



10.15446/fyf.v37n1.104617

Artículos

## REGÍMENES LATENTES DE ERROR EN EL APRENDIZAJE DE LA CONCORDANCIA PLURAL EN ELE\*

## LATENT ERROR REGIMES IN THE LEARNING OF PLURAL AGREEMENT IN SFL

*Pablo Ezequiel Marafioti*<sup>1</sup>

### **Cómo citar este artículo:**

Marafioti, P. E. (2024). Regímenes latentes de error en el aprendizaje de la concordancia plural en ELE. *Forma y Función*, 37(1). <https://doi.org/10.15446/fyf.v37n1.104617>

Este es un artículo publicado en acceso abierto bajo una licencia Creative Commons.

Recibido: 2022-09-04, aceptado: 2023-10-10

---

\* Este artículo se enmarca en el proyecto de tesis de doctorado del autor: «El error de concordancia plural en español L2 desde una perspectiva emergente».

<sup>1</sup> ORCID <https://orcid.org/0000-0001-7299-5324> Universidad Nacional de Córdoba, Córdoba, Argentina, [kenavo38@yahoo.com](mailto:kenavo38@yahoo.com)

## Resumen

Se analiza la dinámica de tres regímenes ocultos de aprendizaje en la producción de concordancia nominal de número en un corpus de cuatro aprendientes italofonos de español LE, en un estudio de caso longitudinal. En el marco de la teoría de los sistemas complejos, se aplicó un modelo bayesiano de descubrimiento de tópicos y un modelo de Markov oculto. La dinámica tendía a permanecer en el régimen asociado al atractor «correcto» (o regresar eventualmente a este) con las concordancias: (1) con controlador concreto, familiar y animado; (2) aplicando estrategias facilitadoras de aprendizaje; (3) con TYPES más frecuentes; (4) con «e» epentética; (5) formadas por más de dos miembros. En cambio, tendía a alejarse del atractor «correcto» con aquellas: (1) a larga distancia; (2) con modificadores que no eran artículos definidos; especialmente en los niveles bajos de competencia (A, B1). La simulación se aproximó a los datos empíricos, pero subestimó las probabilidades de emisión.

**Palabras clave:** *segunda lengua; sistemas dinámicos; concordancia nominal; modelo de Markov oculto; plural.*

## Abstract

The dynamics of three hidden learning regimes in the production of nominal number agreement are analyzed in a corpus of four Italian learners of Spanish FL, in a longitudinal case study. Within the framework of complex systems theory, a Bayesian topic discovery model and a hidden Markov model were applied. The dynamics tended to remain in the regime associated with the «correct» attractor (or eventually return to it) with agreement instances: (1) with concrete, familiar and animated controller; (2) when facilitating learning strategies are applied; (3) with more frequent TYPES; (4) with epenthetic «e»; (5) with more than two members. Instead, it tended to move away from the «correct» attractor with: (1) long distance agreement; (2) instances with modifiers that were not definite articles; in particular at low levels of proficiency (A, B1). The simulation was close to the empirical data, but the emission probabilities were underestimated.

**Keywords:** *second language; dynamical systems; nominal agreement; hidden markov model; plural.*

## I. Introducción

El objetivo principal de este trabajo consiste en detectar regímenes de aprendizaje de la producción oral de concordancia nominal plural en cuatro aprendientes itálofonos de ELE (español como lengua extranjera). Además, se buscará evaluar algunos factores que inciden en la dinámica de la transición entre dichos regímenes.

Aunque el nivel de competencia en ELE provoca una disminución en los errores de género y número, los primeros tienden a persistir incluso luego de muchos años de práctica de producción oral (Muñoz Liceras, Díaz & Mongeon, 2000; Franceschina, 2001; White et al., 2004). La concordancia de género y número del artículo resulta más fácil de adquirir que la del adjetivo tanto en producción como en procesamiento (White et al., 2004; Montrul et al., 2008; Alarcón, 2011; Gillon Dowens et al., 2010). En particular, la concordancia de género parece ser más fácil de producir y procesar (en hablantes de L1 inglesa): (1) con masculino respecto al femenino, ya que las formas de masculino se utilizan en contextos femeninos, o sea, como *defaults* (White et al., 2004; Montrul et al., 2008; McCarthy, 2008; Alarcón, 2011); (2) con controladores<sup>1</sup> de morfología transparente (en «-o»/«-a»: «libro», «maestra») respecto a los menos transparentes (en «-e»: «el puente», «la calle»; en consonante: «el camión», «la canción»; u opuestos: «la mano», «el tema»), en el orden de facilidad: consonante > «-e» > opuestos (Montrul et al., 2008; Alarcón, 2011); (3) con controladores inanimados («hospital») respecto de los animados en correspondencia con el sexo biológico, como en «doctor/a» (Sagarra & Herschensohn [2013]; aunque véase también Alarcón [2009], que halló el efecto contrario en núcleos de SN complejos); (4) con SN complejos del tipo «N1 de N2» («la pared de la cocina está sucia»), cuando el género de N1 coincide con el N2 (Foote, 2015). Por otra parte (y también para anglófonos), con el aumento de distancia estructural (cantidad de nodos sintácticos entre controlador y objetivo) disminuye la sensibilidad a las violaciones de género y la concordancia se procesa más lentamente (Sagarra, 2007; Lichtman, 2009; Keating, 2009, 2010; Foote, 2011; Gillon Dowens et al., 2010). Asimismo, dada suficiente competencia gramatical, el procesamiento del género y del número resulta similar al de los nativos; sin importar si el rasgo está presente o no en la L1 (Alemán Bañón, Fiorentino & Gabriele, 2014). A pesar de ello, el hecho de que se encuentren diferencias de «profundidad» o «esfuerzo» de procesamiento entre dichos rasgos sí depende de si el rasgo se halla en la L1; y, por lo tanto, que los aprendientes puedan procesarlo «mejor» en la L2 reclutando rutinas ya presentes su L1 (inglés y chino [Gillon Dowens et al., 2010, 2011]). Incluso los principiantes logran detectar violaciones de número en el ámbito nominal de modo similar a los nativos cuando el rasgo está presente en L1 y L2 (Gabriele et al., 2021). Por otra parte, Gonzá-

lez et al. (2019) encontraron efectos significativos de aumento de errores de género y número en hablantes de holandés: (1) del plural respecto del singular, (2) del femenino respecto del masculino, (3) en los artículos femeninos (sin importar el rasgo de número). Para italofonos, Marafioti (2021, 2022) halló que los siguientes factores se asociaban de modo significativo tanto a errores de concordancia (chance) como al tiempo hasta que se producía un error (riesgo): adjetivos indefinidos (respecto de los artículos definidos), alta cantidad de errores cometidos hasta la instancia de concordancia, controlador de la concordancia animado (respecto de los inanimados). Por otro lado, la chance/riesgo de error bajaba con: adjetivos calificativos, tipos de concordancia con frecuencia alta en corpus, controladores nominales familiares y/o frecuentes, concordancias que requerían la inserción de «-e-» epentética en ambos términos, con estrategias de aprendizaje facilitadoras para los italofonos (ver más abajo).

En este trabajo se considera al lenguaje como un sistema dinámico, es decir, como sistema que cambia en el tiempo de aprendizaje. Dicho sistema está formado por un conjunto de componentes (subsistemas) que interactúan entre sí para generar un estado particular en un punto del tiempo. Se trata de un proceso iterativo, o sea que el sistema es afectado por flujos provenientes del «ambiente cognitivo» y por el estado previo del sistema. El «ambiente cognitivo» o «contexto» es el origen de la información (recursos) necesaria para que el sistema lleve a cabo la dinámica. Sin embargo, dichos recursos son escasos. Existe un flujo constante de información del ambiente hacia el sistema y viceversa. Esta «apertura» implica que el contexto mismo forma parte del sistema y contribuye a su complejidad. Asimismo, el sistema ajusta de modo constante su respuesta a los cambios del «ambiente cognitivo»; o sea, se adapta a transformaciones del contexto. Además, el cambio se concibe como no lineal: no es constante ni proporcional al input recibido. Esto implica que el aprendizaje no puede concebirse desde la metáfora de una «escalera» transitada en estadios de menor a mayor complejidad. Se produce un incremento de complejidad por medio de la creación de nuevas formas que no se hallaban codificadas en las condiciones iniciales. En consecuencia, el sistema se auto-organiza generando patrones más complejos. Esto se realiza al costo de consumir información disponible en el «ambiente cognitivo». Si bien el input es importante, el aprendiente no lo percibe de manera pasiva. Más bien este se encuentra con oportunidades de aprendizaje (*affordances*): una potencialidad para aprender mientras se halla inmerso en actividades comunicativas. Por ende, un input facilitador (frecuente, saliente, no ambiguo) no necesariamente se adquirirá antes, a no ser que el aprendiente vea en este potencialidades para la acción (Larsen Freeman & Cameron, 2008; Marafioti, 2020).

