, J. P., & Moeschberger, M. L. (2005). *Survival analysis: techniques for censored and truncated data*. New York: Springer.
- Lichtman, K. (2009). Acquisition of Attributive and Predicative Adjective Agreement in L2 Spanish. En M. Bowles, T. Ionin, S. Montrul & A. Tremblay (Eds.), *Proceedings of the 10th Generative Approaches to Second Language Acquisition Conference* (pp. 231–247). Somerville, MA: Cascadilla Proceedings Project.
- Lin, D. Y., & Wei, L. J. (1989). The robust inference for the cox proportional hazards model. *Journal of the American Statistical Association*, 84(408), 1074-1078. doi: 10.1080/01621459.1989.10478874
- Mac Whinney, B. (2020). *The Childes Project: Tools for Analyzing Talk*. 3<sup>rd</sup> Edition. Mahwah, NJ: Lawrence Erlbaum Associates.
- Machin, D., Cheung, Y. B., & Parmar, M. (2006). *Survival Analysis: A Practical Approach*. John Wiley & Sons.
- McCarthy, C. (2008). Morphological variability in the comprehension of agreement: an argument for representation over computation. *Second Language Research*, 24(4), 459-486. doi: 10.1177/0267658308095737.
- Montrul, S., Foote, R., & 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. doi: 10.1111/j.1467-9922.2008.00449.x
- Moore, D. F. (2016). *Applied Survival Analysis Using R*. Switzerland: Springer Cham.
- Muñoz Licerias, J., Díaz Rodríguez, L., & 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 & C. Sanz (Eds.), *Current research on the acquisition of Spanish* (pp. 67–96). Somerville, MA: Cascadilla Press.
- Nerbonne, J., Van Ommen, S., Gooskens, C., & Wieling, M. (2013). Measuring socially motivated pronunciation differences. En L. Borin & A. Saxena (Eds.), *Approaches to Measuring Linguistic Differences* (pp. 107–140). Berlin / Boston: De Gruyter Mouton. doi: 10.1515/9783110305258
- O’Grady, W. (2005). *Syntactic Carpentry: An Emergentist Approach to Syntax*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Oakes, M. P. (1998). *Statistics for Corpus Linguistics*. Edinburgh University Press.
- Peña, D. (2002). *Análisis de Datos Multivariantes*. Madrid: 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 & M.-J. Cabrera (Eds.), *Current Issues in Linguistic Theory Series* (pp. 240–253). Amsterdam: John Benjamins. doi: 10.1075/cilt.287.18sag
- Sagarra, N., & Herschensohn, J. (2013). Processing of gender and number agreement in late Spanish bilinguals. *International Journal of Bilingualism*, 17(5), 607-627. doi: 10.1177/1367006912453810.
- Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). Mclust 5: clustering, classification and density estimation using gaussian finite mixture models. *The R Journal*, 8(1), 289-317. doi: 10.32614/RJ-2016-021
- Tableman, M., & Kim, J. S. (2005). *Survival Analysis Using S: Analysis of Time-to-Event Data*. Londres: Chapman & Hall/CRC.
- Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). Mice: multivariate imputation by chained equations in R. *Journal of Statistical Software*, 45(3), 1-67. doi: 10.18637/jss.v045.i03
- White, L., Valenzuela, E., Kozłowska-Macgregor, M., & Leung, Y. (2004). Gender and number agreement in nonnative Spanish. *Applied Psycholinguistics*, 25(1), 105–133. doi: 10.1017/S0142716404001067