

Modelación de errores de concordancia plural en español LE mediante sistemas dinámicos*

PABLO EZEQUIEL MARAFIOTI**

Recepción: 12 de octubre de 2022

Aprobación: 14 de diciembre de 2022

Forma de citar este artículo: Marafioti, P. (2022). Modelación de errores de concordancia plural en español LE mediante sistemas dinámicos. *Cuadernos de Lingüística Hispánica*, (40), e14700

 <https://doi.org/10.19053/0121053X.n40.2023.14700>

* Artículo de investigación.

** Doctor en Ciencias del Lenguaje por la Universidad Nacional de Córdoba, Argentina; especialista en Lingüística Teórica y Aplicada por la Universidad de Pavia, Italia; licenciado en Letras (orientación en Lingüística) por la Universidad de Buenos Aires, Argentina. Realizó varios posgrados en Estadística. Áreas de especialización: sistemas dinámicos, modelos estadísticos y de minería de datos aplicados al cambio lingüístico y a la adquisición de lenguas extranjeras. kenavo38@yahoo.com  <https://orcid.org/0000-0001-7299-5324>

Resumen

El propósito de este artículo es modelar los errores de concordancias nominales plurales para un corpus de español de cuatro aprendientes de español como lengua extranjera: SONIA (nivel A1/A2), NATI (nivel B1), JAKO (nivel B2), MIRKA (nivel C1). En el marco de los sistemas dinámicos, se postula que las concordancias “correcta” e “incorrecta” constituyen atractores discretos estables a los cuales converge el flujo continuo del sistema en tres modelos simples de simulación: Lokta-Volterra, juego de coordinación (teoría de los juegos evolutiva), ascenso de gradiente. Como resultados, en general, los modelos aproximan mejor el patrón de error global que el de sesiones. En este último, el modelo basado en Lokta-Volterra ofrece mejor desempeño para SONIA, NATI y JAKO. MIRKA obtiene el peor desempeño en todos los modelos, pero su error disminuye con el modelo de ascenso de gradiente. Se concluye que los patrones de aprendizaje se autoorganizan y emergen de la dinámica de microsistemas de concordancia.

Palabras clave: modelo de simulación, adquisición de L2, sistemas dinámicos, número plural.

Modeling of Plural Concordance Errors in Spanish as Foreign Language Using Dynamic Systems

Abstract

The purpose of this paper is to model plural nominal concordance errors for a Spanish corpus of four learners of Spanish as a foreign language: SONIA (level A1/A2), NATI (level B1), JAKO (level B2), MIRKA (level C1). In the framework of dynamical systems, it is postulated that “correct” and “incorrect” concordances constitute stable discrete attractors to which the continuous flow of the system converges in three simple simulation models: Lokta-Volterra, coordination game (evolutionary games theory), gradient ascent. As a result, the models better approximate the global error pattern than the session pattern. In the latter, the model based on Lokta-Volterra offers better performance for SONIA, NATI and JAKO. MIRKA obtains the worst performance in all the models, but its error decreases with the gradient ascent model. It is concluded that learning patterns are self-organized and emerge from the dynamics of concordance microsystems.

Keywords: simulation model, second language acquisition, dynamic systems, plural number.

Modélisation des erreurs d'accord pluriel en espagnol LE par le biais de systèmes dynamiques

Résumé

Les erreurs d'accord nominatifs pluriels sont modélisées pour un corpus d'espagnol langue étrangère de quatre apprenants : SONIA (niveau A1/A2), NATI (niveau B1), JAKO (niveau B2), MIRKA (niveau C1). Dans le cadre des systèmes dynamiques, il est postulé que les concordances “correctes” et “incorrectes” constituent des attractrices discrètes stables vers lesquelles converge le flux continu du système dans trois modèles de simulation simples : Lokta-Volterra, jeu de coordination (théorie des jeux évolutifs), montée en gradient. En général, les modèles se rapprochent mieux au modèle d'erreur global qu'à celui de sessions. Dans ce dernier, le modèle basé sur Lokta-Volterra offre de meilleures performances pour SONIA, NATI et JAKO. MIRKA obtiennent les plus mauvaises performances dans tous les modèles mais leur erreur diminue avec le modèle de montée en gradient. Les schémas d'apprentissage s'auto-organisent et émergent de la dynamique de micro-systèmes d'accord.

Mots clés: modèle de simulation, acquisition L2, systèmes dynamiques, nombre pluriel.

Modelagem de erros de acordo plural em espanhol por meio de sistemas dinâmicos

Resumo

Erros de concordância nominal plural são modelados para um corpus espanhol de quatro aprendizes de espanhol como língua estrangeira: SONIA (nível A1/A2), NATI (nível B1), JAKO (nível B2), MIRKA (nível C1). No âmbito dos sistemas dinâmicos, postula-se que as correspondências “corretas” e “incorretas” constituem atratores discretos estáveis para os quais converge o fluxo contínuo do sistema em três modelos simples de simulação: Lokta-Volterra, jogo de coordenação (teoria dos jogos evolutivos), gradiente ascendente. Em geral, os modelos aproximam melhor o padrão de erro global do que o modelo de sessão. Neste último, o modelo baseado em Lokta-Volterra oferece melhor desempenho para SONIA, NATI e JAKO. MIRKA obtém o pior desempenho em todos os modelos, mas seu erro diminui com o modelo gradiente ascendente. Os padrões de aprendizagem são auto-organizados e emergem da dinâmica de microssistemas de concordância.

Palavras-chave: modelo de simulação, aquisição de L2, sistemas dinâmicos, número plural.

Introducción

En la corriente de pensamiento emergentista (MacWhinney, 2015), la estructura lingüística emerge a partir de patrones de uso repetido en el tiempo. Los niveles del lenguaje se hallan interconexos y se adopta la visión conexionista de interrelación entre módulos permeables y emergentes. Se supone un *input* muy rico que permite el aprendizaje mediante mecanismos generales que abstraen propiedades estadísticas y distribucionales de dicho *input*. Los patrones del lenguaje surgen del uso, la generalización y la autoorganización.