Al aprender una lengua, el aprendiente crea variabilidad, es decir, 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. A medida que transcurre el tiempo el sistema del lenguaje se adapta a los contextos y, si se producen cambios cualitativos en la dinámica (bifurcaciones [por ej., un aumento de competencia lingüística]), los errores disminuyen. Colectivamente emerge un patrón global de error. Es decir que el error se constituye como un fenómeno emergente. La dinámica se representa en un espacio de fase (*phase space*) donde se hallan todos los posibles estados del sistema, correspondiendo cada estado a un único punto *m*-dimensional en dicho espacio. Un «atractor» es una región 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». Aquí se considerará un atractor asociado a «correcto» y otro asociado al «error» de concordancia. Este enfoque se halla centrado en el sujeto: cada aprendiente debe descubrir su propia trayectoria de aprendizaje, la cual está surcada de variabilidad de diferente intensidad entre los diferentes subsistemas del lenguaje. Un estudiante puede desatender un determinado subsistema (por ejemplo, la morfología) produciendo allí más errores/variabilidad, para concentrarse en otro (por ejemplo, la complejidad sintáctica). Entonces el nivel de investigación será el del individuo a través del análisis de series de tiempo.

## 2. Metodología

En el apartado 2.1. se detalla la creación experimental del Corpus a partir de cuatro informantes; así como (1) la variable asociada al error y (2) aquellas que caracterizan a cada instancia de concordancia. El apartado 2.2. hace uso de (1) para implementar las series temporales a través de «motivos» y se explica el método LDA (*Latent Dirichlet Allocation*). En el apartado 2.2. se introduce el modelo de Markov oculto que utiliza como insumos a las series temporales como variable respuesta y a las variables de (2) como predictoras que influyen las probabilidades de transición entre los estados ocultos del modelo.

### 2.1. Participantes y diseño del corpus

Se utilizó un corpus de adquisición de cuatro alumnos adultos, de lengua nativa italiana, estudiantes del Instituto Cervantes de Milán en el año académico 2008/09. Cada

alumno poseía un nivel distinto de competencia lingüística según el *Marco Común Europeo de Referencia*. Se hicieron entrevistas de 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 Mac Whinney (2021). El corpus estaba constituido por los siguientes conjuntos de transcripciones (nombres ficticios): 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. Siempre estaban formadas por dos términos, es decir que, si hubiere más de dos, por ej., «las casas blancas», se codificaron dos ejemplos: «las casas» y «casas blancas». Se clasificaron los errores como variable binaria en ausencia («0») o presencia («1»). Se tuvieron en cuenta las siguientes variables que caracterizaban cada instancia de concordancia en español.

*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).

*FREQ.S.* La frecuencia del TYPE de concordancia de acuerdo con el corpus del español *EsTenTen* de *Sketch Engine* (Kilgarriff et al., 2014). 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). Se discretizó usando clustering por mezcla de gaussianas (Scrucca et al., 2016), en los niveles: 1 = frecuencia alta, 0 = frecuencia baja.

*EP.* Se especificó si en el controlador, en el objetivo, o en ambos, había una desinencia que requería la inserción de «e» epentética [«-(e)s»]. El razonamiento fue que realizar concordancia con dos operaciones de este tipo resulta más complicado que con una o con ninguna; según: 0 = sin «e» epentética; 1 = con «e» epentética en un término; 2 = con «e» epentética en ambos términos.

*FAM.LEX.* Índice a partir de PCA (*Principal Component Analysis* [Peña, 2002]) combinando los siguientes rasgos del controlador nominal extraídos de la base de

datos BuscaPalabras (Davis & Perea, 2005): (1) Familiaridad (FAM): índice subjetivo que indica cuán frecuentemente una palabra es oída, leída o producida diariamente; (2) Frecuencia (LEX): frecuencia de la palabra en el corpus BuscaPalabras, en escala por mil. Se discretizó el índice PCA mediante clustering por mezcla de gaussianas, en los niveles: 1 = alto, 0 = bajo.

*IMA.CONC.* Índice a partir de PCA combinando los siguientes rasgos del controlador nominal (base de datos BuscaPalabras): (1) Imaginabilidad (IMA): índice subjetivo que indica a la intensidad con la que una palabra evoca imágenes; (2) Concretud (CONC): índice subjetivo que indica cuán concreta es una palabra de menos (+ abstracta) a más (+ concreta). Se discretizó el índice PCA mediante clustering por mezcla de gaussianas, en los niveles: 1 = alto, 0 = bajo.

*GRAM.* Si la concordancia constaba de dos o más términos: 0 = dos; 1 = más de dos.

*LDA.* Si la concordancia entre controlador y objetivo se establecía a larga distancia: 0 = no, 1 = sí.

*AN.* Animicidad del controlador de la concordancia: 0 = inanimado, 1 = animado.

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 (0 en ambos). Dichas estrategias buscaron identificar casos que facilitaran la producción de concordancias<sup>2</sup>, y que además resultaron estar significativamente asociadas a una baja en la chance/riesgo de error en estudios anteriores (Marafioti, 2021, 2022).

- Estrategia 1 (*EST1*): si la palabra plural del italiano termina en «-i» poner en español plural en «-os»: «i libri» > «los libros».
- Estrategia 2 (*EST2*): si la palabra plural del italiano termina en «-e» poner en español plural en «-as»: «le case» > «las casas».
- Estrategia 5 (*EST5*): si la palabra singular del italiano termina en «-e», poner en español el plural en «-es». Por ejemplo, la palabra «sole» ('sol') podría ser la base para formar el plural español agregando «s»: «sole» > «soles». Se trata de casos en los cuales el español coincide con la aplicación del plural con «-e-» epentética.

A modo de 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 dos términos (GRAM = 0) de frecuencia alta (FREQ.S = 1) y a corta distancia (LDA = 0); el plural de «alemán» requiere insertar una «-e-» epentética (EP = 1), «alemán» conlleva una familiaridad/frecuencia léxica baja (FAM.LEX = 0),

familiaridad/concretud baja ( $\text{IMA.CONC} = 0$ ) y es animado ( $\text{AN} = 1$ ); no es aplicable ninguna estrategia ( $\text{EST}_1 = \text{EST}_2 = \text{EST}_5 = 0$ ).

## 2.2. Motivos y Latent Dirichlet Allocation

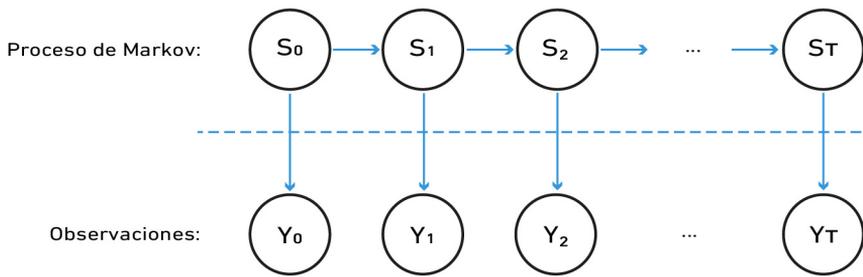
Los «motivos» son series de símbolos discretos, aquí pertenecientes al alfabeto:  $A = \{a, b\}$ . Para descubrirlos se corrió, para cada sesión, una ventana de largo tres a la secuencia de la variable binaria de error. Los motivos descubiertos fueron (entre corchetes la secuencia de serie de error): *aaa* (000), *aab* (001), *aba* (010), *abb* (011), *baa* (100), *bab* (101), *bba* (110), *bbb* (111). Se consideró que cada motivo representaba un «estado» diferente de dinámica, con dos atractores «a = correcto» y «b = error» según: (1) atractor «a» fuerte: *aaa*; (2) atractor «a» débil: *aab*, *aba*, *baa*; (3) atractor «b» fuerte: *bbb*; (2) atractor «b» débil: *bba*, *bab*, *abb*. Se interpretó «fuerte» como permanencia en el atractor y «débil» como un solo apartamiento del atractor hacia el otro. El objetivo consistió en clasificar las sesiones en regímenes de dinámica similar, según «estados» compartidos. Se consideró a cada sesión como un «texto» cuyas palabras eran los «motivos».

No se consideró la primera sesión de SONIA por tener solamente cuatro instancias, con lo cual hubo 51 sesiones/«textos». A continuación, se hizo un análisis para descubrir «tópicos», como si cada tópico fuera un régimen de dinámica diferente. Para ello se aplicó *Latent Dirichlet Allocation* (LDA [Blei et al., 2003]). Esta técnica bayesiana permite descubrir los tópicos de los que hablan los textos, asignando una probabilidad a cada tópico. En concreto, en LDA cada «texto» se representa como una mezcla de distribuciones sobre tópicos latentes (no observados) y cada tópico se caracteriza por una distribución asimétrica sobre palabras, en donde solo algunas palabras del vocabulario tendrán alta probabilidad. Se ajustó un modelo con tres tópicos. Para evaluar la agrupación encontrada, se formaron los grupos  $g_i [i=1,2,3]$  de referencia utilizando las frecuencias relativas de error ( $f$ ) de las sesiones:  $g_1 = \{f \leq 0.2\}$ ;  $g_2 = \{f \in [0.21, 0.35]\}$ ;  $g_3 = \{f > 0.35\}$ . Se calculó la tabla de confusión con los tópicos predichos en las filas y los grupos de referencia en las columnas. De ella se derivaron las medidas de desempeño de la clasificación.

