

ANÁLISIS DE LA EVOLUCIÓN DE ERRORES DE CONCORDANCIA EN CUATRO APRENDIENTES ITALIANOS DE ELE USANDO REDES COMPLEJAS

Pablo Ezequiel Marafioti
Universidad Nacional de Córdoba (Argentina)
kenavo38@yahoo.com

Recibido: 11/08/2020 - **Aprobado:** 26/09/2020

DOI: doi.org/10.17533/udea.lyl.n79a10

Resumen: En el siguiente estudio observacional de casos se identificaron algunos factores influyentes en la producción de concordancia plural en cuatro aprendientes italianos de español como lengua extranjera (ELE). Para ello se creó un corpus de datos longitudinal a partir de un conjunto de entrevistas orales y se construyeron redes complejas en las que cada palabra/nodo tuviera un enlace con otra si se hallaba una relación de concordancia errónea entre ellas. Además, se analizaron los efectos estadísticos temporales en redes de acumulación de errores y se encontraron efectos para algunos factores.

Palabras clave: Español como Lengua Extranjera (ELE); análisis de errores; concordancia nominal; redes complejas; plural.

EVOLUTION ANALYSIS OF AGREEMENT ERRORS IN FOUR ITALIAN LEARNERS OF SPANISH L2 USING COMPLEX NETWORKS

Abstract: In the following observational case study, some influencing factors were identified in the production of plural agreement in four Italian learners of Spanish as a foreign language (ELE). To do this, a longitudinal data corpus was created from a set of oral interviews and complex networks were built in which each word / node had a link with another if an erroneous concordance relationship was found between them. In addition, the temporal statistical effects were examined in networks with cumulative errors and some factors were found as statistically significant.

Key words: Spanish as a Foreign Language; error analysis; nominal agreement; complex networks; plural.

1. Introducción

1.1. La concordancia: definición y antecedentes en adquisición de ELE

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) denominó *controlador* al ítem léxico que determina la concordancia y *objetivo* al elemento cuya forma es determinada por aquel. Además, definió al *dominio* como el entorno sintáctico en el cual ocurre la concordancia. Esta se establece por una covarianza sistemática de rasgos. En este trabajo el controlador será nominal, presente y expresará sus rasgos abiertamente. Por otra parte, los objetivos consistirán en artículos (definidos e indefinidos), adjetivos y pronombres (demostrativos, posesivos, indefinidos); además, concordarán 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—.

La literatura de adquisición ha arrojado los siguientes resultados en lo que respecta a la producción. Fernández-García (1999) investigó la concordancia de género en el sintagma nominal en siete aprendientes ingleses de nivel intermedio a partir de entrevistas grabadas. En general, hubo menos errores en contextos masculinos. Se detectó, en primer lugar, que el masculino se usó mayormente como *default* en artículos indefinidos y adjetivos. Además, se verificó más precisión con artículos que con adjetivos. Los sustantivos que marcaban las clases en *-o /-a (libro)* lograron mayor precisión que los no marcados (*flor*). También existieron menos errores para aquellos motivados semánticamente en el sexo biológico. Asimismo, los alumnos produjeron menos errores con los sustantivos animados (*tío*) que con los inanimados (*tenedor*). Por su parte, Licerias, Díaz y Mongeon (2000) compararon la adquisición de género y número en L1 y L2 en el contexto del sintagma nominal y de elipsis (*las faldas negras / las negras*) en cuatro niños con (edades entre los 4 y 8 años). Los errores de concordancia de género (*el zapatos; el cama*) en L1 y L2 tenían porcentajes muy similares, desapareciendo eventualmente en L1 y persistiendo en L2. En cuanto al número, los errores en L2 eran siempre menos que los de género y eventualmente desaparecían, como sucede en L1. En otro estudio, Francheschina (2001) siguió a un hablante de inglés que había alcanzado nivel superior. El participante exhibió un alto grado de precisión en la concordancia de artículos, adjetivos, demostrativos y pronombres. Sin embargo, los errores de género eran siempre más que los de número, concentrándose estos últimos en los adjetivos. Asimismo, el sujeto observado extendió más veces el uso del masculino a contextos femeninos que viceversa en adjetivos y artículos. Entretanto, Bruhn de Garabito y White (2002) estudiaron el género en estudiantes de secundaria cuya lengua materna era el francés, divididos en dos grupos según estuvieran terminando el primer o el segundo año de español. Aunque la precisión en general fue alta, esta resultó mejor en los artículos que en los adjetivos. Sumado a esto, se verificaron más errores en el artículo definido que en el indefinido. Dichos errores fueron mayores con controladores de género motivados

semánticamente que con aquellos motivados morfológicamente —sugiriendo que los alumnos estaban prestando atención a la morfología—. Nuevamente, hubo más sobreextensiones del masculino a contextos femeninos, es decir, una preferencia de masculino como *default*. Asimismo, White *et al.* (2004) investigaron la producción y la comprensión en la concordancia de género y número con estudiantes franceses e ingleses de ELE, divididos a su vez en grupos de nivel inicial, medio y superior. En lo que atañe a la producción, tanto tomando en cuenta bigramas *Artículo + Nombre*, como trigramas *Art. + N + Adj.*, la precisión en general fue alta. Se volvió a observar mayor facilidad para adquirir la concordancia del artículo respecto de los adjetivos por un lado y más precisión en el número respecto al género, por el otro. También se notó nuevamente la tendencia a utilizar el masculino como *default*. En cuanto a la incurrencia en errores, se constató que desaparecerían a medida que subía el nivel. En suma, pues, se detectó un efecto de competencia —la precisión difería significativamente de acuerdo con el nivel— y uno de rasgo —diferencia significativa en la producción de (valores de) rasgos correctos requeridos por el contexto, por ejemplo, masculino correcto en el adjetivo si el controlador era masculino—. Sin embargo, no hubo un influjo de la lengua materna, dado que los resultados de los aprendientes ingleses no diferían significativamente de los obtenidos por los franceses. Por otra parte, Echevarría & Prévost (2004) investigaron la concordancia de plural en los adjetivos atributivos (pronominales) y en los cuantificadores *mucho*, *bastante* y *demasiado*, para alumnos de francés como lengua materna de nivel intermedio y superior. En lo que atañe a los adjetivos, la concordancia obtuvo altos niveles de precisión. La precisión en *mucho* resultó muy alta, por lo que en la tarea de producción se observó el orden *mucho > demasiado > bastante* en los niveles de corrección. Posteriormente, Bruhn de Garavito (2008) analizó el plural en hablantes de francés L1 divididos en dos grupos de edades entre los 14 y 17 años pertenecientes, respectivamente, al nivel 1 y 2 de ELE. El investigador encontró que la adquisición del plural seguía las etapas plural nulo > plural en /-s/ > plural en /-es/. En otro estudio, Montrul *et al.* (2008) se enfocaron en las diferencias de la concordancia de género entre estudiantes de ELE de habla inglesa (grupo 1) y aquellos nacidos en EE. UU. hijos de padres mexicanos expuestos al inglés antes de los cinco años (grupo 2). El grupo de control estaba conformado por hablantes nativos de inglés. En la tarea de producción de sintagmas nominales [*Art. + N + Adj.*] apareció un efecto de grupo, obteniendo el grupo 2 mejores resultados que el 1. Se verificó también un efecto de animidad: mayor precisión en los sustantivos transparentes (-o / -a como en *vaso*, *maestra*) animados que en los inanimados. También se halló un efecto de morfología: los sustantivos terminados en consonante (*el camión*, *la canción*) con mejores resultados que los terminados en -e (*el puente*, *la suerte*); además, los opuestos (*la mano*) fueron los que se erraron más, especialmente con los adjetivos. Asimismo, se observó un efecto de valor de rasgo: masculino mejor que femenino; pero ningún efecto de modificador (artículo / adjetivo). De modo similar, Alarcón (2011) replicó la tarea de producción del estudio anterior con estudiantes de L2 ingleses e hijos de inmigrantes nacidos en EE. UU. expuestos al español antes de los 5 años. Ambos grupos tenían una competencia similar de nivel superior; durante la investigación, participaron hablantes nativos incluidos dentro de un grupo de control. Se halló significancia en los siguientes factores: a) grupo: mejor en los hijos de inmigrantes que en los estudiantes de L2; b) rasgo: mejor en el masculino que en el femenino; c) morfología: mejor en sustantivos transparentes que