En este trabajo se considera el lenguaje como un sistema dinámico, es decir, como sistema que cambia en el tiempo de procesamiento *online* en el momento del habla. Al aprender una lengua, el aprendiente crea variabilidad, esto es, nuevas formas (“errores”) que no estaban codificadas en las condiciones iniciales del *input* que recibe (o en su L1). El contexto/ambiente cognitivo de producción (características de la forma seleccionada, memoria a corto plazo, motivación, similitud de rasgos con su L1, etc.) influye para que el hablante profiera formas “correctas” o “incorrectas”. Por ejemplo, realizar una concordancia en un contexto de dependencia a larga distancia podría aumentar la chance de error. Cada forma constituye el resultado de un sendero continuo de activación mental que desemboca en una forma lingüística discreta proferida. Pero, a medida que pasa el tiempo, el sistema del lenguaje se adapta a los contextos y, si se producen cambios de dinámica/bifurcaciones —por ejemplo, un aumento de competencia lingüística— los errores disminuyen. Colectivamente emerge un patrón global de error. O sea que el error se constituye como un fenómeno emergente.

Las reglas que gobiernan el sistema se especifican como ecuaciones (diferenciales ordinarias no lineales). La dinámica se representa en un espacio de fase (*phase space*) de m dimensiones, donde se hallan todos los posibles estados del sistema, y cada estado corresponde a un único punto m -dimensional en dicho espacio. A medida que pasa el tiempo, el vector de estado cambia posición en el espacio de fase, creando así una trayectoria o flujo. Un punto fijo es aquel donde el sistema no cambia (la tasa de crecimiento es constante). En sistemas de flujos unidimensionales, un punto fijo puede ser un “atractor” (estable), un “repulsor” (inestable) o ser un “punto silla” (*saddle*). Un “atractor” es una región/punto del espacio de fase al cual el sistema es atraído. Por el contrario, un punto del espacio de fase del cual el flujo se aleja constituye un “repulsor”. En un “punto silla”, el punto fijo atrae en una dirección y repele en la otra. El conjunto de condiciones iniciales que llevan a un atractor se llama “cuenca de atracción”. Una bifurcación es un cambio cualitativo en el comportamiento del sistema (Kaplan & Glass, 1995).

No existen investigaciones de modelado con sistemas dinámicos aplicados a la adquisición de lenguas segundas. Smith *et al.* (2018) propusieron un modelo para el fenómeno de atracción en la concordancia sujeto-verbo en L1, en inglés, focalizándose en sujetos que eran construcciones pseudopartitivas del tipo “una caja

[N1] de chocolates [N2]”. Utilizaron un modelo generalizado de competición de Lokta-Volterra. El sistema terminaba atraído a una configuración con el N1 (concordancia correcta) o a otra con el N2 (incorrecta). Otra vertiente de investigación son los modelos de ascenso de gradiente. Dichos modelos definen una función de “armonía” global que indica el grado de “buena formación” de los estados del sistema. Estos estados se definen en términos de configuraciones de rasgos. Los “picos” de la función se corresponden a diferentes tipos de estructuras lingüísticas simbólicas, estén bien o mal formadas. Los modelos “recorren” la función de armonía en dirección ascendente y se detienen en uno de los picos, que constituyen atractores. Debido a la acción del “ruido”, el sistema puede ser empujado hacia atractores de estructuras mal formadas, o sea, ser inducido a error. En dicha dirección están los trabajos de Smith y Tabor (2018), Smith *et al.* (2021) y el paradigma de *Gradient Symbolic Computation* (Smolensky et al., 2014; Cho et al., 2017, 2018, 2020; Brehm et al., 2022). Por otra parte, se encuentran los modelos que aplican teoría de los juegos evolutiva al lenguaje. Deo (2015) utilizó dicha teoría para modelar el ciclo progresivo (“estaba comiendo”) – imperfecto (“comía”) en el cambio semántico diacrónico de los verbos en pasado mediante cuatro estadios/estrategias.

Aquí se propone ver las concordancias (nominales) correctas e incorrectas producidas por los aprendientes de español como lengua extranjera (ELE) como atractores de un sistema dinámico. Se intentará modelar la dinámica en las trayectorias de “activación mental” que lleven el flujo hacia alguno de los atractores. Cada punto en el estado de fase será un vector que describirá la configuración de activación “mental”. Se considerará que inicialmente el aprendiente se encuentra en un estado mental “mixto” entre los posibles atractores estables; esto es, en un punto fijo inestable del cual se debe alejar para entrar en la cuenca de atracción de alguno de los atractores mencionados. Diversas características de los ítems experimentales (características del controlador nominal, si es a larga distancia, etc.) determinarán hacia cuál cuenca habrá inclinación, lo cual sesgará a su vez qué atractor resultará ganador. Esta dinámica se producirá para cada instancia de concordancia producida por el aprendiente. Los atractores serán símbolos discretos que emergen de una dinámica continua. Si en cada instancia de concordancia se aplica la dinámica, colectivamente emergerá un patrón de proporción global de error. Las variables relevantes se identificaron con ayuda de modelos estadísticos aplicados al corpus de datos. Estas formaron parte de un mecanismo de distancia a configuraciones de variables con chance de error baja, que permitió “romper” el estado mixto inicial e inclinar el flujo hacia alguno de los atractores. El error de concordancia es representado como un atractor simbólico (estado puro) que emerge de la dinámica de estados mixtos entre los diferentes atractores de estructuras correctas y erróneas. Se realizaron simulaciones para aproximar los patrones observados de error global y por series temporales a partir de datos longitudinales de cuatro aprendientes con diferente competencia en ELE.

Método

Participantes y diseño del corpus

Se utilizó un corpus de adquisición de cuatro aprendientes italianos de ELE. Se trató de cuatro alumnos adultos, de lengua nativa italiana, estudiantes del Instituto Cervantes de Milán en el año académico 2008-2009. Cada alumno poseía un nivel distinto de competencia lingüística según el *Marco Común Europeo de Referencia* (los nombres son ficticios para preservar la identidad de los sujetos). Se hicieron entrevistas de treinta 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. Cada alumno realizaba simultáneamente el curso de español del nivel. Hubo entre doce y catorce entrevistas por alumno. La codificación y transcripción de los datos se hizo mediante el formato CHAT, siguiendo a MacWhinney (2021). 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.