## 2.3. Hidden Markov Models (HMM) y simulación

Para cada sujeto consideremos como variable respuesta  $Y_{i:T} = (Y_1, Y_2, \dots, Y_T)$  la serie temporal de «motivos» o «estados visibles» de la dinámica. En un modelo de Markov oculto (*Hidden Markov Model* o HMM) cada estado visible es generado por un proceso estocástico de estados ocultos:  $S_{i:T} = (S_1, S_2, \dots, S_T)$ . Para cada estado oculto existe una

distribución de probabilidad sobre los estados visibles (Figura 1). Un HMM de primer orden consiste en los siguientes componentes: (1) la variable respuesta  $Y_{t:T}$  con  $M$  niveles; (2) una variable categorial latente  $S_{t:T}$  con  $K$  niveles; (3) una matriz de transición  $A$  de orden  $K \times K$  que indica la probabilidad de pasar de un estado a otro que le sigue; (4) las probabilidades de emisión, o sea, la probabilidad de observar un determinado nivel de la variable respuesta cuando el proceso se encuentra en un determinado nivel de estado oculto; (5) la distribución inicial de probabilidades  $\pi$  para los  $k$  estados ocultos en  $t=1$  (Visser, 2011).



**Figura 1. Representación de un modelo de Markov oculto**

El modelo ajustado para cada sujeto fue el siguiente:

$$Y_t \sim \text{Multinomial}(1; p_1, \dots, p_M);$$

$$P(Y = m | S_t = i) = \frac{e^{\beta_{0,m}}}{\sum_{j=1}^M e^{\beta_{0,j}}}; m = 1, \dots, M$$

$$P(S_t = j | S_{t-1} = i) = \frac{e^{\eta_j}}{\sum_{j=1}^K e^{\eta_j}}; j = 1, \dots, K$$

$$\eta_j = \beta_0 + \beta_1 LDA + \beta_2 EP_1 + \beta_3 EP_2 + \beta_4 GRAM + \beta_5 MOD_1 + \beta_6 MOD_2 + \beta_7 MOD_3 + \beta_8 IMA.CONC + \beta_9 FAM.LEX + \beta_{10} AN + \beta_{11} EST1 + \beta_{12} EST2 + \beta_{13} EST5 + \beta_{14} FREQ.S$$

Las respuestas se distribuyen como multinomial. Las predictoras descritas en el apartado anterior influyen la probabilidad de transición de un estado oculto a otro, pero no influyen la probabilidad de una determinada respuesta (dado un determinado estado oculto). Se calcularon: (1) las probabilidades de emisión y (2) las probabilidades de transición para el caso en que todas las predictoras valen cero y para el caso en el que cada predictora vale uno y las demás, cero. Notar que los subscripts en MOD y EP indican las indicadoras (cada variable cualitativa con  $l$  niveles está representada con  $l-1$

indicadoras). Si la influencia de una variable en la dinámica es positiva entonces dicha dinámica tenderá a permanecer en el régimen oculto asociado al atractor «correcto», o bien regresará eventualmente a este en caso de abandonar dicho régimen. En cambio, si la influencia es negativa, la dinámica tenderá a salir del régimen asociado al atractor «correcto» y a circular entre regímenes cercanos al atractor «error». Se hipotetiza que las variables de influencia positiva (cuando valen uno) indican concordancias: (1) con controlador más concreto y/o imaginable (IMA.CONC); (2) con controlador más familiar y/o frecuente (FAM.LEX); (2) en donde se puedan aplicar estrategias facilitadoras de aprendizaje (EST1, EST2, EST5); (3) cuyos TYPES sean más frecuentes (FREQ.S). En cambio, ejercerían influencia negativa las concordancias: (1) a larga distancia (LDA); (2) con «-e-» epentética de uno o ambos miembros (EP); (3) aquellas que tienen modificadores que no sean artículos definidos (MOD); (4) las que tengan controladores animados (AN); (5) aquellas que consten de más de dos miembros (GRAM).

Además, para cada sujeto, se simularon las respuestas (motivos) a partir de los modelos ajustados. Para ello: (1) se calcularon las matrices de transición teniendo en cuenta los valores de las predictoras para cada concordancia; (2) se generó el estado oculto correspondiente; (3) se usaron las probabilidades de emisión para generar la respuesta dado el estado oculto elegido. Se comparó la distribución empírica de motivos de cada aprendiente con aquella simulada mediante la medida de divergencia de Kullback-Leibler (que idealmente sería cero si la distribución empírica y la simulada fueran iguales). El material suplementario y el código de R se puede consultar en Github:

[https://github.com/pablomarafioti/PabloMarafioti/tree/master/articulo\\_HMM](https://github.com/pablomarafioti/PabloMarafioti/tree/master/articulo_HMM)

Dentro del enfoque de los sistemas complejos en L2, Chan et al. (2015) utilizaron HMM para revelar momentos de auto-organización en series de tiempo de medidas de complejidad sintáctica en dos aprendientes gemelas chinas de L2 de nivel inicial, tanto en modalidad oral como escrita. Hallaron que la complejidad sintáctica se desarrolló antes en lenguaje oral respecto del escrito y que los patrones de aprendizaje observados para cada sujeto eran diferentes, a pesar de las similitudes en el contexto de aprendizaje de las gemelas.

### 3. Resultados

#### 3.1. Latent Dirichlet Allocation (LDA)

A continuación, se muestran la matriz de confusión (Tabla 2) y las medidas derivadas (Tabla 1). La precisión global fue de 0.882 (C.I. 95%: [0.761; 0.956]) y el coeficiente Kappa de 0.82. Con respecto a las medidas por clase, todas se hallan por encima de 80%.

Por otra parte, el área bajo la curva (AUC) de una gráfica de la sensibilidad versus las falsas alarmas (1-especificidad) arroja:  $AUC(1,2) = 0.887$ ;  $AUC(1,3) = 0.997$ ;  $AUC(2,3) = 0.954$ . Ya que  $AUC = 1$  indica clasificación perfecta, se nota que el modelo distingue menos entre los tópicos 1 y 2.

**Tabla 1. Medidas de desempeño de la predicción**

	Tópico 1	Tópico 2	Tópico 3
Sensibilidad	0.9474	0.8261	0.8889
Especificidad	0.8750	0.9286	1
Valor predictivo positivo	0.8182	0.9048	1
Valor predictivo negativo	0.9655	0.8667	0.9767
F1	0.8780	0.8636	0.9412
Precisión balanceada	0.9112	0.8773	0.9444

**Tabla 2. Matriz de confusión**

	Grupo 1	Grupo 2	Grupo 3
Tópico 1	18	4	0
Tópico 2	1	19	1
Tópico 3	0	0	8

Se muestran gráficamente los tópicos (Figura 2), indicando las frecuencias de los «motivos» que los componen (en rojo: frecuencia absoluta; en celeste: total). Como el tópico 1 coincide con proporciones de error menores al 20%, se observa que la mayor cantidad de «motivos» tienen que ver con el atractor «a» (aaa, aba). En este grupo, el estado fuerte del atractor «correcto» es el que tiene más instancias. En el otro extremo se encuentra el tópico 3, el atractor «b» se asocia a probabilidades de error mayores al 35%, los estados «bbb» (atractor «error» fuerte), «bba», «abb», «bab» (atractor «error» débil) registran más frecuencia comparados con los otros dos tópicos y el atractor «aaa» resulta el de menos frecuencia en comparación. Por último, el tópico 2, asociado a proporción de errores de 21 a 35%, registra más frecuencia en los estados débiles del atractor «a» (baa, aab, aba) comparado con el tópico 1. Es decir que se evidencia una dinámica intermedia. En resumen, los tópicos logran captar tres tipos de dinámica asociadas a franjas de probabilidad de error de las sesiones.