en menos transparentes; d) rasgo*grupo: la diferencia en precisión entre masculino y femenino fue mayor en el caso de los aprendientes ingleses; e) rasgo*morfología: la diferencia en precisión entre nombres más y menos transparentes resultó mayor para el femenino; f) rasgo*modificador: la precisión fue significativamente mayor para los artículos que para los adjetivos en el caso del masculino; g) grupo*morfología: ambos grupos con mejores resultados en sustantivos transparentes que menos transparentes pero, al igual que en d), con diferencia mayor entre los aprendientes ingleses. Recientemente, Foote (2015) investigó el rol de la transparencia morfológica del controlador en la concordancia de género (singular), con aprendientes ingleses de ELE de nivel intermedio y superior. En el experimento se utilizaba una tarea de completamiento del tipo: *la pared (N1) del baño (N2) está... limpia/*o*. Las variables manipuladas fueron: i) género: el género del núcleo N1 del SN (masculino / femenino); ii) morfología: la transparencia de marca de género en el N1 (transparente / opaca); iii) correspondencia: la igualdad o desigualdad entre los géneros del N1 y el N2 (igual / distinto); iv) grupo: el nivel de competencia (intermedio / superior / nativo). Los resultados arrojaron una interacción significativa grupo*correspondencia: los aprendientes de nivel intermedio cometían significativamente más errores que los de nivel superior; y estos, más que los nativos. Además, hubo más errores en la condición de no igualdad de género entre N1 y N2, por lo que no hubo un efecto de transparencia morfológica ni de género.

1.2. Objetivos

El objetivo general de este trabajo fue identificar algunos factores influyentes en la producción de concordancia plural en cuatro aprendientes italianos de ELE. Para ello, se utilizó el marco de las redes complejas (Newman, 2010), un método poco aplicado en adquisición de L2. Para cada alumno, se crearon redes «de amistad» en las cuales las palabras / actores establecen enlaces / relaciones de concordancia erróneas con otras palabras / actores. A partir de este planteamiento, se intentó encontrar efectos respecto de los atributos de las palabras que influyeran en la aparición o desaparición de enlaces de concordancia con error en la evolución temporal de la adquisición¹.

2. Recolección de datos

En esta sección se analizarán los datos obtenidos de cuatro casos de estudiantes adultos, de lengua nativa italiana, estudiantes de ELE en el Instituto Cervantes de Milán durante el año académico 2008-2009. Cada alumno poseía un nivel distinto de competencia lingüística según el Marco Común Europeo de Referencia. Inicialmente, se hicieron entrevistas de 30 minutos entre el alumno y el investigador autor de este trabajo. La tarea consistió en sostener conversaciones no estructuradas sobre temas acordes al nivel de competencia de los participantes. Dichas entrevistas tuvieron lugar durante aproximadamente cada 20 días, según la disponibilidad de los alumnos.

1. El código de R empleado para el análisis se puede encontrar en el siguiente enlace: https://github.com/pablomarafioti/PabloMarafioti/tree/master/redes_complejas/R%20script

Cada aprendiente realizaba simultáneamente el curso de español que le correspondía. Durante el tiempo de la investigación, hubo entre doce y catorce entrevistas con cada sujeto. De este modo, el *corpus* está constituido por los siguientes conjuntos de transcripciones: SONIA (nivel A1/A2), con 12 transcripciones; NATI (nivel B1), con 14 transcripciones; JAKO (nivel B2), con 14 transcripciones; y MIRKA (nivel C1), con 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, aunque era posible que hubiese más términos «objetivo»: por ejemplo, en *los libros azules* se codificaron dos instancias: *los libros* y *libros azules*. Para la realización del conteo, se utilizó el programa CLAN. A modo de ilustración, 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 se si me explico.
52 *STU: los [*0] seres humanos [*0] ehh@fp tenemos +...
53 *STU: es algo un poco malo que decir.
54 *STU: pero tenemos [*] más derechos que los animales.
55 %err: tenemos
56 *STU: en el sentido que [*] ehh@fp <me>[/] me doy cuenta que a veces los animales [*0] son muy sensibles [*0].

3. Redes de concordancia y atributos

Se crearon cuatro redes acumulando los casos de concordancia con error de las sesiones. Las redes eran dirigidas, donde la flecha indicaba el origen del error (Ej.: *la casas* recibiría la representación, mostrando que el error se ubica en el artículo). Si el error estaba en ambos términos, entonces la flecha era bidireccional. Así, se tomaron solamente los casos de concordancia de error, es decir, se trataba de redes de crecimiento de error. Para ello, se agruparon las sesiones en tres etapas (de crecimiento acumulado) según el siguiente esquema:

1. SONIA y MIRKA: i) hasta la sesión 4; ii) hasta la sesión 8; iii) hasta la sesión 12.
2. NATI y JAKO: i) hasta la sesión 5; ii) hasta la sesión 9; iii) hasta la sesión 14.