Se extrajeron las concordancias plurales en el dominio nominal, verbal (predicativos) y en oraciones subordinadas. La concordancia se consideró asimétrica con controlador nominal. Además, siempre estaban formadas por dos términos, es decir que si hubiere más de dos, por ejemplo, *las casas blancas*, se codificaron dos ejemplos: *las casas* y *casas blancas*. Se clasificaron los errores del siguiente modo: (1) ausencia de error; (2) errores en el género; por ejemplo: *muchos personas* [Sonia, sesión 1, línea 56] (error por: *muchas personas*); (3) errores debidos al uso del alomorfo *-es*, por ejemplo: *muchos trenos* [Sonia, sesión 7, línea 148] (error por *muchos trenes*); (4) errores de plural, o sea ausencia del alomorfo *-s*; por ejemplo: *los veneciano* [Sonia, sesión 5, línea 293] (error por *los venecianos*); (5) errores mixtos por acumulación de los anteriores; por ejemplo: *les joven* [Sonia, sesión 2, línea 144] (error por *los jóvenes*). El conteo se hizo con el programa CLAN. Se crearon variables que caracterizaban cada instancia objetivo de concordancia en español (con independencia de que la instancia efectivamente producida contuviera o no error). Se aplicaron dos modelos estadísticos a la variable respuesta de tipos de errores: (a) un modelo multinomial bayesiano (Marafioti, 2021); (b) un modelo de riesgos competitivos utilizando el tiempo hasta que ocurre el evento “tipo de error” (Marafioti, 2022). Las siguientes variables resultaron en una disminución de la chance de los tipos de 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).

Fabs. La frecuencia del TYPE de concordancia. Cada TYPE especificaba el contexto de la concordancia. 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 propios [variable *Fabs_c*] y de un corpus del español electrónico en línea [variable *Fabs_s*]. Para esto último, se apeló al corpus del español *EsTenTen* de *Sketch Engine* (Kilgarriff et al., 2014). Primero se transformaron según: $Fabs_C = \log(Fabs_C + 1)$ 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). Luego, dicho índice se discretizó usando *clustering* por mezcla de gaussianas (Scrucca et al., 2016), en los niveles: 1 = frecuencia alta, 0 = frecuencia baja.

Epen. Se especificó si en el controlador, en el objetivo, o en ambos, había una desinencia que requería la inserción de “e” epentética [-(e)s]; según [ES]: 0 = sin “e” epentética; 1 = con “e” epentética en un término; 2 = con “e” epentética en ambos términos.

FamLex. Índice a partir de PCA combinando los siguientes rasgos del controlador nominal extraídos de la base de datos “BuscaPalabras” (Davis & Perea, 2005): (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.

Morf. Se usó la distancia de Levenstein (Nerbonne et al., 2013) con el objetivo de medir la similitud entre los alomorfos de género y número plural. El algoritmo de Levenstein calcula la distancia entre dos secuencias de caracteres como el número mínimo de operaciones necesarias para transformar una secuencia en la otra. Se las discretizó mediante *clustering* por mezcla de gaussianas, según: 0 = distancia baja, 1 = distancia media, 2 = distancia alta.

Estrategias. Se crearon 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. Entre estas: (a) *EST1*: si la palabra plural del italiano termina en *-i* poner en español plural en *-os*; (b) *EST2*: si la palabra plural del italiano termina en *-e* poner en español plural en *-as*; (c) *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*. Se trata de casos en los cuales el español coincide con la aplicación del plural con *-e-* epentética.

Se recolectaron 1857 casos de concordancia en total. Sin embargo, los atributos relacionados con el controlador a veces no tenían datos registrados en la base de datos de “BuscaPalabras”. Debido a ello, hubo 161 casos en los que faltaban

datos en una o más de estas variables. Los casos faltantes representaron el 8.6 % de la base de datos. Se utilizó el paquete *mice* (Multivariate Imputation by Chained Equations) de *R* (Van Buuren & Groothuis-Oudshoorn, 2011), que realiza imputación múltiple. Por ejemplo, en la instancia “los alemanos” (error por “los alemanes”) [SONIA, sesión 3, línea 178] el modificador es un artículo definido [Mod = 0]; se trata de una concordancia de frecuencia alta [Fabs = 1]; el plural de “alemán” requiere insertar una “e” epentética [Epen = 1], “alemán” conlleva una familiaridad / frecuencia léxica baja [FamLex = 0], la distancia con las desinencias del italiano (“i tedeschi”) es alta [Morf = 2]; no es aplicable ninguna estrategia [Est1 = Est2 = Est5 = 0].

Los siguientes son los niveles de las variables asociadas a disminución de chance de error de concordancia plural, según tipo de error; halladas en Marafioti (2021, 2022): (i) adjetivos [Mod = “3”] (en errores de género); (ii) frecuencia alta de TYPE de concordancia [Fabs = “1”] (en errores de género, plural y mixto); (iii) concordancias cuyas terminaciones tienen similitud media [Morf = “1”] (error de “e” epentética y plural) y alta con el italiano [Morf = “2”] (error de plural); (iv) concordancias cuyo controlador tiene alta familiaridad y/o frecuencia léxica (errores de “e” epentética, plural y mixto); (v) concordancias en las que ambas terminaciones llevan “e” epentética [Epen = “2”] (error de plural); (vi) las concordancias en donde es posible aplicar las estrategias EST1 (error de género y mixto), EST2 (error de “e” epentética), EST5 (error de “e” epentética, mixto).

También se tuvo en cuenta, para cada instancia, la proporción de errores cometidos hasta la instancia anterior, contando desde el inicio de cada sesión. Por ejemplo, en SONIA, la primera sesión consta de las siguientes instancias: [“muchos personas” (error), “muchos pueblos” (correcto), “mucho pensadores” (error), “otros autores” (correcto)] y tiene las siguientes proporciones de errores anteriores a una instancia en cuestión: [0; 1; .5; .66]. Es decir que en “otros autores” hubo dos errores sobre tres instancias producidas antes, contando desde el inicio de la sesión.

Modelado

Las simulaciones que siguen tienen siempre dos atractores estables: “correcto” y “error”. Para cada instancia se intenta recrear un sendero de “activación” en el “paisaje de eventos mentales” hasta que el sistema se acerque a uno de los atractores. Las variables especificadas en cada tipo de error se toman como un “contexto” que sesga el flujo de activación hacia uno de los atractores. Para lograr esto, se plantean diferentes mecanismos de similitud entre las variables de la instancia que se procesa y cada uno de los contextos que disminuyen el error. O sea, la idea fue combinar la lejanía a los contextos que favorecen el buen desempeño junto con los errores cometidos anteriormente, para sesgar el flujo de activación hacia uno de los dos atractores. A continuación, se detallan los modelos. El material complementario con el detalle técnico y el código de *R* empleado puede consultarse en github (<https://github.com/pablomarafioti/PabloMarafioti/tree/master/Simulacion>).