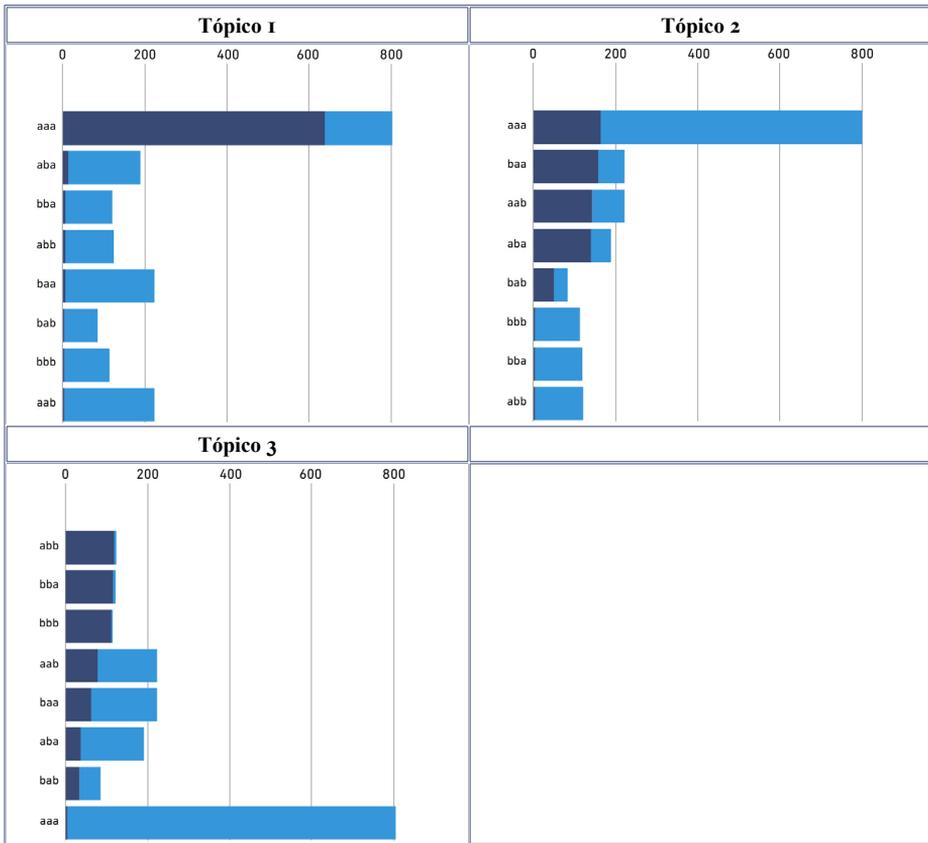


Figura 2. Tópicos y motivos que los componen

La Tabla 3 muestra a las sesiones de cada alumno agrupadas por tópico de dinámica similar. SONIA y JAKO tienen más sesiones en el tópico 1 (error < 0.20) y solamente una en el 3 (error > 0.35). En contraste, NATI y MIRKA tienen la mitad o más de las sesiones en el tópico 3 y como máximo dos en el tópico 2. Es decir que tanto SONIA como JAKO están en el atractor «correcto», y luego de una perturbación vuelven a este.

Tabla 3. Sesiones según tópicos de dinámica similar

Alumno	Tópico 1	Tópico 2	Tópico 3
SONIA	3, 5, 7, 8, 9, 10, 12	2, 4, 11	6
NATI	4, 14	3, 9, 12, 13	1, 2, 5, 6, 7, 8, 10, 11
JAKO	1, 2, 3, 5, 6, 7, 9, 11, 12, 13, 14	4, 8	10
MIRKA	1, 4, 6, 8	3, 5	2, 7, 9, 10, 11, 12

### 3.2. Modelo de Markov oculto

La Figura 3 muestra las probabilidades de emisión para cada estado observado en un determinado estado oculto, según sujeto.

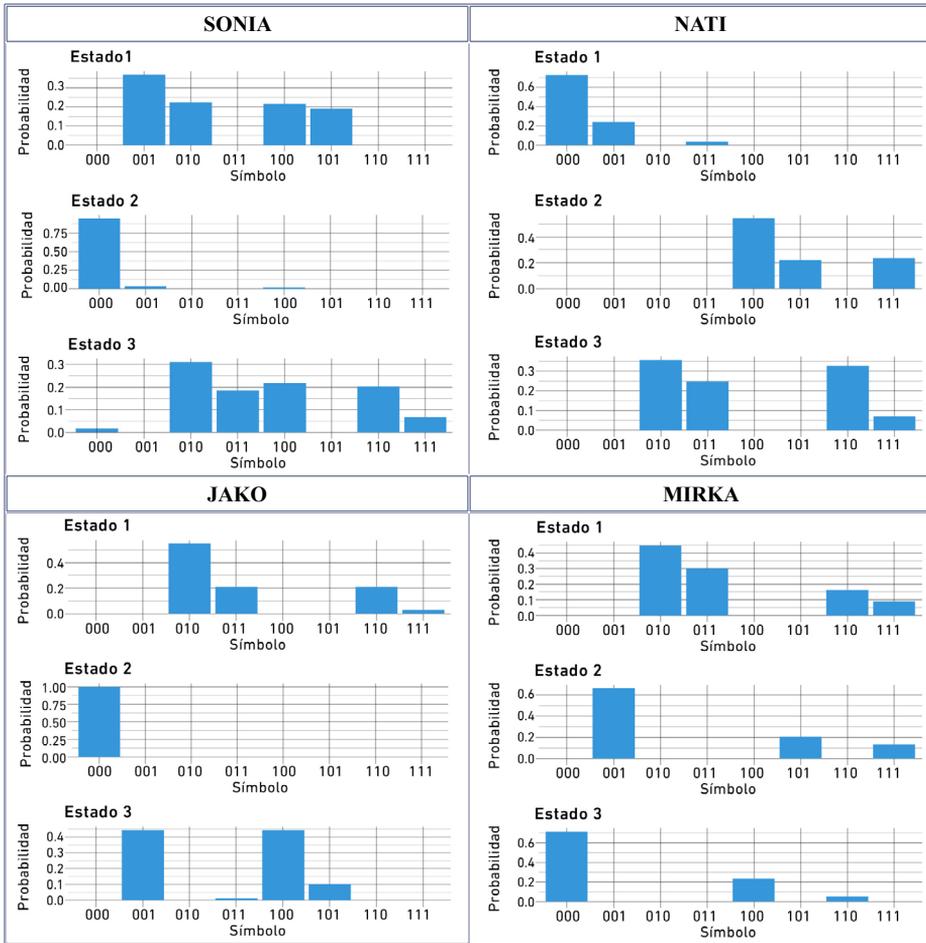


Figura 3. Probabilidades de emisión

En SONIA, el estado oculto 2 está dominado por el atractor «correcto» ('000'). El estado oculto 1 contiene mayor probabilidad para estados pertenecientes a un atractor «correcto» más débil ('001', '010', '100'). El estado oculto 3 apunta a una dinámica mixta entre el atractor «error» y «correcto». En NATI el atractor «correcto» domina el estado oculto 1, hallándose los otros dos estados en una dinámica mixta con presencia creciente de estados observados del atractor «error» desde el estado oculto 2 al 3. JAKO tiene

el estado oculto 2 completamente en el atractor «correcto» y el estado 3 dominado por estados observados cercanos a aquel. El estado 1 acusa una dinámica mixta entre ambos atractores. En MIRKA domina el atractor «correcto» en el estado oculto 3; siendo los otros dos de dinámica mixta, pero con más presencia de estados observados asociados al atractor de «error» en el estado oculto 1. En suma, todos los sujetos acusan un estado oculto más cercano al atractor «correcto»: el segundo para SONIA y JAKO, el primero para NATI y el tercero para SONIA. Los otros dos estados ocultos son en general de dinámica intermedia en un gradiente de estados observados cada vez más cercanos al atractor de «error».

En la Figura 4 se indican las probabilidades de transición cuando todas las predictoras valen cero (probabilidades redondeadas a dos decimales; el cero indica:  $p < 0.0001$ ; el uno indica:  $p > 0.9999$ ). En SONIA, la dinámica determina que hay probabilidad alta de permanecer en el estado oculto 2 (asociado al atractor «correcto»); pero si se deja dicho estado se pasa al estado de 1 (atractor «correcto» más débil) para luego pasar al estado oculto 3 (influenciado por estados asociados al atractor «error») pero dejándolo inmediatamente para volver al estado oculto 2. La dinámica está dominada por estados cercanos al atractor «correcto». En NATI, la probabilidad de permanecer en el atractor «correcto» (estado oculto 1) es más baja. Si se lo deja, se pasa al estado oculto 3, con predominio de estados asociados al atractor «error»; habiendo una pequeña probabilidad de permanecer incluso en dicho estado. Abandonándolo de inmediato, se pasa al estado oculto 2, con menos influencia de estados asociados al atractor «error». O bien hay una probabilidad baja de volver al estado oculto 1, o bien se retorna más probablemente al estado oculto 3. Por ende, la dinámica está dominada por estados ocultos mixtos, más asociados al atractor «error». En JAKO la probabilidad de quedarse en el estado oculto 2 (atractor correcto «000», sin errores) es muy alta. Pero en el evento de dejarlo, se pasa al estado oculto 3, igualmente dominado por estados asociados al atractor «correcto». Existe una probabilidad moderada de volver al estado oculto 2; y una más alta de pasar al estado oculto 1, con baja influencia de estados asociados al atractor «error»; pero regresando de inmediato al estado oculto 3. En MIRKA hay una probabilidad moderada de permanecer en el estado oculto 3, donde domina el atractor «correcto». Dejándolo, se pasa al estado oculto 2, el cual es abandonado de inmediato para pasar al estado oculto 1, cuyos estados se asocian más al atractor «error». De allí lo más probable es retornar al estado oculto 3. Notar que en todos los casos nunca se pasa directamente del estado oculto más asociado al atractor «correcto» al estado oculto más asociado al atractor «error» (y la probabilidad de permanencia en este último es muy baja).