Nótese que las redes mantienen constantes los nodos / palabras y lo que se agrega en cada etapa son las aristas. La Figura 1 muestra la red de SONIA (hasta la sesión 12). Por otro lado, la Figura 2 ilustra las redes de todos los aprendientes.

Por otra parte, se crearon variables que caracterizaban cada término dentro de una concordancia (nodo en la red), las cuales son descritas a continuación (el primer nivel se considera el de referencia):

—ERROR. Cantidad de errores de concordancia en los que interviene el término. Por ejemplo, el plural *profesores* puede estar involucrado en las concordancias producidas: *los profesores*, *los profesore*, *mucho profesores*; con lo cual estaría involucrada en dos concordancias erróneas.

—ESPA. Concordancia en español (sin error).

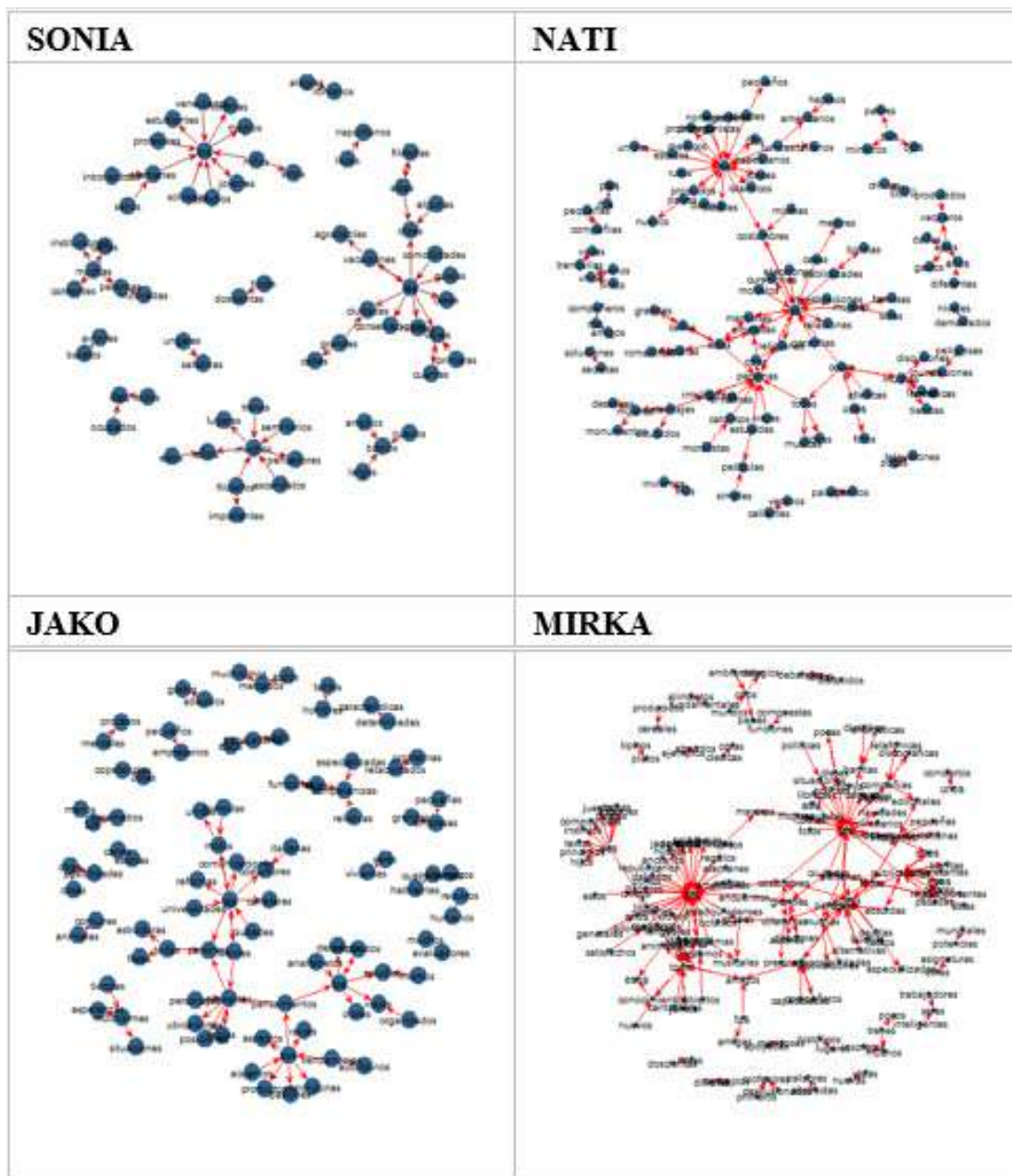


Figura 2. *Redes de crecimiento de error de los cuatro estudiantes de ELE*

—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, según: 0 = no es controlador; 1 = inanimado, 2 = animado.

—ESP1. Se codificó la vocal final de la palabra más la desinencia de plural en español del primer término. Según: 1 = *us* [Ej.: *sus*]; 2 = *is* [Ej.: *mis*]; 3 = *os* [ej.: *rojos*]; 4 = *as* [ej.: *blancas*]; 5 = *es* [ej.: *azules*]; 6 = **es* [ej.: útiles (-e- epentética)]

—ESP2. Se codificó la vocal final de la palabra más la desinencia de plural en español del segundo término (mismos niveles que ESP1).

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

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

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 del español y el italiano; y entre los morfemas de género y número plural. El algoritmo de Levenstein calcula la distancia entre dos secuencias de caracteres como el número mínimo de operaciones necesarias para transformar una secuencia en la otra. Estas operaciones son: DELETE (borrar); SUBSTITUTE (sustituir); INSERT (insertar). A partir de esto, se asignaron los siguientes pesos para las operaciones: i) DELETE = 0.3; ii) SUBSTITUTE = 0.6; INSERT = 1. Por ejemplo, en la instancia *much-os amig-os* [SONIA, sesión 1, línea 7] (it. *molt-i amic-i*) la similitud entre las terminaciones de ambos términos de la concordancia se calculó como: . En ambos términos es necesario operar una sustitución (= 0.6) de *i* por *o* y una inserción de *s* (= 1), cuya suma es 1.6. También se practicó una corrección sumando 0.2 al valor arrojado por el algoritmo en los casos en los que interviene una *-e-* epentética. Esto es así porque el algoritmo no distingue entre una terminación *-es* donde la *e* es epentética [*canciones*] y otra en que no lo es [*calles*]. A modo de ejemplo, en el caso de *las lecciones* [SONIA, sesión 8, línea 370] (it. *le lezioni*) donde el segundo término tiene *-e-* epentética, se calculó:.