Modelo competitivo de Lotka-Volterra

Los modelos de competición de Lotka-Volterra intentan captar la interacción entre (dos) especies. Dichas especies se interfieren entre sí, disminuyendo (las unas a las otras) las tasas de crecimiento de abundancias poblacionales (Kot, 2001; Murray, 2002). Se trata de una extensión a varias especies del modelo logístico de una especie. Sean entonces dos especies, cada una con abundancias N_i , $i = 1; 2$. En ausencia de la otra, la abundancia de cada especie sigue un crecimiento logístico cuya tasa decrece linealmente con el tamaño poblacional. Cada especie tiene su propia tasa constante de crecimiento r_i y su propia capacidad de carga ambiental K_i (“carring capacity”, la máxima capacidad de “carga” del ambiente, respecto de cada especie). El modelo logístico prescribe un sistema dinámico de dos dimensiones, mediante las siguientes ecuaciones diferenciales ordinarias:

$$\frac{dN_1}{dt} = r_1 N_1 \left(1 - \frac{N_1}{K_1}\right) ; \quad \frac{dN_2}{dt} = r_2 N_2 \left(1 - \frac{N_2}{K_2}\right) \tag{1}$$

El término cuadrático representa la competencia intraespecífica, o sea, entre miembros de la misma especie. A medida que $t \rightarrow \infty$, $N_i \rightarrow K_i$, para toda abundancia inicial $N_{0i} > 0$. Ahora introduzcase un efecto de competición interespecífica, en el que cada especie causa un decrecimiento en la abundancia de la otra especie. Ya que las especies son diferentes, el efecto de disminución puede ser más fuerte o más débil. Dicho efecto se representa en dos parámetros de competencia con a_{ij} : la “fuerza” del efecto de competencia de la especie j respecto de la especie i ; a_{12} y a_{21} . El modelo queda:

$$\frac{dN_1}{dt} = r_1 N_1 \left(1 - \frac{N_1}{K_1} - a_{12} \frac{N_2}{K_1}\right) = \frac{r_1}{K_1} N_1 (K_1 - N_1 - a_{12} N_2) \tag{2}$$

$$\frac{dN_2}{dt} = r_2 N_2 \left(1 - \frac{N_2}{K_2} - a_{21} \frac{N_1}{K_2}\right) = \frac{r_2}{K_2} N_2 (K_2 - N_2 - a_{21} N_1) \tag{3}$$

En el ámbito de la adquisición, las abundancias de las dos especies serán: (i) N_1 : la abundancia de especie “correcto”; (ii) N_2 : la abundancia de la especie “error”. Las nulclinas del sistema son los puntos (N_1, N_2) en los que el flujo del campo vector apunta en una determinada dirección; allí la ecuación diferencial de una determinada variable se anula: $\dot{N}_i = \frac{dN_i}{dt} = 0$. Las nulclinas correspondientes a $N_1 = 0$ son: (a) $N_1 = 0$ y (b) $N_1 = K_1 - a_{12} N_2$. Por otra parte, las nulclinas correspondientes a $N_2 = 0$ son: (c) $N_2 = 0$ y (d) $N_2 = 0$ y (d) $N_2 = K_2 - a_{21} N_1$. Los puntos fijos (p. f.) del sistema son aquellos puntos en donde las nulclinas se intersectan entre sí o intersectan los ejes del plano de fase: $p. f. = \{P_1 = (0,0); P_2 = (K_1, 0); P_3 = (0, K_2); P_4 = \left(\frac{K_1 - a_{12} K_2}{1 - a_{12} a_{21}}, \frac{K_2 - a_{21} K_1}{1 - a_{21} a_{12}}\right)\}$.

El sistema permite cuatro tipos de dinámicas posibles. Sin embargo, nos interesa el caso con dos atractores, ya que la hipótesis de este trabajo es que “correcto” y “error” atraen el flujo de la dinámica. Para dicho caso debe suceder que: $a_{12} > \frac{k_1}{k_2}$ y $a_{21} > \frac{k_2}{k_1}$ y. El punto fijo interior P_4 es un nodo “silla” (inestable). El punto

P_1 es un nodo repulsor (inestable) y los puntos P_2 y P_3 son nodos atractores (estables). Hay una *separatrix* que delimita el comportamiento de la dinámica: arriba de esta las órbitas son atraídas hacia P_2 y debajo de esta son atraídas hacia P_3 . Las órbitas que comienzan en la *separatrix* (excepto el origen) convergen hacia P_4 . La competencia interespecífica es fuerte y eventualmente una de las especies gana y la otra se extingue, dependiendo de las condiciones iniciales. Se tomaron los siguientes parámetros: $r_1 = r_2 = .5$; $a_{12} = a_{21} = 1.5$; $K_1 = K_2 = 1$. Con lo cual el punto interior “silla” fue $P_4 = \left(\frac{1-1.5}{1-1.5^2}, \frac{1-1.5}{1-1.5^2}\right) = (.4, .4)$ y los atractores son: $P_3 = (0.1) = \text{“error”}$ y $P_2 = (1.0) = \text{“correcto”}$. La condición inicial del sistema se establece por un mecanismo que involucra distancias de los valores de las variables asociadas a la instancia en cuestión y los contextos descritos. Además, se tiene en cuenta la cantidad de errores producidos hasta el momento de proferir la concordancia plural en cuestión (ver material complementario). La Figura 1 ilustra el espacio de fase del sistema con la recorrida del flujo para dos instancias “muchos pueblos” [SONIA, sesión 1, línea 94] y “les idiomas” [SONIA, sesión 2, línea 111]. Cada punto (N_2, N_1) indica un estado del sistema con las abundancias mixtas de “correcto” (N_1) y “error” (N_2). Es decir que se trata de estados que indican “incertidumbre” respecto de la producción de una determinada concordancia plural. Las flechas indican la dirección del flujo. Las rectas negras son las nulclinas y su intersección indica los puntos fijos. El círculo indica la condición inicial del sistema, determinado por el mecanismo de distancia a los contextos. Para “les idiomas” (línea punteada) la condición inicial está arriba de la *separatrix*, con lo cual el flujo llevará el sistema al atractor “error” (0,1). En el caso de “muchos pueblos” (línea discontinua) la condición inicial se halla debajo de la *separatrix*; por ende, el flujo el flujo conducirá hacia el atractor “correcto” (1,0). La Figura 2 indica las abundancias de “correcto” y “error” a medida que pasa el tiempo. Mientras que las abundancias de una especie aumentan, las otras caen a cero. Eventualmente, una especie alcanza el atractor (“correcto” o “error”) y la otra “se extingue”.