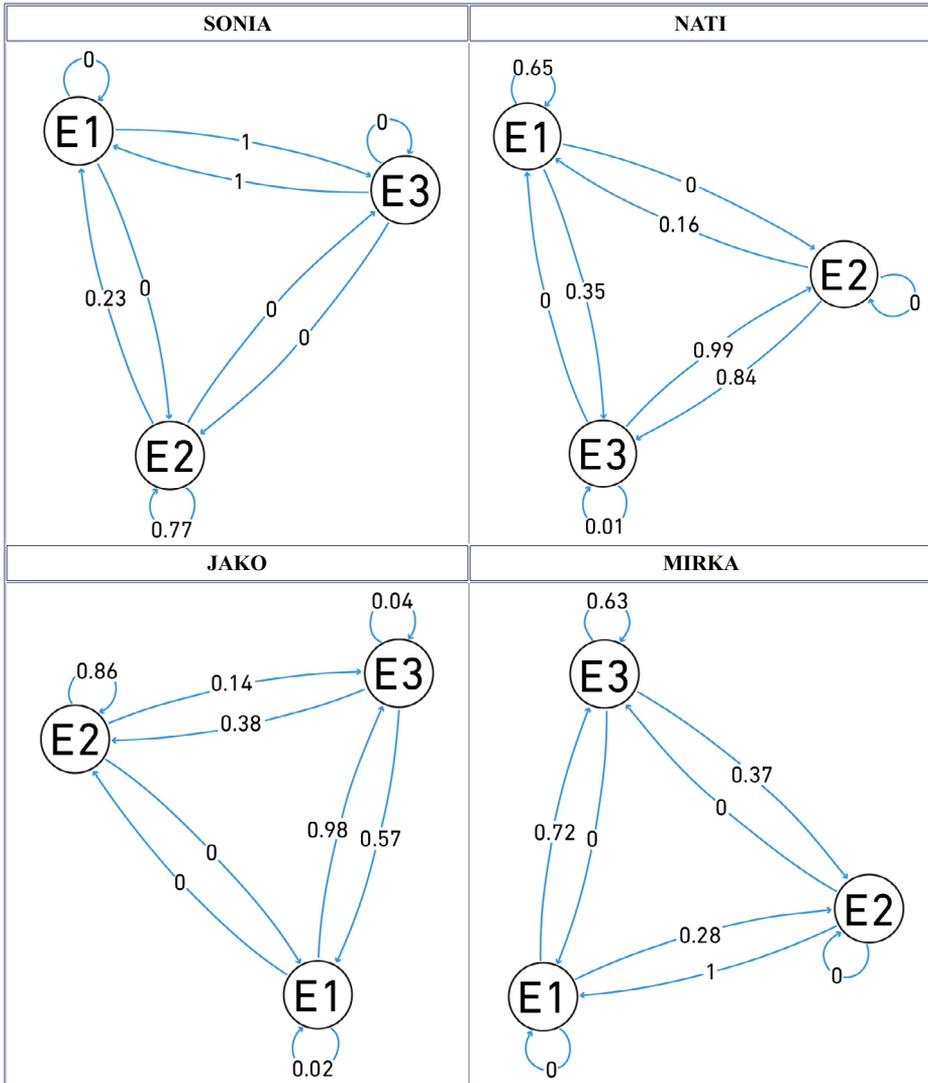


Figura 4. Probabilidades de transición cuando todas la predictoras valen cero

A continuación, se examinarán las probabilidades de transmisión cuando una determinada predictora vale uno y las demás valen cero. Esto permite observar el efecto de cada predictora en la dinámica: (1) ¿qué variables tienden a quedarse en el estado oculto más asociado al atractor «correcto»?; (2) si se deja dicho estado, ¿qué sucede con la dinámica?; (3) ¿cuáles variables están más influenciadas por el atractor «error»? La Tabla 4 muestra las probabilidades de transmisión para SONIA (en las Tablas 4 a 7

se redondea a dos decimales; el cero indica siempre una probabilidad bajísima pero no nula: < 0.0001; el uno indica probabilidad altísima: > 0.9999).

**Tabla 4. Probabilidades de transición para SONIA, según predictoras**

	1 → 1	1 → 2	1 → 3	2 → 1	2 → 2	2 → 3	3 → 1	3 → 2	3 → 3
LDA	0	0	1	0	0	1	0	0	1
EP1	1	0	0	0.01	0.99	0	0	1	0
EP2	1	0	0	1	0	0	1	0	0
GRAM	0	0	1	0.03	0.97	0	0	0	1
MOD2	0	0	1	0.63	0.37	0	1	0	0
MOD3	1	0	0	0.62	0.38	0	0	0	1
IMA.CONC	1	0	0	0.43	0.57	0	1	0	0
FAM.LEX	0	0	1	0.36	0.64	0	1	0	0
AN	0	0	1	0.01	0.99	0	1	0	0
EST1	0	1	0	0.08	0.92	0	0	1	0
EST2	1	0	0	0.01	0.99	0	0	1	0
EST5	1	0	0	0	1	0	1	0	0
FREQ.S	0	0	1	0.23	0.77	0	0	0	1

Cuando las concordancias son a larga distancia (LDA = «1»), toda la dinámica va hacia el estado oculto 3, dominado por el atractor «error». La alta probabilidad de permanecer en el estado oculto 2 (relacionado con el atractor «correcto») se asocia a concordancias: (1) con un miembro con «e» epentética (EP = «1»), (2) con más de dos términos (GRAM = «1»), (3) con controlador animado (AN = «1»), (4) con las tres estrategias de aprendizaje (EST1 = EST2 = EST5 = «1»). Las concordancias cuyo TYPE posee alta frecuencia también acusan una probabilidad relativamente alta de permanecer en el estado oculto 2. Sin embargo, si se lo deja se va hacia el estado oculto 1 y de allí inmediatamente al estado oculto 3 (más asociado al atractor «error») y se permanece allí (GRAM exhibe la misma dinámica, pero con una probabilidad muy baja de salir del estado oculto 2). Las concordancias con controladores con más imaginabilidad y/o concretud (IMA.CONC = «1») y aquellas con controladores más familiares y/o frecuentes (FAM.LEX = «1») poseen una probabilidad moderada de permanecer en el estado oculto 2. No obstante no es baja la probabilidad de pasar del estado oculto 2 al 1 (asociado a un atractor «correcto» débil) y de permanecer allí. Para las concordancias con determinantes (MOD = «2») y adjetivos (MOD = «3») resulta más probable pasar del estado oculto 2 al 1 que permanecer en el 2. Notar que para las concordancias cuyos dos miembros contienen

«-e-» epentética (EP = «2») la dinámica siempre conduce al estado oculto 1. En suma, a excepción de LDA, las demás variables se hallan relacionadas a una dinámica cuyos estados tienen que ver con el atractor «correcto» en mayor o menor medida (estados ocultos 1 y 2). La Tabla 5 muestra las probabilidades de transición para NATI.

**Tabla 5. Probabilidades de transición para NATI, según predictoras**

	1 → 1	1 → 2	1 → 3	2 → 1	2 → 2	2 → 3	3 → 1	3 → 2	3 → 3
LDA	0.81	0	0.19	0.33	0	0.67	0	0	1
EP1	0.69	0	0.31	0.04	0	0.96	0	0.08	0.92
EP2	1	0	0	1	0	0	0	1	0
GRAM	0.63	0	0.37	0.97	0	0.03	0	1	0
MOD1	0	0	1	0	0	1	0	0	1
MOD2	0.42	0	0.58	0.01	0	0.99	0	1	0
MOD3	0.33	0	0.67	0.24	0	0.76	0	1	0
IMA.CONC	0.66	0	0.34	0.3	0	0.7	0	1	0
FAM.LEX	0.83	0	0.17	0.14	0	0.86	0	0.95	0.05
AN	0.54	0	0.46	0.04	0	0.96	0	0.08	0.92
EST1	0.92	0	0.08	0.36	0	0.64	0	0.86	0.14
EST2	0.86	0	0.14	0.1	0	0.9	0	0.58	0.42
EST5	0.89	0	0.11	0.44	0	0.56	0	1	0
FREQ.S	0.42	0	0.58	0.77	0	0.23	0	1	0

En NATI el atractor «correcto» está más asociado al estado oculto 1. Tienen probabilidades altas de permanecer en dicho estado las concordancias: (1) a larga distancia; (2) con controlador es más frecuente y/o familiar; (3) con las tres estrategias de aprendizaje. Notar que en el caso (1) hay una probabilidad no despreciable de que la dinámica se dirija al estado oculto 3, más asociado al atractor «error». En el caso de la «-e-» epentética, si esta se halla en los dos miembros la dinámica conduce al estado oculto 1. Pero si se encuentra en un solo miembro, hay posibilidad de que la dinámica vaya hacia el estado oculto 3. Tanto GRAM como IMA.CONC tienen probabilidades de permanecer en el estado oculto 1 que rondan 0.6. No obstante ello, en el primer caso si la dinámica sale de dicho estado es altamente probable que regrese. En contraste, si en IMA.CONC (y FAM.LEX) se deja el estado oculto 1 es más probable que la dinámica circule entre los estados oculto 2 y 3, influenciados por el atractor «error». Si el artículo es indefinido (MOD = «1») la dinámica siempre se dirige al estado oculto 3. Si se trata de un determinante o un adjetivo es más probable que la dinámica deje el estado oculto

1 y se pose en los estados 2 y 3, más asociados al atractor «error». La animicidad del controlador induce a que la dinámica pueda salir del estado oculto 1 y se dirija al 3 con muy alta probabilidad de permanecer allí. Por último, en los TYPES de alta frecuencia, si bien es más probable que se deje el primer estado oculto, igualmente la probabilidad de regresar a este es alta. En suma, las variables más asociadas al atractor «error» son MOD y AN; y en menor medida IMA.CONC y (EP = «1»).