Además de la similitud entre las desinencias, se calculó la similitud entre las raíces léxicas de ambos términos. Para ello, se extrajo la raíz de cada uno de los términos y se aplicó la distancia de Levenstein. Por ejemplo, en *much-os person-as* [SONIA, sesión 1, línea 56] (it. *molt-e person-e*) se calculó como: .

Por otra parte, se realizaron las siguientes operaciones de pre-procesamiento. En primer lugar, se transformó al logaritmo la frecuencia del controlador, sumándole una unidad (Corpus “BuscaPalabras”): .Como resultado de estas operaciones, se recolectaron en total 1857 casos de concordancia. 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. Para tal efecto, se utilizó el paquete *mice* [*Multivariate Imputation by Chained Equations*] de (Van Buuren & Groothuis-Oudshoorn, 2011), que realiza imputación múltiple.

Por otra parte, se empleó la correlación de Spearman entre las variables extraídas de BuscaPalabras: Imaginabilidad (IMA), Concretud (CONC), Familiaridad (FAM)y Frecuencia en escala logarítmica (LEXESP). Las variables fueron escaladas en la ecuación (). Una vez realizadas, se observó una correlación positiva moderada

(0.68) entre Imaginabilidad y Concretud, como era de suponerse: entre más concreta resulte una palabra, más fácil resulta evocar imágenes de esta. Además, existió una correlación positiva entre Familiaridad y Frecuencia (0.50): entre más frecuente resulta una palabra, más familiar parece. Se aplicó, pues, un Análisis de Componentes Principales (PCA, por sus siglas en inglés). La técnica permite obtener nuevas variables ortogonales llamadas componentes principales, que se calculan como combinación lineal de las variables cuantitativas originales (Peña, 2002). En la primera componente, denominada IMA.CONC, cargaban las variables Imaginabilidad y Concretud; y en la segunda, Familiaridad y LEXESP, denominada FAM.LEX. En suma, se resolvió la colinealidad pasando de cuatro variables de controlador con correlación a dos componentes sin correlación entre sí.

Por último, se decidió discretizar los atributos cuantitativos utilizando *clustering* por mezcla de gaussianas. Para ello, se utilizó el paquete *mclust* de R (Strucca *et al.*, 2016) con el fin de hacer el agrupamiento. Durante el proceso, se discretizaron los atributos: IMA.CONC y FAM.LEX. El atributo FAM.LEX se discretizó en dos categorías poniendo como punto de corte a la mediana; ya que el *clustering* no resultó efectivo. A continuación, en la Tabla 1 se ilustran los resultados obtenidos, mientras que en la Tabla 2 se describen los atributos de los nodos.

Atributo	Descripción	Discretización	Ejemplos del Corpus
IMA.CONC.f	PCA1	0=[-3.48;0.58]	<i>nuevos conocimientos</i> (-3.11); <i>los servicios</i> (-1.19)
		1=[0.58;2.35]	
			<i>muchas personas</i> (1.31); <i>los hospitales</i> (2.17)
FAM.LEX.f	PCA2	0=[-4.24;0.17]	<i>los sultanes</i> (-4.42); <i>las comodidades</i> (-0.64)
		1=[0.17;1.98]	
[corte: mediana]			<i>los años</i> (1.22); <i>los hombres</i> (1.89)

Tabla 1. *Discretización de atributos utilizando clustering por mixtura de gaussianas*

Variable	Descripción	Clase	Niveles
MOD	Tipo de modificador del controlador	Cualitativa	0 = no es modificador, 1 = artículo definido; 2 = artículo indefinido;
			3 = determinante; 4 = adjetivo
ERROR	Cantidad de errores de concordancia en los que interviene la palabra.	Cuantitativa	[0, 1, 2...]
ES	¿Tiene la palabra plural con <i>e</i> epentética?	Cualitativa	0 = NO, 1 = SI
ANIM	¿Es la palabra <i>controlador</i> animado?	Cualitativa	0 = no es controlador; 1 = controlador inanimado;
			2 = controlador animado.
FAM.LEX	Índice PCA de familiaridad y frecuencia léxica del controlador (discretizado)	Cualitativa	0 = no es controlador; 1 = baja; 0 = alta
IMA.CONC	Índice PCA de imaginabilidad y concretud (discretizado)	Cualitativa	0 = no es controlador; 1 = baja; 0 = alta

	Léxica del controlador	Cualitativa	
ESP1	Tipos de terminaciones del «objetivo»	Cualitativa	1 = <i>us</i> [ej.: <i>sus</i>]; 2 = <i>is</i> [ej.: <i>mis</i>]; 3 = <i>os</i> [ej.: <i>rojos</i>];
			4 = <i>as</i> [ej.: <i>blancas</i>]; 5 = <i>es</i> [ej.: <i>azules</i>];
			6 = <i>es*</i> [ej.: <i>útiles</i> (-e- epentética)]
ESP2	Tipos de terminaciones del «controlador»	Cualitativa	1 = <i>us</i> [ej.: <i>sus</i>]; 2 = <i>is</i> [ej.: <i>mis</i>]; 3 = <i>os</i> [ej.: <i>rojos</i>];
			4 = <i>as</i> [ej.: <i>blancas</i>]; 5 = <i>es</i> [ej.: <i>azules</i>];
			6 = <i>es*</i> [ej.: <i>útiles</i> (-e- epentética)]
MORF	Similitud de la palabra entre terminaciones de español e italiano	Cuantitativa	0.6, 2]
STEM	Similitud de la palabra entre las raíces léxicas de español e italiano	Cuantitativa	[0.7, 6]

Tabla 2. Atributos de los nodos formulados para el análisis

4. Análisis estadístico de evolución temporal

En este punto, se aplicó el llamado *Actor-Based Stochastic Model* (Snijders, 1996; Snijders *et al.*, 2007, 2010), un modelo estadístico de evolución temporal de redes. El modelo trata a cada nodo de una red como un actor. Ya que los enlaces de la red se modelan como si fueran estados, se considera el cambio como un proceso de Markov —en cada punto del tiempo el estado actual de la red determina probabilísticamente el estado de la red en el instante sucesivo, sin influencia de estados pasados—. En este modelo, el proceso es continuo, es decir, la red observada en cada punto es el resultado de microcambios continuos realizados en minipasos. En cada uno de estos, dicho actor debe decidir si preserva, elimina o establece enlaces con otros actores. Este cambio se halla influenciado por factores endógenos a la red —por ejemplo, la tendencia a la reciprocidad de los enlaces: *si soy tu amigo, eres mi amigo*—; o exógenos a ésta: covariables del actor —los atributos del nodo— o covariables diádicas —interacciones entre atributos de dos nodos conectados por un enlace, como presencia de la característica en ambos nodos—. Además, en la evolución temporal pueden cambiar características de los actores influenciados por la estructura de la red.² Por lo tanto, se pueden seguir dos dinámicas: una de la red y otra de características comportamentales de los actores. Las relaciones entre los actores están representadas en la matriz de adyacencia

$$X_{n \times n} = (x_{ij}) \quad (n \text{ actores}); \text{ con } x_{ij} = 1 \text{ si hay un enlace entre los actores } i \text{ y } j, \text{ y } x_{ij} = 0 \text{ si no lo hay.}^3$$