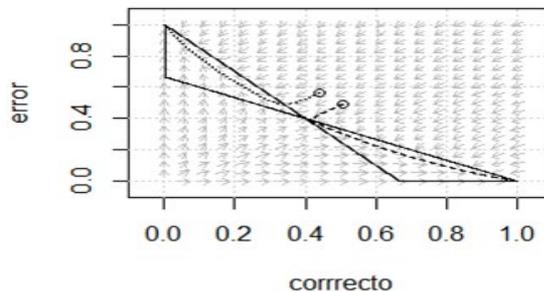


Figura 1. Espacio de fase del modelo competitivo de Lotka-Volterra. Línea punteada: flujo de “les idiomas”; línea discontinua: flujo de “muchos pueblos”.

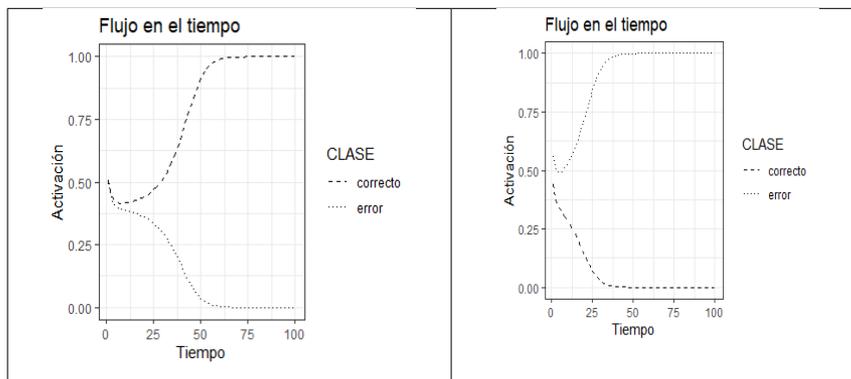


Figura 2. Flujo en el tiempo para las instancias “muchos pueblos” (izquierda) y “les idiomas” (derecha).

Modelo basado en la teoría de los juegos evolutiva

Se considera un “juego de coordinación” con dos jugadores ($i=1,2$), cada uno de ellos con dos estrategias puras y una mixta (Weibull, 1997). El conjunto de estrategias puras de cada uno es $s_1=s_2=\{e_1,e_2\}$, donde la primera estrategia es “error” y la segunda, “correcto”. La estrategia mixta \hat{x} es una distribución de probabilidad sobre el conjunto de estrategias puras. Aquí: $\hat{x} = (\frac{1}{4}, \frac{3}{4})$; o sea que los jugadores usan la primera estrategia (“error”) un 25 % de las veces (en coincidencia con el error total de los datos) y la segunda, un 75 %. Hay una función de utilidad que asocia un pago a cada estrategia. Aquí, la matriz de pagos será doblemente simétrica, dada por: $A = \begin{bmatrix} a_1 & 0 \\ 0 & a_2 \end{bmatrix} = \begin{bmatrix} 3 & 0 \\ 0 & 1 \end{bmatrix}$; $a_1 > 0, a_2 > 0$.

Por ejemplo, supóngase que el primer jugador corresponde a las filas y el segundo, a las columnas. Si ambos juegan la estrategia (primera fila / columna), obtienen un pago de 3. Notar que para la estrategia mixta ocurre que $\hat{x} = (\hat{x}_1, \hat{x}_2) = (\frac{a_2}{a_1+a_2}, \frac{a_1}{a_1+a_2})$. Ahora bien, existe equilibrio de Nash [EN] si un jugador supone que el otro se queda con una determinada estrategia de una fila/columna, al otro no le conviene pasarse de fila/columna. En A , $EN = \{e_1, e_2, \hat{x}\}$; es decir que ambos deberían jugar la estrategia “correcto”, “error” o la mixta. Supóngase ahora una población grande de individuos que son seleccionados aleatoriamente para jugar un juego simétrico de a dos. Cada individuo está programado para jugar las mismas estrategias mencionadas. Cada estrategia es jugada por una fracción de individuos de la población. Una estrategia se dice “evolutivamente estable” si es resistente a las perturbaciones. O sea, si otras estrategias mutantes que se introducen en la población eventualmente se extinguen porque nadie las juega. En el juego que nos ocupa las estrategias puras son estables y la mixta es inestable. Notar que si una estrategia es evolutivamente estable, entonces es un equilibrio de Nash. Ambos conceptos suponen equilibrios estáticos. Pero supóngase ahora que cada individuo está programado

para usar una estrategia pura, la cual replica a su descendencia. La dinámica del replicador muestra cómo la distribución de estrategias puras de la población cambia con el tiempo siguiendo un proceso de selección, y que dicha distribución converge a una estrategia evolutivamente estable (punto fijo). Denótese $x_1(t)$ a la fracción de individuos que juega la estrategia pura “error” en el tiempo t . En consecuencia, $x_2(t) = 1 - x_1(t)$ será la fracción de individuos que juega la estrategia pura “correcto” en el tiempo t . Por ende, el estado de la población en t estará dado por $x(t) = (x_1(t), 1 - x_1(t))$. El pago promedio (“fitness”) de un individuo que usa la estrategia “error” es $u(e_1, x)$; el pago promedio de un individuo contra cualquier otro (promedio o “fitness” poblacional) es $u(x, x)$. Las ecuaciones diferenciales para la dinámica del replicador para la matriz A son:

$$\frac{dx_1}{dt} = x_1[u(e_1, x) - u(x, x)] = x_1(1 - x_1)[a_1x_1 - a_2(1 - x_1)] \quad (4)$$

$$\frac{dx_2}{dt} = -\frac{dx_1}{dt} \quad (5)$$

La tasa de crecimiento de la fracción poblacional de la estrategia “error” es proporcional a la diferencia entre su “fitness” y el “fitness” poblacional. Si esta diferencia es positiva, la tasa de crecimiento aumentará y habrá más individuos jugando dicha estrategia en la próxima generación, en $t + dt$. Ya que $x_2 = -x_1$, se trata de un sistema unidimensional. Por lo tanto, se concentrará el análisis en la primera ecuación. Sus puntos fijos serán los que hagan que $x_1 = 0$, lo cual sucede cuando: $p.f. = \{P_1 = x_1 = 1, P_2 = x_1 = 0, P_3 = x_1 = \frac{a_2}{a_1 + a_2}\}$.