La Tabla 6 muestra las probabilidades de transmisión para JAKO. Las probabilidades de permanecer en el atractor «correcto» («000»), coincidente con el segundo estado oculto, son altas para todas las variables y la probabilidad de pasar de allí al estado oculto 1, más asociado al atractor «error» son ínfimas. Las concordancias donde se puede aplicar la estrategia EST2, aquellas con controladores animados y cuyos modificadores son artículos indefinidos tienen una probabilidad del 0.25-0.3 de dejar el estado oculto 2 y de que luego la dinámica vaya hacia al estado oculto 3.

**Tabla 6. Probabilidades de transición para JAKO, según predictoras**

	1 → 1	1 → 2	1 → 3	2 → 1	2 → 2	2 → 3	3 → 1	3 → 2	3 → 3
LDA	0.84	0	0.16	0	1	0	1	0	0
EP1	0	0	1	0	0.87	0.13	0.55	0.44	0
EP2	0	0	1	0	1	0	0.49	0.5	0.01
GRAM	0.1	0	0.9	0	0.84	0.16	0.52	0.47	0
MOD1	0	0	1	0	0.71	0.29	0	0	1
MOD2	0.27	0	0.73	0	0.88	0.12	0.76	0.24	0
MOD3	0.28	0	0.72	0	0.85	0.15	0.64	0.24	0.11
IMA.CONC	0.02	0	0.98	0	0.93	0.07	0.32	0.1	0.59
FAM.LEX	0.03	0	0.97	0	0.94	0.06	0.35	0.62	0.03
AN	0.06	0	0.94	0	0.75	0.25	0.57	0.35	0.08
EST1	0.01	0	0.99	0	0.83	0.17	0.57	0.3	0.12
EST2	0.04	0	0.96	0	0.72	0.28	0.67	0.33	0
EST5	1	0	0	0	0.84	0.16	0.41	0.35	0.25
FREQ.S1	0.01	0	0.99	0	0.87	0.13	0.49	0.47	0.04

La Tabla 7 muestra las probabilidades de transmisión para MIRKA. En dicho sujeto el estado oculto más asociado al atractor «correcto» es el tercero. Las probabilidades de permanecer en dicho estado son altas para las concordancias: (1) a larga distancia; (2) con «e» epentética (EP = «1», «2»); (3) con más de dos miembros; (4) con artículos indefinidos o determinantes; (5) con TYPES frecuentes. Si la dinámica escapa al

estado oculto 3 entonces regresa al mismo estado, a excepción de los casos: (1) con «-e-» epentética en un término, en donde hay más probabilidad de posarse en el estado oculto 2; (2) con determinantes, en cuyo caso la dinámica se dirige al estado oculto 1, el más influenciado por el atractor «error». En el orden de probabilidad de 0.6 a 0.7 de permanecer en el estado oculto 3 se hallan las concordancias: (1) con adjetivos; (2) con controladores más frecuentes y/o familiares; (4) con controladores más concretos y/o imaginables; (5) con controladores animados. A excepción del caso (1), en donde dejar el estado oculto 3 implica ir hacia el estado oculto 1; en el resto la dinámica regresa al estado oculto 3 con más probabilidad. Las tres estrategias de aprendizaje tienen la probabilidad más baja de permanecer en el estado oculto 3. En el evento de dejarlo, en las estrategias EST1 y EST2 la dinámica tiende a quedarse en el estado oculto 2, más asociado al atractor «error»; y EST5 tiende a volver al 3. En suma, las variables más asociadas al atractor «error» son EST1, EST2, y, en menor medida (MOD = «2», «3»).

**Tabla 7. Probabilidades de transición para MIRKA, según predictoras**

	1 → 1	1 → 2	1 → 3	2 → 1	2 → 2	2 → 3	3 → 1	3 → 2	3 → 3
LDA	0	0.28	0.72	0.37	0	0.63	0	0.2	0.8
EP1	0	0.72	0.28	1	0	0	0	0.26	0.74
EP2	0	0.61	0.39	0	0	1	0	0	1
GRAM	0	0.15	0.85	1	0	0	0	0.26	0.74
MOD1	0	0.03	0.97	1	0	0	0	0.06	0.94
MOD2	1	0	0	1	0	0	0	0.16	0.84
MOD3	1	0	0	1	0	0	0	0.4	0.6
IMA.CONC	0	0.2	0.8	1	0	0	0	0.37	0.63
FAM.LEX	0	0.4	0.6	1	0	0	0	0.37	0.63
AN	0	0.32	0.68	1	0	0	0	0.32	0.68
EST1	0	0.58	0.42	1	0	0	0	0.43	0.57
EST2	0	0.73	0.27	1	0	0	0	0.51	0.49
EST5	0	0.08	0.92	0.36	0.64	0	0	0.66	0.34
FREQ.S	0	0.08	0.92	1	0	0	0	0.23	0.77

### 3.3. Simulación

La Figura 5 compara las proporciones de cada estado observado y simulado. La divergencia de *Kullback-Leibler* (KL) entre las distribuciones de probabilidades simuladas y empíricas arrojó:  $KL_{SONIA} = 0.0357$ ;  $KL_{NATI} = 0.0387$ ;  $KL_{JAKO} = 0.0886$ ;  $KL_{MIRKA} = 0.0430$ .

Si bien las divergencias son bajas y cercanas a cero, el modelo simula peor en el caso de JAKO y en particular para el estado observado «000», el cual se subestima en todos los sujetos.

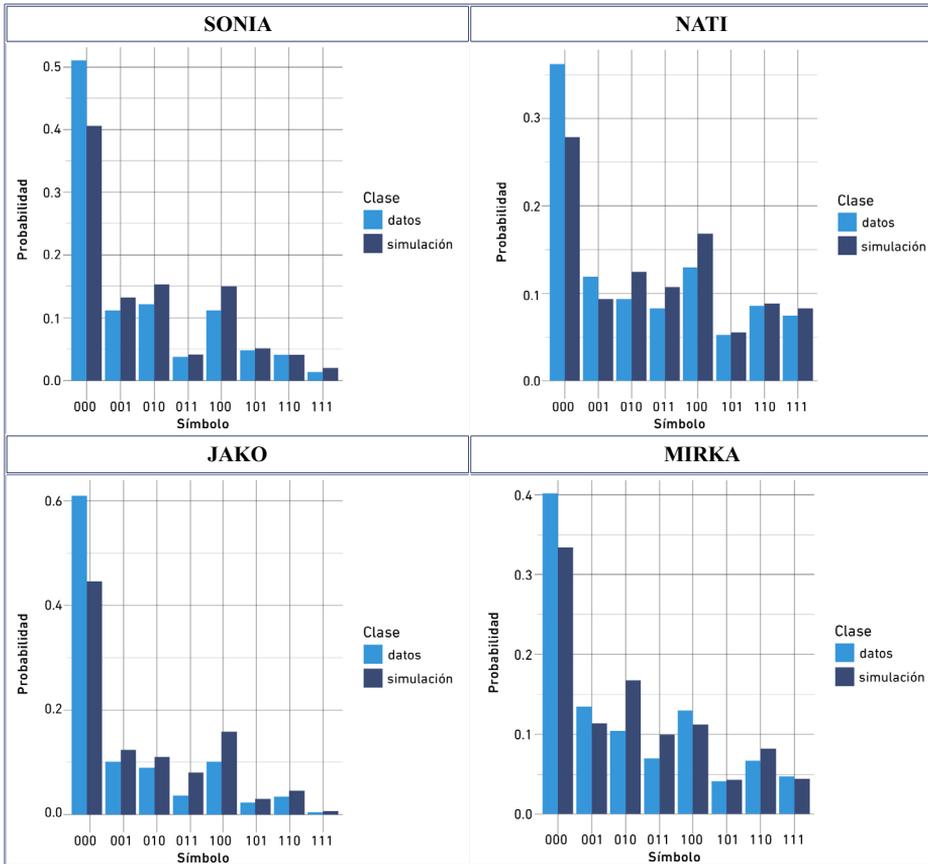


Figura 5. Distribución de estados observados simulada y empírica

## 4. Conclusiones

Este es el primer estudio en ELE en considerar la dinámica en estados de aprendizaje ocultos que gobiernan secuencias observadas («motivos») de patrones de error mediante cadenas de Markov. Usando LDA se identificaron tres tópicos que lograban captar tres tipos de dinámica asociadas a franjas de probabilidad de error (de concordancia) en las sesiones. El primer tópico, relacionado con baja cantidad de error, se asociaba a motivos cercanos al atractor «a» (dominado por el atractor fuerte «aaa»). En el extremo