Los atributos de comportamiento de los actores deben ser categóricos u ordinales y se los denotará como z_{hi} ($h = 1, \dots, H$), el valor del h -ésimo atributo del i -ésimo actor. Por lo tanto, cada actor tendrá las observaciones $Y(t) = (x_{ij}(t), z_{1i}(t), \dots, z_{Hi}(t))$, en cada tiempo discreto. El proceso de cambio se descompone en dos subprocesos.

2. Un estudiante que fuma puede ser amigo de otro que fuma [homofilia]; o bien no fumar inicialmente, pero comenzar a hacerlo si la mayoría de su grupo de amigos es fumador [asimilación].

3. Las redes con pesos categóricos se tratan como modelos con redes separadas.

El primero modela la frecuencia con la que suceden los cambios (de enlaces y del comportamiento de los actores) en los minipasos. Es una función de tasa de arribo que depende del efecto de la tasa del periodo, de efectos de covariables del actor y de efectos de posición del actor en la red (por ejemplo, la cantidad de enlaces de un nodo). Son:

$$\lambda_i^X(Y, m) = \rho_m^X \left\{ \exp \left(\sum_k \alpha_k^X a[x]_{ki}(Y(t)) \right) \right\}$$

$$\lambda_i^{Z_h}(Y, m) = \rho_m^{Z_h} \left\{ \exp \left(\sum_k \alpha_k^{Z_h} a[Z_h]_{ki}(Y(t)) \right) \right\}$$

Por tanto, los tiempos de espera entre minipasos de cada actor sigue una distribución exponencial de parámetro (). Si hay tasa constante: , . Por otra parte, se modela el subproceso de cambio determinístico; es decir, en cuál enlace o atributo comportamental se produce el cambio. Los atributos de comportamiento cambian siempre en una unidad —crecen o decrecen—. La función se descompone en tres partes: una función de evaluación, una función de satisfacción y un término de error (modelado a partir de la distribución de Gumbel): . La segunda función penaliza la pérdida de un enlace o el decrecimiento de un valor comportamental como pérdida de satisfacción. Sin embargo, no será usado dicho componente en el análisis. La parte fundamental es, pues, la función de evaluación, que especifica los efectos de estructura o de comportamiento. En este punto, el actor seleccionará el cambio que maximiza dicha función. Al igual que con la función de tasa de arribo, hay dos; una para cada dinámica:

$$f_i^X(\beta, x) = \sum_{k=1}^n \beta_k^X s_{ik}^X(x)$$

$$f_i^{Z_h}(\beta, z) = \sum_{k=1}^n \beta_k^{Z_h} s_{ik}^{Z_h}(x, z)$$

En donde son parámetros, son funciones que dependen de la estructura de la red y son funciones que dependen del comportamiento del actor, pero también del comportamiento o características estructurales (de red) de sus vecinos.

Por fin, la probabilidad de que un actor cambie de estado (enlace o comportamental) se define como:

$$P_{ij}(x(i \rightarrow j), x(t), z(t)) = \frac{\exp\{[f + g]_i^X(\beta^X, \gamma^X, x(i \rightarrow j)(t))\}}{\sum_k \exp\{[f + g]_i^X(\beta^X, \gamma^X, x(i \rightarrow k)(t))\}}$$

$$P_{i\delta}(z(i \downarrow_h \delta) | x(t), z(t)) = \frac{\exp\{[f + g]_i^{Z_h}(\beta^{Z_h}, \gamma^{Z_h}, z(i \downarrow_h \delta)(t))\}}{\sum_{\tau \in \{-1, 0, 1\}} \exp\{[f + g]_i^{Z_h}(\beta^{Z_h}, \gamma^{Z_h}, z(i \downarrow_h \tau)(t))\}}$$

A continuación, se describen los efectos de estructura incluidos en el modelo elegido para este estudio.

—*Outdegree-Popularity*. Es el número de enlaces salientes de los nodos conectados a un nodo i -ésimo.

$$s_{i1}(x) = \sum_j x_{ij} x_{j+}; x_{j+} = \sum_h x_{jh}.$$

—*Indegree-Popularity*. Es el número de enlaces entrantes de los nodos conectados a un nodo i -ésimo.

$$s_{i2}(x) = \sum_j x_{ij} x_{+j}; x_{+j} = \sum_h x_{hj}.$$

—*Reciprocity*. Es el número de enlaces recíprocos entre dos nodos.

$$s_{i3}(x) = \sum_j x_{ij} x_{ji}.$$

—*Anti isolates*. Efecto de conexión con otros nodos, los cuales estarían aislados sin dicha conexión.

$$s_{i4}(x) = \sum_j I\{x_{+j} \geq 1, x_{j+} = 0\}.$$

—*Isolate - popularity*. Efecto de estar conectado con actores cuyo grado de enlaces entrantes es 1.

$$s_{i5}(x) = \sum_j x_{ij} I\{x_{+j} = 1, x_{j+} = 0\}.$$

—*out-in degree^{1/2} assortativity*. La tendencia de que los nodos con alto grado de enlaces salientes estén conectados a nodos con alto grado de enlaces entrantes.

$$s_{i6}(x) = \sum_j x_{ij} x_{i+}^{1/2} x_{+j}^{1/2}.$$

—*Balance*. Similitud entre los enlaces salientes del nodo i -ésimo y los enlaces salientes de los otros nodos con los cuales se conecta dicho nodo i -ésimo.

$$s_{i7}(x) = \sum_{j=1}^n x_{ij} d_{ij}; d_{ij} = \sum_{h=1; h \neq i, j}^n (b_0 - |x_{ih} - x_{jh}|).$$

—*In-struct equivalence*. Similitud entre los enlaces entrantes del nodo i -ésimo y los enlaces entrantes de los otros nodos con los cuales se conecta dicho nodo i -ésimo.