Con la condición $a_1 > 0$, $a_2 > 0$, existe un nodo interior repulsor P_3 , que coincide con la probabilidad de la estrategia “error” del equilibrio de Nash mixto $\hat{x} = (\hat{x}_1, \hat{x}_2)$. Recuérdese que en el juego de coordinación la estrategia mixta \hat{x} no es evolutivamente estable (es un equilibrio inestable). Si la condición inicial $x_1^0 > \hat{x}_1$, el flujo es atraído al punto fijo estable $P_1 = 1$ (gana la estrategia “error”: e_1). Por otra parte, si la condición inicial $x_1^0 < \hat{x}_1$ el flujo es atraído al punto fijo estable $P_2 = 0$ (gana la estrategia “correcto”: e_2). Por lo tanto, las cuencas de atracción (CA) (“basin of attraction”) de ambos atractores son: $CA(P_1) = (\hat{x}, 1)$ y $CA(P_2) = (0, \hat{x})$. La Figura 3 ilustra, a la izquierda, el flujo en el tiempo para el ejemplo “lo estudios” [SONIA, sesión 4, línea 179]. A la derecha se muestra el espacio de fase del modelo. Ya que $x_1^0 = .74 > \hat{x}_1 = \frac{1}{4}$, el flujo escapa del punto inestable interno y se dirige al punto fijo estable de la estrategia pura de “error”.

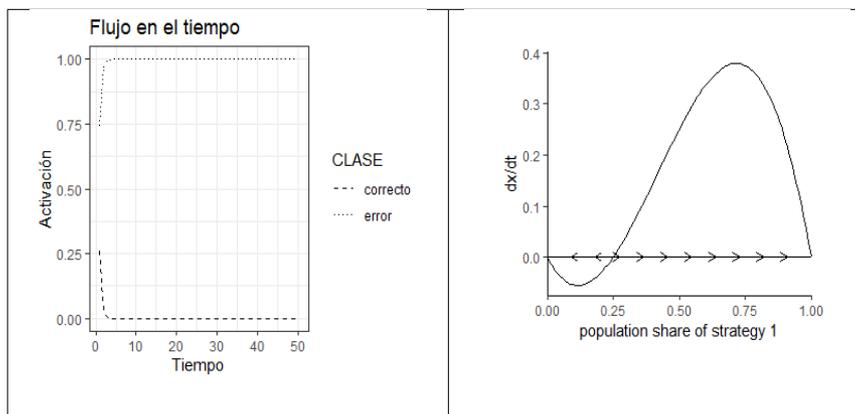


Figura 3. Flujo en el tiempo para la instancia “los estudio” (izquierda) y espacio de fase del modelo (derecha).

2.2.3 Modelo de ascenso de gradiente de armonía

En el paradigma de *Gradient Symbolic Computation* se postulan sistemas dinámicos estocásticos para tiempo y estados continuos. Se construyen gradualmente estructuras simbólicas discretas (atractores) a partir de un espacio de estados continuo. Considérese un modelo para una dimensión donde la variable “ x ” constituye un continuo de activación entre $[-1, 1]$. En los extremos “ x ” representa dos estados “puros” o atractores discretos: $x = -1$ es el símbolo “correcto” y $x = 1$ representa el símbolo “error”. Los valores de $x \in (-1, 1)$ son estados “mezcla” entre ambos símbolos. Se desea que el sistema comience en un estado arbitrario y la activación vaya hacia alguno de los estados “puros”. Al inicio, la función de armonía $H(x, q, e)$ tiene un único atractor que representa dicho estado “mezcla”, intermedio entre “error” y “correcto”. A medida que el parámetro $q \in [0, 5]$ cambia, se produce una bifurcación y aparecen dos atractores estables. Por otro lado, el parámetro $e \in [-1, 1]$ es el valor que sesga el flujo hacia alguno de los atractores. Se lo calcula mediante un mecanismo de similitud a los contextos ya especificados (ver material complementario). A medida que pasa el tiempo, el sistema escala el gradiente de la función de armonía $\nabla_x H(x, q, e)$ para terminar cerca del pico de alguno de los dos atractores. Sin embargo, si el ruido es grande, el sistema podría, por ejemplo, terminar en el atractor “error” aunque estuviera sesgado hacia el atractor “correcto”. El modelo estocástico se muestra en (6); la función de armonía, en (7) y su gradiente, en (8). En (6) W es el proceso de Wiener, T es el nivel de ruido (que puede ser constante o decrecer con el tiempo de modo exponencial) y ∇_x es el gradiente de la armonía total evaluada en valor de activación “ x ”. El primer término es el determinístico (*drift*) y el segundo es la parte estocástica.

$$\frac{dx}{dt} = \nabla_x H(x, q, e)dt + \sqrt{2T} dW \quad (6)$$

$$H(x, q, e) = -2x^2 - q[(x - 1)(x + 1)]^2 + ex \quad (7)$$

$$\nabla_x H(x, q, e) = -4x - 4qx(x - 1)(x + 1) + e \quad (8)$$

Ahora bien, supóngase $e = 1$; es decir que el sistema está sesgado hacia el atractor “error” (si $e = 0$ no hay sesgo). La Figura 4 ilustra el comportamiento del sistema a medida que cambia el parámetro dinámico “q”. Si $q = 0$ la “superficie” de armonía tiene la forma de una parábola con máximo global en $x = \frac{1}{4}$. Por lo tanto, dicho máximo es el único atractor y representa un estado “mezcla”, intermedio entre “error” y “correcto”, con un poco de sesgo hacia el “error”. Si se aumenta el valor de “q”, se observará que aparecen dos máximos locales. Ha sucedido una bifurcación y se ha pasado de un atractor a dos atractores. A medida que “q” sigue aumentando, el “valle” entre ambos máximos locales se hace más “profundo”. El sistema “escalará” el gradiente de armonía y eventualmente se aproximará $x = 1$, el atractor del “error”. Si la cantidad de ruido T que se agrega a la parte determinística de la ecuación es pequeño, será difícil que haya valores que estén atraídos hacia el atractor “correcto” y superar “hacia atrás” el valle entre los máximos locales.