opuesto, el tercer tópico (relacionado a alta cantidad de errores) se hallaba dominado por motivos cercanos al atractor de error «b». El segundo tópico exhibía una dinámica intermedia asociada a estados débiles del atractor «a». SONIA y JAKO se encuentran cerca del atractor «correcto». NATI y MIRKA, cerca entre el atractor «error» y los estados débiles del atractor «a». La no linealidad del aprendizaje supuesta por la teoría de sistemas dinámicos no predice etapas fijas de adquisición. Notar que en NATI y MIRKA se encuentra evidencia de un régimen cercano al atractor «a» tanto hacia las primeras sesiones como hacia las últimas, es decir que no hay un cambio lineal «de menor a mayor».

Las probabilidades de emisión del HMM permitieron analizar los tres regímenes de error para cada sujeto. Todos los sujetos mostraban un estado oculto más cercano al atractor «correcto»: el segundo para SONIA y JAKO, el primero para NATI y el tercero para SONIA. Los otros dos estados ocultos revelaron una dinámica intermedia con un gradiente de estados observados cada vez más cercanos al atractor de «error». Cuando todas las predictoras valían cero, la dinámica de las matrices de transición entre estados ocultos manifestaban mayor (SONIA, JAKO) o menor (NATI, MIRKA) probabilidad de permanecer en el estado oculto más cercano al atractor «correcto». Pero desde este nunca se pasaba al estado oculto cercano al atractor «error». Es más, en dicho evento, se permanecía poco en el régimen de «error» para pasar al régimen oculto de dinámica intermedia.

La Tabla 8 compara las predicciones del análisis con los resultados hallados para cada variable. Se contrasta la influencia predicha negativa (N) o positiva (P) con aquella observada en la dinámica de los estados ocultos para cada aprendiente. Un hallazgo bastante establecido en la literatura sobre adquisición de la concordancia en L2 es que la concordancia de género y número del artículo es más fácil de adquirir que la del adjetivo. En NATI se observó que los artículos indefinidos promueven que la dinámica vaya hacia un régimen cercano al atractor «error» (igualmente, en JAKO, pero con alta probabilidad de permanecer cerca del atractor «correcto»). En MIRKA la influencia de dicho factor fue positiva. Para SONIA y NATI (niveles de competencia más bajos) los adjetivos indefinidos y los calificativos favorecían que se dejara el régimen más cercano al atractor «correcto» y que se asentara una de error intermedia. En JAKO hubo efecto facilitador. En cuanto a MIRKA, si bien ambos factores inducían a la permanencia en el régimen cercano al atractor «correcto», si se salía de este la dinámica no volvía a él. En suma, la influencia negativa predicha se verificó solamente en los aprendientes de menos nivel.

Tabla 8. Comparación, para cada variable, entre la influencia predicha y los resultados

Variable	Influencia	SONIA (A1/A2)	NATI (B1)	JAKO (B2)	MIRKA (C1)
LDA	N	N	P/N	P	P
EP = «1»	N	P	P/N	P	P
EP = «2»	N	N	P	P	P
GRAM	N	P	P	P	P
MOD = «1»	N	-	N	P/N	P
MOD = «2»	N	N	N	P	P/N
MOD = «3»	N	N	N	P	P/N
IMA.CONC	P	P/N	P/N	P	P
FAM.LEX	P	P/N	P/N	P	P
AN	N	P	P/N	P/N	P
EST1, EST2, EST5	P	P	P	P	P/N
FREQ.S	P	P	P	P	P

En la literatura también se reporta para la concordancia de género un efecto facilitador de los controladores inanimados respecto de los animados. Para NATI y JAKO se observó que la animicidad provoca que, una vez dejado el atractor «correcto» (con más probabilidad de permanencia), la dinámica vaya hacia alguno de los otros dos estados ocultos. En SONIA es muy improbable que se abandone dicho atractor y en MIRKA la dinámica regresa a este. Por ende, contrariamente a lo predicho, la animicidad favoreció una dinámica con influencia positiva. Otro factor crucial para la concordancia parece ser la distancia (lineal o estructural) entre controlador y objetivo. En SONIA las concordancias a larga distancia llevan a que la dinámica se acerque al atractor «error». En los demás sujetos la variable LDA conlleva probabilidad alta de permanecer en el régimen cercano al atractor «correcto»; pero, si escapase de este, la dinámica no regresa en el caso de NATI pero sí lo hace para los otros dos aprendientes. Notar que la influencia negativa de la larga distancia en la dinámica se ejerce solamente en los aprendientes de menor competencia (A, B1). Se hipotetizó que las otras características del controlador (IMA.CONC y FAM.LEX) influyeran positivamente en mantener la dinámica cerca del atractor «correcto». En general, las probabilidades de permanecer en el régimen asociado al atractor «correcto» es del orden medio/alto. En el evento de abandonarlo, la dinámica se aleja de este en el caso de SONIA/NATI pero retorna en el caso de JAKO/MIRKA. La inserción de una «-e-» epentética se consideró como factor de influencia negativa. No obstante ello, excepto para el caso de ambos miembros (EP

= «2») en SONIA, las probabilidades de permanecer cerca del atractor «correcto» resultaron ser altas. Tampoco influyó negativamente el hecho de que una concordancia tuviera más de dos términos. Se crearon tres estrategias que se supuso que podrían estar poniendo en acto los alumnos cuando creaban las concordancias. A excepción de MIRKA, dichas estrategias lograron un efecto facilitador, logrando mantener a la dinámica cerca del atractor «correcto». La interpretación de este resultado es que los hablantes sacan provecho del parecido de las palabras singulares en español e italiano y forman el plural agregando una «-s» a la palabra singular italiana (EST5: «elezione» > «elecciones»). Como en español estas coinciden con palabras terminadas en consonante que requieren plural en «-es», sobrepasan de esta forma la dificultad de insertar plural con «-e-» epentética. Este factor quizás debe haber influido en el hecho de no haber hallado efecto negativo para la dinámica con la variable EP. La alta frecuencia de los TYPES de concordancia acusó, como se predijo, un efecto facilitador. La dinámica tendía a permanecer en el régimen cercano al atractor correcto, excepto en NATI; aunque, aun en este caso la dinámica regresaba a dicho régimen.

La simulación aproximó bien en general a los datos, aunque con resultados menos favorables para JAKO y, en particular, para el estado observado «000». Es decir que el modelo se aparta más del atractor «correcto» que los datos empíricos.

En el enfoque de los sistemas dinámicos propuesto el nivel de análisis es el individuo, por ello se usan series temporales. Se trata de identificar regiones de mayor variabilidad porque se hipotetiza que es precursora de cambios cualitativos de la dinámica de aprendizaje (bifurcaciones). Sacar un promedio de precisión de errores oculta la variabilidad en cada individuo: podría dar un promedio alto, pero no se sabría cuáles son los periodos de intermitencia de error. Agregar (promediar) datos puede ocultar diferencias importantes en la dinámica de cada aprendiente. La teoría de sistemas dinámicos complejos trata sobre transiciones. Lo importante es identificar los factores que influyen en las trayectorias individuales y usarlos para guiar el proceso de aprendizaje; el acento no está puesto en realizar predicciones para la adquisición (imposibles dada la naturaleza no lineal del fenómeno). La influencia de una determinada variable depende de la dinámica de cada individuo. En pos de incrementar la escalabilidad, se podrían tomar más informantes itálofonos para verificar comportamientos de dinámica similar a las halladas o bien influencias novedosas. Por ejemplo, si LDA ejerce influencia negativa en niveles iniciales para hablantes de italiano LI; allí deberán dirigirse las estrategias didácticas. Además, se podría ampliar la LI de los informantes a lenguas tipológicamente diversas para corroborar si algunas dinámicas se mantienen.

En este trabajo de naturaleza observacional se optó por un enfoque dinámico para el error de concordancia. En lugar de tomar una variable respuesta estática de orden binario «error/no error», se analizaron secuencias de errores, las cuales conformaban patrones generados por la dinámica de aprendizaje en tres regímenes ocultos. El proceso no lineal que se aduce responsable del desarrollo lingüístico causa alta variabilidad. Si se considera al 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, estudios longitudinales y simulaciones. Se espera que con esta contribución se logre un paso más en pos de dicho cambio de perspectiva.