$$s_{i8}(x) = \sum_{i=1}^n x_{ij} d_{ij}; d_{ij} = \sum_{h=1; h \neq i, j}^n (b_0 - |x_{hi} - x_{hj}|).$$

—*SameX*. Mismo comportamiento entre los actores (nodos) y [homofilia respecto de atributos cualitativos].

$$s_{i9}(x) = \sum_j x_{ij} I\{v_i = v_j\}.$$

—*sameXRecip*. Mismo comportamiento (homofilia) en enlaces bidireccionales.

$$s_{i10}(x) = \sum_j x_{ij} x_{ji} I\{v_i = v_j\}.$$

—*SimX*. Suma de las similitudes centradas (respecto de un determinado atributo v : sim_{ij}^v) entre i y los otros actores j con los que está conectado [Asortatividad para atributos cuantitativos].

$$s_{i11}(x) = \sum_j x_{ij} (sim_{ij}^v - \widehat{sim}^v).$$

—*SimRecipX*. Similitud respecto de enlaces bidireccionales.

$$s_{i12}(x) = \sum_j x_{ij} x_{ji} (sim_{ij}^v - \widehat{sim}^v).$$

—*EgoX*. El número de enlaces salientes pesados por un valor de covariable del actor.

$$s_{i13}(x) = v_i x_{i+}$$

—*AltX*. La suma de la covariable sobre todos los actores con los cuales tiene un enlace.

$$s_{i14}(x) = \sum_j x_{ij} v_j$$

—*EffFrom*. Dependencia de un atributo de otras variables de comportamiento

$$h' (h \neq h'). s_{i1}(x, z_h) = z_{ih} z_{ih'}$$

Entonces, las funciones de evaluación para la dinámica de la red y la del comportamiento se definen como:

$$f_i^X(\beta, x) = \exp \{ \beta_1 s_{i1}(x) + \beta_2 s_{i2}(x) + \beta_3 s_{i3}(x) + \beta_4 s_{i4}(x) + \beta_5 s_{i5}(x) + \beta_6 s_{i6}(x) + \beta_7 s_{i7}(x) + \beta_8 s_{i8}(x) + \left(\sum_v \beta_9 s_{i9}(x) + \beta_{10} s_{i10}(x) + \beta_{11} s_{i11}(x) + \beta_{12} s_{i12}(x) + \beta_{13} s_{i13}(x) + \beta_{14} s_{i14}(x) \right) \}$$

$$f_i^{Z_h}(\beta, Z_h) = \exp \left\{ \sum_{Z_h} \beta_1 s_{i1}(x, Z_h) \right\}$$

Las tasas de arribo son constantes (sin efectos de covariables o de estructura): ; . Para la dinámica de comportamiento se considera la cantidad de errores de una palabra / nodo.⁴ Se utilizó *RSiena* (Ripley *et al.*, 2018).

Se testearon las siguientes hipótesis:

- i) El nodo (palabra) es origen de error, según un determinado atributo [*EgoX*].
- ii) El nodo (palabra) no es origen de error, según un determinado atributo [*AltX*].
- iii) Un determinado atributo es similar en ambos nodos (palabra) [*SimX*, *SameX*].
- iv) El error es recíproco, según un determinado atributo [*SimRecipX*, *sameXRecip*].
- v) La acumulación de errores en una palabra crece, según un determinado atributo [*EffFrom*].

Las Tablas 4 a 7 muestran los resultados significativos del modelo estimado para cada aprendiente.⁵ En estas aparecen las palabras de controlador con alta imaginabilidad y concretud, así como aquellas que funcionan como adjetivos y las que conllevan -e- epentética susceptibles a la incurrencia en errores. De modo similar, las palabras con controlador de familiaridad / frecuencia léxica alta tienden tanto a generar como a recibir errores. Por otra parte, las palabras terminadas en *es* (ej., *cantantes*, *amables*), las que conllevan -e- epentética y las que tienen alta similitud de raíz / desinencia con el italiano tienden a tener valores similares con el otro término de la concordancia errónea. Solamente en el caso de MIRKA se halló un efecto de influencia en la acumulación de error, respecto de controladores con valores imaginabilidad / concretud altos. Por lo tanto, no hubo efectos significativos de reciprocidad de errores. En la Tabla 3 se resume lo dicho.

4. Los nodos que todavía no son parte de la sesión tienen valor faltante (NA).

5. Se considera una convergencia aceptable para el ajuste de cada coeficiente si ρ y la *ratio* de convergencia máximo del modelo es menor a 0.25.

Hipótesis	SONIA	NATI	JAKO	MIRKA
i)	FAM.LEX	MOD = «4»; ES	IMA.CONC	-
ii)	FAM.LEX	-	-	-
iii)	ES; ESP = «es»; STEM	ESP = «es»	MORF	ESP = «es»; ES
iv)	-	-	-	-
v)	-	-	-	IMA.CONC

Tabla 3. Resumen de resultados para cada hipótesis

Una manera de evaluar la bondad de ajuste del modelo consiste en comparar el modelo observado con otros resultados posibles simulados a partir del modelo ajustado. Se simularon 1000 redes generadas a partir de los parámetros del modelo y se calculó para cada una la probabilidad de cada grado entrante [*in degree*: cantidad de enlaces entrantes a un nodo], tomando de cero a tres grados. Los resultados se muestran mediante un *violin plot* para cada grado mostrando la distribución sobre los mil valores obtenidos. En el punto rojo se indica el valor observado. El *p*-valor, por su parte, muestra la probabilidad de observar la distribución de grado observada en las mil simulaciones —o más extrema—. Cuanto más alto, mejor ajuste es decir, que la hipótesis nula es «el modelo ajusta bien» y, por ende, no debe rechazarse []. Se puede notar, pues, que ningún modelo ajustado rechaza la hipótesis nula (ver Figura 3).

Function	Effect	Coef	StEr	Pval	Sig	Conv
Red	constant red rate (period 1)	1.162	0.239	$p < 0.001$	***	0.086
Red	constant red rate (period 2)	0.238	0.103	0.021	*	0.045
Red	Reciprocity	3.366	1.483	0.0232	*	0.070
Red	indegree - popularity	0.761	0.249	0.0022	**	0.019
Red	anti isolates	3.938	1.429	0.0058	**	0.007
Red	stem similarity	2.256	1.056	0.0326	*	0.075
Red	same esp_3	4.235	1.705	0.013	*	0.060
Red	same es	1.873	0.903	0.038	*	0.089
Red	fam.lex alter	0.695	0.329	0.0343	*	0.001
Red	fam.lex ego	12.922	3.065	$p < 0.001$	***	0.058
Error	rate error (period 1)	0.604	0.206	0.0034	**	0.011
Error	rate error (period 2)	0.310	0.102	0.0025	**	0.035

Tabla 4. Resultados del análisis de errores de SONIA (Maximum Convergence Ratio: 0.236).