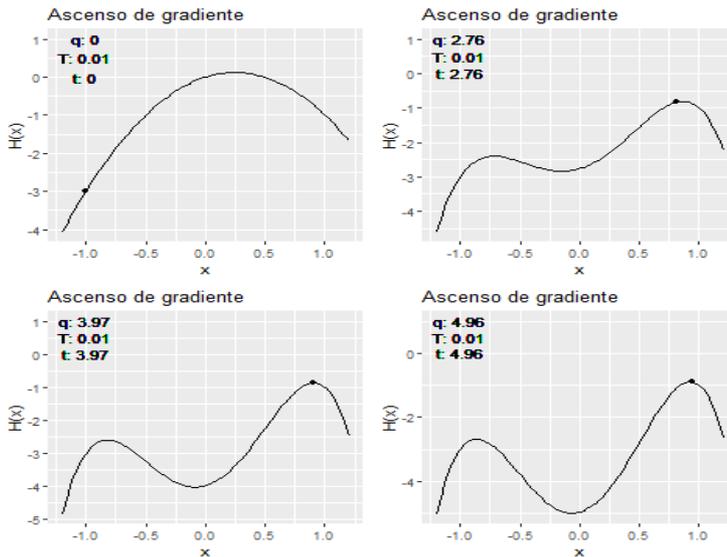


Figura 4. Ascenso de gradiente de armonía con sesgo hacia el atractor “error” ($e = 1$); según $q = 0$ (arriba izquierda), $q = 2.76$ (arriba derecha), $q = 3.97$ (abajo izquierda), $q = 4.96$ (abajo derecha).

Resultados

Globalmente, el patrón de error fue: JAKO < SONIA < MIRKA < NATI. El porcentaje de error en todo el corpus fue del 24.7 %. En primera instancia se presentan los resultados de la simulación con el modelo competitivo de Lokta-Volterra. La Tabla 1 muestra: (i) las proporciones globales de “error” entre la simulación y los datos; (ii) los p-valores de un test asintótico de diferencia entre proporciones entre los errores simulados y los errores de los datos (se debería observar que $p > .05$); (iii) los valores de la distancia de Minkowski $D = \sqrt[p]{(x_i - y_i)^p}$ ($p = 3$), entre los errores simulados y reales para la serie temporal de errores por sesiones (a menor distancia, mejor).

Se observa que SONIA y MIRKA subestiman el error global y JAKO lo sobreestima. Sin embargo, las diferencias entre las proporciones de error no resultan significativas. En cuanto al patrón por sesiones, SONIA retiene el mejor desempeño; MIRKA, el peor; y NATI y JAKO se aproximan de igual modo a las series de tiempo de los datos. El error de simulación total promedio es $\overline{error} = .245$, muy cercano al real. Por otro lado, la distancia promedio arroja $\overline{D} = 11.301$.

Tabla 1. Resultados del modelo competitivo de Lokta-Volterra

Sujeto	Simulación	Datos	P-valor	D
Sonia	.176	.215	.267	6.495
Nati	.349	.336	.763	8.604
Jako	.196	.164	.231	8.631
Mirka	.259	.290	.206	21.475

En lo que atañe al modelo basado en dinámica del replicador, en la Tabla 2 se observa que SONIA y MIRKA subestiman al error y que JAKO lo sobreestima. No obstante, las diferencias entre las proporciones de error no arrojan significatividad. En lo que respecta al patrón por sesiones, la distancia D aumenta con el nivel. SONIA resulta ser la de mejor desempeño; MIRKA obtiene el peor; y NATI / JAKO aproximan las proporciones de error por sesiones de manera similar. El error de simulación total promedio es $\overline{error} = .241$, muy cercano al real. Por otro lado, la distancia promedio arroja $\overline{D} = 12.961$.

Tabla 2. Resultados del modelo basado en dinámica del replicador

Sujeto	Simulación	Datos	P-valor	D
Sonia	.176	.215	.267	.686
Nati	.333	.336	1.000	9.929
Jako	.178	.164	.662	2.218
Mirka	.279	.290	.677	22.011

En lo concerniente al modelo basado en ascenso de gradiente de armonía, la Tabla 3 muestra: (i) la media de las proporciones de error extraídas de $n = 50$ simulaciones; (ii) su error estándar ($ee = s/\sqrt{n}$); (iii) la proporción de error en los datos; (iv) la frecuencia relativa de veces (en 50 simulaciones) en las que los p-valores de un test asintótico de diferencia entre proporciones entre los errores simulados y los errores de los datos son menores a .05 (o sea que habría diferencia significativa entre el patrón global simulado y el de los datos); (v) la distancia media D para medir la distancia entre el patrón por sesiones simulado y real. En el patrón global, las simulaciones que más se acercan, en media, son las de JAKO y MIRKA. Se observa que las proporciones simuladas varían muy poco. En general, no hay diferencias significativas entre las proporciones simuladas y las reales. En el patrón por sesiones, SONIA es la que mejor aproxima el patrón real. Los otros tres aprendientes evidencian un desempeño similar. Notar que este modelo es el que mejor aproxima el patrón por sesiones para MIRKA.