## 5. Referencias

- Alarcón, I. (2009). The processing of gender agreement in L1 and L2 Spanish: Evidence from Reaction Time Data. *Hispania*, 92(4), 814-828. <https://www.jstor.org/stable/40648463>
- Alarcón, I. (2011). Spanish gender agreement under complete and incomplete acquisition: Early and late bilinguals' linguistic behavior within the noun phrase. *Bilingualism: Language and Cognition*, 14(3), 332-350. <https://doi.org/10.1017/S1366728910000222>
- Alemán Bañón, J., Fiorentino, R., & Gabriele, A. (2014). Morphosyntactic processing in advanced second language (L2) learners: An event-related potential investigation of the effects of L1-L2 similarity and structural distance. *Second Language Research*, 30(3), 275-306. <https://doi.org/10.1177/0267658313515671>
- Blei, D. M., Ng, A. Y., & Jordan, M. I. (2003). Latent Dirichlet Allocation. *Journal of Machine Learning Research*, 3(5), 993-1022.
- Chan, H., Veerspoor, M., & Vahtrick, L. (2015). Dynamic Development in Speaking Versus Writing in Identical Twins. *Language Learning*, 65(2), 298-325. <https://doi.org/10.1111/lang.12107>
- Corbett, G. (2006). *Agreement*. Cambridge: Cambridge University Press.
- 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. <https://doi.org/10.3758/BF03192738>
- Foote, R. (2011). Integrated knowledge of agreement in early and late English-Spanish bilinguals. *Applied Psycholinguistics*, 32(1), 187-220. <https://doi.org/10.1017/S0142716410000342>
- Foote, R. (2015). The production of gender agreement in native and L2 Spanish: The role of morphophonological form. *Second Language Research*, 31(3), 343-373. <https://doi.org/10.1177/0267658314565691>

- Franceschina, F. (2001). Morphological or syntactic deficit in near-native speakers? An assessment of some current proposals. *Second Language Research*, 17(3), 213-247. <https://doi.org/10.1177/026765830101700301>
- Gabriele, A., Alemán Bañón, J., Hoffmann, L., Coveyd, L., Rossomondoe, A., & Fiorentino, R. (2021). Examining variability in the processing of agreement in novice learners: Evidence from event-related potentials. *Journal of Experimental Psychology Learning Memory and Cognition*, 47(7), 1106-1140. <https://doi.org/10.1037/xlm0000983>
- Gillon Dowens, M., Guo, J., Barber, H., & Carreiras, M. (2011). Gender and Number Processing in Chinese Learners of Spanish: Evidence from Event Related Potentials. *Neuropsychologia* 49(7), 1651-1659. <https://doi.org/10.1016/j.neuropsychologia.2011.02.034>
- Gillon Dowens, M., Vergara, M., Barber, H., & Carreiras, M. (2010). Morphosyntactic processing in late second-language learners. *Journal of Cognitive Neuroscience*, 22(8), 1870-1887. <https://doi.org/10.1162/jocn.2009.21304>
- González, P., Mayans, D., & Van der Bergh, H. (2019). Nominal agreement in the interlanguage of Dutch L2 learners of Spanish. *International Review of Applied Linguistics in Language Teaching*. Advance online publication. <https://doi.org/10.1515/iral-2017-0174>
- Keating, G. D. (2009). Sensitivity to violations of gender agreement in native and nonnative Spanish: An eye-movement investigation. *Language Learning*, 59(3), 503-535. <https://doi.org/10.1111/j.1467-9922.2009.00516.x>
- Keating, G. D. (2010). The effects of linear distance and working memory on the processing of gender agreement in Spanish. En VanPatten, B., & Jegerski, J. (eds.), *Research in Second Language Processing and Parsing* (pp. 113-134). John Benjamins. <https://doi.org/10.1075/lald.53.05kea>
- 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>
- Larsen Freeman, D., & Cameron, L. (2008). *Complex Systems and Applied linguistics*. Oxford University Press.
- Lichtman, K. (2009). Acquisition of Attributive and Predicative Adjective Agreement in L2 Spanish. En Bowles, M., Ionin, T., Montrul, S., & Tremblay, A. (Eds.), *Proceedings of the 10th Generative Approaches to Second Language Acquisition Conference* (pp. 231-247). <http://www.lingref.com/cpp/gasla/10/abstract2273.html>
- Mac Whinney, B. (2021). *The Childes Project: Tools for Analyzing Talk. Part 1 y 2* [en línea]. <https://talkbank.org/manuals/CHAT.pdf>

- Marafioti, P. E. (2020). Dinámica del error de concordancias plurales en un corpus de cuatro aprendientes italianos de español L2. *Revista Nebrija de Lingüística Aplicada a la Enseñanza de Lenguas (RNAEL)*, 14(29). <https://doi.org/29.10.26378/rnlael1429409>
- 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), 116-128. <https://revistas.unc.edu.ar/index.php/racc/article/view/32001>
- McCarthy, C. (2008). Morphological variability in the comprehension of agreement: An argument for representation over computation. *Second Language Research*, 24(4), 459-486. <https://doi.org/10.1177/0267658308095737>
- 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. <https://doi.org/10.1111/j.1467-9922.2008.00449.x>
- Muñoz, J., Díaz, L., & Mongeon, C. (2000). N-drop and determiners in native and non-native Spanish: More on the role of morphology in the acquisition of syntactic knowledge. En Leow, R. P., & Sanz, C. (eds.), *Current research on the acquisition of Spanish* (pp. 67-96). Cascadilla Press.
- Peña, D. (2002). *Análisis de Datos Multivariantes*. Mc Graw Hill.
- Sagarra, N. (2007). Online processing of gender agreement in low proficient english-spanish late bilinguals. En Camacho, J., Flores-Ferrán, N., Sánchez, L., Déprez, V., & Cabrera, M. J. (eds.), *Current Issues in Linguistic Theory Series* (pp. 240-253). <https://doi.org/10.1075/cilt.287.18sag>
- Sagarra, N., & Herschensohn, J. (2013). Processing of gender and number agreement in late spanish bilinguals. *International Journal of Bilingualism*, 17(5), 607-627. <https://doi.org/10.1177/1367006912453810>
- Scrucca, L., Fop, M., Murphy, T. B., & Raftery, A. E. (2016). Mclust 5: Clustering, classification and density estimation using Gaussian finite mixture models. *The R Journal*, 8(1), 289-317. <https://doi.org/10.32614/RJ-2016-021>
- Visser, I. (2011). Seven things to remember about hidden Markov models: A tutorial on Markovian models for time series. *Journal of Mathematical Psychology*, 55(6), 403-415. <https://doi.org/10.1016/j.jmp.2011.08.002>
- White, L., Valenzuela, E., Kozłowska-Macgregor, M., & Leung, Y. (2004). Gender and number agreement in nonnative Spanish. *Applied Psycholinguistics*, 25(1), 105-133. <https://doi.org/10.1017/S0142716404001067>

## NOTAS

---

- 1 Corbett (2006) denomina controlador al ítem léxico que determina la concordancia (aquí será nominal) y objetivo al elemento cuya forma es determinada por aquel (aquí serán artículos y adjetivos).
- 2 Las otras estrategias creadas fueron las siguientes: (1) Estrategia 3 (EST3): si la palabra plural del italiano termina en «-o» u en «-a» no acentuada («le foto» [‘las fotos’], «le ossa» [‘los huesos’]), poner el plural del italiano; (2) Estrategia 4 (EST4): si la palabra plural del italiano termina en «-e», poner en español el plural en «-es». Por ejemplo: «vacanze» > «vacaciones»; «strade» > «calles»; «volte» > «veces»; (3) Estrategia 6 (EST6): si la palabra singular del italiano termina en «-e», poner en español el plural en «-es». Por ejemplo, la palabra «grande» [‘grande’] o «studente» [‘estudiante’] podrían formar plural (y singular) a partir de una base singular en italiano: grandes, estudiantes. Otros casos: «fonte» > «fuentes»; «abitudine» > «costumbres»; «dolce» > «dulces». Son casos que no coinciden con «-e-» epentética; (4) Estrategia 7 (EST7): si la palabra plural del italiano termina en «-a» acentuada («università» [‘universidades’]) o es invariante terminada en consonante («i film» [‘las películas’]) poner, en general, plural en «-es». Respecto de la noción de «transfer» (la aplicación de rasgos lingüísticos de una lengua a otra), vale la pena notar que las estrategias EST1, EST2, EST4, EST5, EST6 implicarían un «transfer» positivo, ayudando a que el alumno no cometa errores de concordancia. Sin embargo en Marafioti (2021, 2022) las EST4, EST6 y EST7 no obtuvieron resultados significativos, razón por la cual se las excluyó.