Function	Effect	Coef	StEr	Pval	Sig	Conv
Red	constant red rate (period 1)	2.741	0.783	5e-04	***	0.076
Red	constant red rate (period 2)	0.591	0.202	0.0034	**	0.018
Red	Reciprocity	2.628	0.517	$p < 0.001$	***	0.038

Red	outdegree – popularity	0.625	0.121	$p < 0.001$	***	0.036
Red	isolate – popularity	2.478	0.536	$p < 0.001$	***	0.009
Red	mod_4 ego	1.569	0.699	0.0248	*	0.050
Red	same esp_3	1.748	0.523	8e-04	***	0.060
Red	es ego	45.032	4.495	$p < 0.001$	***	0.008
Error	rate error (period 1)	1.401	0.249	$p < 0.001$	***	0.031
Error	rate error (period 2)	0.355	0.085	$p < 0.001$	***	0.002

Tabla 5. Resultados del análisis de errores de NATI (Maximum Convergence Ratio: 0.126).

Function	Effect	Coef	StEr	Pval	Sig	Conv
Red	constant red rate (period 1)	1.523	0.485	0.0017	**	0.017
Red	constant red rate (period 2)	0.359	0.120	0.0027	**	0.009
Red	Reciprocity	2.575	0.680	2e-04	***	0.009
Red	out-in degree ^(1/2) assortativity	2.411	0.440	$p < 0.001$	***	0.029
Red	in-struct equivalence	0.683	0.353	0.0532	.	0.065
Red	morf similarity	2.672	1.000	0.0075	**	0.013
Red	ima.conc ego	3.288	0.881	2e-04	***	0.005
Error	rate error (period 1)	0.556	0.168	9e-04	***	0.006
Error	rate error (period 2)	0.184	0.068	0.0071	**	0.036

Tabla 6. Resultados del análisis de errores de JAKO (Maximum Convergence Ratio: 0.107).

Function	Effect	Coef	StEr	Pval	Sig	Conv
Red	constant red rate (period 1)	0.849	0.113	$p < 0.001$	***	0.015
Red	constant red rate (period 2)	0.893	0.104	$p < 0.001$	***	0.086
Red	Reciprocity	3.309	0.387	$p < 0.001$	***	0.007
Red	Balance	0.533	0.142	2e-04	***	0.014
Red	anti in-isolates	0.769	0.451	0.0884	.	0.023
Red	same esp_3	2.728	0.864	0.0016	**	0.007
Red	same es	0.801	0.295	0.0065	**	0.006
Red	ima.conc alter	0.212	0.136	0.1192		0.008
Error	rate error (period 1)	0.568	0.122	$p < 0.001$	***	0.025
Error	rate error (period 2)	0.821	0.146	$p < 0.001$	***	0.042
Error	error: effect from ima.conc	0.479	0.206	0.0199	*	0.029

Tabla 7. Resultados del análisis de errores de MIRKA (Maximum Convergence Ratio: 0.387).

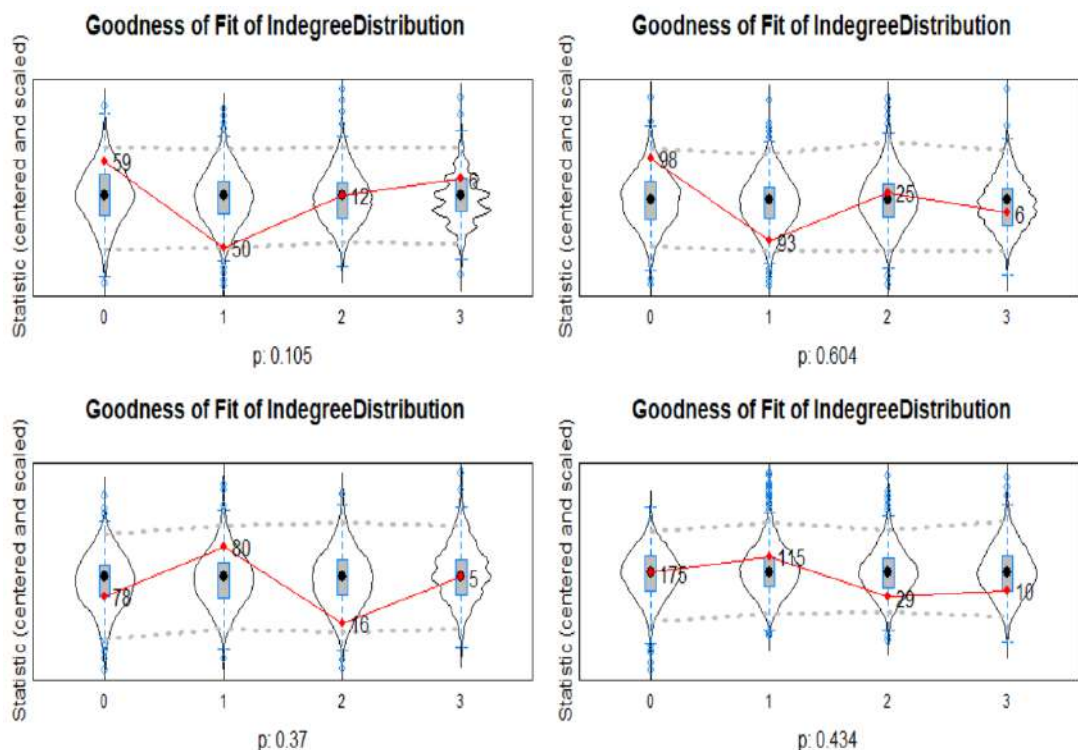


Figura 3. *Bondad de ajuste*. Arriba: SONIA (izq.), NATI (der.); Abajo: JAKO (izq.), MIRKA (der.).