Tabla 3. Resultados del modelo de ascenso de gradiente de armonía

Sujeto	Simulación (media)	Simulación (error est.)	Datos	Frec. (p<.05)	D (media)
Sonia	.183	.0023	.215	.020	7.670
Nati	.370	.0024	.336	.000	12.854
Jako	.177	.0015	.164	.000	11.877
Mirka	.279	.0016	.290	.000	13.846

Conclusiones

Si cada concordancia fuera un “individuo” que toma decisiones de acuerdo con la dinámica de los modelos especificados; entonces, a partir del conjunto de dichas decisiones emerge un patrón global de error y otro por serie de tiempo. Es decir que los patrones de aprendizaje se autoorganizan y emergen de la dinámica ejercida por los individuos. En general, los modelos aproximan bien los datos globales de error reales. En cuanto a la aproximación a las series de sesiones reales, SONIA, NATI y JAKO obtienen un mejor desempeño con el de Lokta-Volterra, y MIRKA alcanza un resultado de menor distancia media (D) con el modelo de ascenso de gradiente. Comparativamente, MIRKA resultó la de peor desempeño en todos los modelos. Todos los modelos dinámicos analizados consideraban al error como atractor que se alcanzaba (o no) en el último paso temporal. Es decir que el error se trató como una representación continua y solamente al final, en el atractor (o cerca de este), como una representación discreta. Por lo tanto, «Error» o «Correcto» son propiedades emergentes de los sistemas dinámicos, al igual que los patrones globales de error agregados y por series temporales. Asimismo, la cognición no es vista como una maquinaria representacional que procesa información y está guiada por una sintaxis computacional. En cambio, se adscribe a una concepción en términos de

un flujo continuo de procesos. La mayoría de los formalismos de la teoría lingüística (minimalismo, gramáticas a unificación, etc.) están pensados para la sincronía, donde las distribuciones de probabilidad de los patrones se han estabilizado. No parece útil aplicarlos a la adquisición, donde la lengua se encuentra en flujo constante. Las ciencias de la complejidad iluminan otro camino: pasar de teorías que manipulan símbolos discretos a otra donde lo importante es un flujo de estados “mezcla” con patrones/configuraciones que modifican sus distribuciones de probabilidad hasta alcanzar atractores discretos estables. Si se considera el lenguaje como un flujo no lineal, continuo, fractal y adaptativo, entonces las técnicas de análisis deben pasar por las series de tiempo (no lineales), el análisis de la variabilidad y cambios de régimen, los estudios longitudinales y las simulaciones.

Referencias

- Brehm, L., Cho, P. W., Smolensky, P., & Goldrick, M. (2022). PIPS: A Parallel Planning Model of Sentence Production. *Cognitive Science*, 46(2). <https://doi.org/10.1111/cogs.13079>
- Cho, P. W., Goldrick, M., & Smolensky, P. (2017). Incremental Parsing in a Continuous Dynamical System: Sentence Processing in Gradient Symbolic Computation. *Linguistics Vanguard*, 3(1). <https://doi.org/1.1515/lingvan-2016-0105>
- Cho, P. W., Goldrick, M., Lewis, R., & Smolensky, P. (2018). Dynamic Encoding of Structural Uncertainty in Gradient Symbols. In A. Sayeed, C. Jacobs, T. Linzen & M. van Schijndel (eds.), *Proceedings of the 8th Workshop on Cognitive Modeling and Computational Linguistics* (pp. 19-28). Association for Computational Linguistics. <https://aclanthology.org/W19-29.pdf>
- Cho, P. W., Goldrick, M., & Smolensky, P. (2020). *Parallel Parsing in a Gradient Symbolic Computation Parser*. <https://files.osf.io/v1/resources/utcv/providers/osfstorage/5ea19ab935f4bf00cf1fd27f?action=download&direct&version=1>
- 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.
- Deo, A. (2015). The Semantic and Pragmatic Underpinnings of Grammaticalization Paths: The Progressive to Imperfective Shift. *Semantics and Pragmatics*, 8, 1-52.
- Kaplan, D., & Glass, L. (1995). *Understanding Nonlinear Dynamics*. Springer.
- Kilgarriff, A., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P., & Suchomel, V. (2014). The Sketch Engine: Ten Years on. *Lexicography*, 1(1), 7-36. <https://doi.org/10.1007/s40607-014-0009-9>
- Kot, M. (2001). *Elements of Mathematical Ecology*. Cambridge University Press.

- Nerbonne, J., Van Ommen, S., Gooskens, C., & Wieling, M. (2013). Measuring socially motivated pronunciation differences. In L. Borin & A. Saxena (eds.), *Approaches to Measuring Linguistic Differences* (pp. 107–140). Mouton. 10.1515/9783110305258.107
- MacWhinney, B. (2015). Introduction. In B. MacWhinney & W. O'Grady (eds.), *The Handbook of Language Emergence* (pp. 1-31). Wiley & Sons.
- MacWhinney, B. (2021). *The Childes Project: Tools for Analyzing Talk. Part 1 y 2*. TalkBank. <https://talkbank.org/manuals/CHAT.pdf>
- Marafioti, P. E. (2021). Factores influyentes en tipos de errores de concordancia nominal en un corpus de cuatro aprendientes italianos de español LE. *Lenguaje*, 49(2), 365-393. <https://doi.org/10.25100/lenguaje.v49i2.10915>
- 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). <https://revistas.unc.edu.ar/index.php/racc/article/view/32001>
- Murray, J. D. (2002). *Mathematical Biology I. An Introduction* (3rd ed.). Springer.
- Peña, D. (2002). *Análisis de datos multivariantes*. Mc Graw Hill.
- R Development Core Team. (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria. <http://www.R-project.org>.
- Smith, G., Franck, J., & Tabor, W. (2018). A Self-Organizing Approach to Subject-Verb Number Agreement. *Cognitive Science*, 42(S4), 1043-1074. <https://doi.org/1.1111/cogs.12591>.
- Smith, G., Franck, J., & Tabor, W. (2021). Encoding Interference Effects Support Self-Organized Sentence Processing. *Cognitive Psychology*, 124(101356). <https://doi.org/1.1016/j.cogpsych.202.101356>
- Smith, G., & Tabor, W. (2018). Toward a Theory of Timing Effects in Self-Organized Sentence Processing. In I. Juvina, J. Hout & C. Myers (eds.), *Proceedings of the 16th International Conference on Cognitive Modeling* (pp. 138-143). University of Wisconsin. <http://acs.ist.psu.edu/papers/ICCM%202018%20Proceedings.pdf>
- Smolensky, P, Goldrick, M., & Mathis, D. (2014). Optimization and Quantization in Gradient Symbol Systems: A Framework for Integrating the Continuous and the Discrete in Cognition. *Cognitive Science*, 38(6), 1102-1138. <https://doi.org/1.1111/cogs.12047>

- 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), 289-317. <https://doi.org/1.32614/RJ-2016-021>
- Van Buuren, S., & Groothuis-Oudshoorn, K. (2011). Mice: Multivariate Imputation by Chained Equations in R. *Journal of Statistical Software*, 45(3), 1-67. <https://doi.org/10.18637/jss.v045.i03>
- Weibull, J. W. (1997). *Evolutionary Game Theory*. MIT Press.