5. Discusión

El objetivo de este trabajo fue identificar algunos factores influyentes en la aparición de errores en un *corpus* longitudinal de concordancias plurales para cuatro alumnos de ELE. Para ello, se crearon redes complejas según tres oleadas de acumulación de errores. En estas, un enlace indicaba la aparición de una concordancia errónea entre dos términos / nodos. Inicialmente, se caracterizó cada nodo con una serie de atributos. Para ello, se usó una técnica estadística de evolución de redes para predecir la aparición de un enlace (error de concordancia) en la siguiente oleada según la influencia de los atributos especificados en los nodos. Un efecto significativo reportado en la revisión de literatura es que las concordancias con adjetivos son más difíciles que aquellas con artículo (Fernández-García, 1999; White *et al.*, 2004; Montrul *et al.*, 2008; Alarcón, 2011). En NATI, se puede notar que los términos (nodos) que funcionan como adjetivos tienden a crear relaciones de concordancias erróneas con otros términos, a medida que pasa el tiempo, y en la dirección de origen del error. Otro hecho constatado es que el plural en *-es* [-*e*- epentética] se adquiere luego del plural en *-s* (Bruhn de Garavito, 2008). Es razonable suponer que las concordancias que involucren plurales en *-es* resulten más desafiantes. Se halló en NATI, además, que las palabras con plurales que involucran *-e*- epentética son origen de error. Es más, en SONIA y MIRKA se tiende a establecer una relación errónea si ambos términos conllevan *-e*- epentética [*pantalones azules*]. Sin embargo, en SONIA, NATI y MIRKA también se encontró el efecto de similitud con terminaciones en *s* en los dos

términos con singulares no canónicos terminados en *-e* [*fuentes grandes*]. Por otra parte, la revisión de literatura también se reporta un efecto de animicidad para el género (Fernández-García, 1999; Montrul *et al.*, 2008), ausente en el presente análisis. No obstante, se incluyeron otras características del controlador, encontrándose que las palabras de controlador con alta imaginabilidad y concretud tendieron a ser origen del error en JAKO. Asimismo, ejercieron un efecto en el aumento acumulado del error en MIRKA. Por otro lado, las palabras con controlador de familiaridad / frecuencia léxica alta tendieron tanto a generar como a recibir errores. Dichos efectos abogan por la inclusión en los estudios de más características del controlador, además de la animicidad. Por último, las palabras que conllevan alta similitud de raíz / desinencia con el italiano tienden a tener valores similares con el otro término de la concordancia errónea en SONIAy JAKO.

A modo de cierre, se considera que este trabajo ha contribuido al estudio de la adquisición de ELE incorporando una herramienta de análisis temporal; es decir, que tenga en cuenta la naturaleza dinámica del aprendizaje de una lengua extranjera.

Referencias bibliográficas

1. 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.
2. Bruhn de Garavito, J. (2008). Acquisition of the Spanish Plural by French L1 Speakers: The Role of Transfer. In J. Liceras, H. Zobl, & H. Goodluck (Eds.), *The Role of Features in Second Language Acquisition* (pp. 270-298). Mahwah, NJ: Lawrence Erlbaum.
3. Bruhn de Garavito, J. & White, L. (2002). The Second Language Acquisition of Spanish DPs: The Status of Grammatical Features. In A. T. Pérez-Leroux & J. Muñoz Liceras (Eds.), *The Acquisition of Spanish Morphosyntax: The L1/L2 Connection* (pp. 153-178). Dordrecht: Kluwer.
4. Corbett, G. (2006). *Agreement*. Cambridge: Cambridge University Press.
5. Davis, C. & 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.
6. Enders, C. (2010). *Applied Missing Data Analysis*. New York: Guilford Press.
7. Español-Echevarría, M. & Prévost, P. (2004). Acquiring Number Specification on L2 Spanish Quantifiers: Evidence Against the Rich Agreement Hypothesis. En J. Van Kampen & S. Baauw (Eds.), *Proceedings of the 2003 Conference on Generative Approaches to Language Acquisition* (pp. 151-162). Utrecht: LOT.
8. Fernández-García, M. (1999). Patterns of Gender Agreement in the Speech of Second Language Learners. In 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.
9. Foote, R. (2015). The Production of Gender Agreement in Native and L2 Spanish: The Role of Morphophonological Form. *Second Language Research*, 31, 343-373.
10. Franceschina, F. (2001). Morphological or Syntactic Deficit in Near-native Speakers? An Assessment of Some Current Proposals. *Second Language Research*, 17, 213-247.
11. Muñoz Liceras, 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. In R. P. Leow & C. Sanz (Eds.), *Current Research on the Acquisition of Spanish* (pp. 67-96). Somerville, MA: Cascadilla Press.
12. Mac Whinney, B. (2020). *The Childes Project: Tools for Analyzing Talk. Part 1 & 2*. Electronic Edition. <https://talkbank.org/manuals>
13. 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.
14. Nerbonne, J., Van Ommen, S., Gooskens, C. y Wieling, M. (2013). Measuring Socially Motivated Pronunciation Differences. In L. Borin & A. Saxena (Eds.), *Approaches to Measuring Linguistic Differences* (pp. 107-140). De Gruyter Mouton.

15. Newman, M. (2010). *Networks: An Introduction*. Oxford University Press.
16. Oakes, M. P. (1998). *Statistics for Corpus Linguistics*. Edinburgh University Press.
17. O'Grady, W. (2005). *Syntactic Carpentry: An Emergentist Approach to Syntax*. Mahwah, NJ: Lawrence Erlbaum Associates.
18. Peña, D. (2002). *Análisis de Datos Multivariantes*. Madrid: Mc Graw Hill.
19. Ripley, R., Snijders, T., Boda, Z., Voros, A., & Preciado, P. (2018). *Manual for Siena Version 4.0; R Package Version 1.2-12*. <https://www.cran.r-project.org/web/packages/RSiena/>
20. Scrucca, L., Fop, M., Murphy, T. & Raftery, A. (2016). Mclust 5: Clustering, Classification and Density Estimation Using Gaussian Finite Mixture Models. *The R Journal*, 8(1), 205-233.
21. Snijders, T. (1996). Stochastic Actor-Oriented Models for Network Change. *Journal of Mathematical Sociology*, 21, 149-176.
22. Snijders, T., Steglich, C. & Schweinberger, M. (2007). Modeling the Co-evolution of Networks and Behavior. In K. Van Montfort, J. Oud & A. Satorra (Eds.), *Longitudinal Models in the Behavioral and Related sciences* (pp. 41-71). Mahwah, NJ: Lawrence Erlbaum.
23. Snijders, T., Van de Bunt, G. & Steglich, C. (2010). Introduction to Stochastic Actor-Based Models for Network Dynamics. *Social Networks*, 32(1), 44-60.
24. Van Buuren, S. & Groothuis-Oudshoorn, K. (2011). Mice: Multivariate Imputation by Chained equations in R. *Journal of Statistical Software*, 45(3), 1-67.
25. White, L., Valenzuela, E., Kozłowska-Macgregor, M., & Leung, Y. (2004). Gender and Number Agreement in Nonnative Spanish. *Applied Psycholinguistics*, 25(2), 105-133.