

Revista
Nebrija

de
Lingüística
Aplicada a la
Enseñanza
de las
Lenguas



UNIVERSIDAD
NEBRIJA

Revista
Nebrija

de
Lingüística
Aplicada a la
Enseñanza
de las
Lenguas



Número 29

Equipo de redacción

Directora

Dra. Susana Martín Leralta

Editora responsable

Dra. María Cecilia Ainciburu

Editores temáticos invitados

Anita Ferreira y Jéssica Elejalde

Editores adjuntos

Dra. Ocarina Masid Blanco

Dra. Margarita Planelles Almeida

Dra. Eirini Mavrou

Editores internacionales asociados

Dr. Joseba Ezeiza Ramos (Universidad del País Vasco)

Dr. Carlos de Pablos Ortega (Universidad de East Anglia, Inglaterra)

Dr. Javier Pérez Ruiz (Wenzao Ursuline College of Languages, Taiwán);

Dra. Claudia Villar (Universidad de Manheim, Alemania)

Dr. Joseba Ezeiza Ramos (Universidad del País Vasco).

Consejo editorial

Leonor Acuña	Universidad de Buenos Aires
Kris Buyse	Katholieke Universiteit te Leuven
Laura Colantoni	Universidad de Toronto, Canadá
Florencio Del Barrio	Universidad Ca' Foscari, Venecia
Raquel Fernández Fuertes	Universidad de Valladolid
José Gómez Asencio	Universidad de Salamanca
Anita Ferreira Cabrera	Universidad de Concepción, Chile
Lu Jingsheng	Shanghai International Studies University
Manel Lacorte	Universidad de Maryland
Marina Larionova	Universidad MGIMO de Moscú
Michael H. Long	University of Maryland
Silvina Montrul	University of Chicago, USA
Francisco Moreno Fernández	Instituto Cervantes en Harvard University
Juana Muñoz Licerias	University of Ottawa
Florentino Paredes García	Universidad de Alcalá
Susana Pastor Cesteros	Universidad de Alicante
Maria Luisa Regueiro Rodríguez	Universidad Complutense de Madrid
Cristina Sanz	Georgetown University
Graciela Vázquez	Freie Universität Berlín

Universidad Antonio de Nebrija

ISSN 1699-6569

Este número se terminó de editar el 15 de diciembre de 2020

Sumario del Número 29

SECCIÓN TEMÁTICA

- Avances en los Estudios sobre Corpus de Aprendientes en formato Computacional. Presentación de la Editoras
Curricular innovations for the teaching of L2 in the 21st century. Presentation of the editors 8
Anita Ferreira Cabrera y Jessica Elejalde Gómez
- El aprendizaje de la pragmática a través de un corpus multimodal de nativos y aprendientes de español
Learning Pragmatics through a Multimodal Corpus of Spanish Learners and Native Speakers 12
Marta Vacas Matos
- Estudio comparativo sobre el uso de recursos metadiscursivos en corpus de hablantes nativos y no nativos grecófonos de español
A comparative corpus-based study on the use of metadiscourse resources between native and Greek non-native speakers of Spanish 36
Panagiota Salapata y Angélica Alexopoulou
- La expresión de la actitud en un corpus oral de presentaciones académicas de estudiantes ELE y hablantes de herencia de español en contexto universitario en Estados Unidos
The expression of attitude in an oral corpus of academic presentations of L2 Students and Heritage Speakers of Spanish at the university context in the United States 54
Eva Gómez García
- Análisis de Errores en el Corpus CAELE: estudio sobre la concordancia gramatical en el verbo "ser" en aprendientes francófonos y anglosajones
Error Analysis in the CAELE Corpus: study on the grammatical agreement in the verb "ser" in Francophone and Anglo-Saxon learners 78
Anita Ferreira Cabrera, Lorena Blanco San Martín y Jessica Elejalde Gómez
- Dinámica del error de concordancias plurales en un corpus de cuatro aprendientes italianos de español L2
The dynamics of agreement errors in a corpus of four Italian learners of Spanish L2 92
Pablo Ezequiel Marafioti
- Estudio comparativo de métodos de transcripción para corpus orales: el caso del español
Comparative study of transcription methods for spoken corpus: the case of Spanish 113
Marimar Rufino Morales

MISCELÁNEA

Using Dialogic Talks in EFL Primary Teacher Education: An Experience
*Una Experiencia sobre Tertulias Dialógicas en Magisterio en Educación
Primaria con la especialidad de Lengua Extranjera*

Raquel Fernández Fernández

152

Panorama general y sistematización de un conjunto de errores en el
aprendizaje escolar del maya en México
*General outlook and systematization of a set of errors in the school learning
of Mayan language in Mexico*

Hamlet Antonio García Zúñiga y Didier A. Chan Quijano

174

Avances en los Estudios sobre Corpus de Aprendientes en formato Computacional

Presentación de la Editoras

Advances in Computerized Learner Corpus Studies

Presentation of the Editors

Anita Ferreira

Universidad de Concepción

aferreir@udec.cl

Jessica Elejalde

Universidad de Concepción

jelejalde@udec.cl

Esta sección temática está circunscrita en el área de la Lingüística Aplicada denominada Corpus de Aprendientes en formato computacional (del inglés, *Computer Learner Corpora*, CLC). El enfoque de investigación de dicha línea es interdisciplinario como resultado de la interacción de dos disciplinas, hasta hace pocos años desvinculadas: la lingüística de corpus (LC) y la Adquisición de Segundas Lenguas (ASL) (Granger, 2002). Actualmente, la investigación sobre Corpus de Aprendientes está centrada en el análisis de la interlengua y en la producción auténtica de los aprendientes, logrando un creciente beneficio en el ámbito de ASL (Granger, 2017).

El interés por los corpus de aprendientes se desarrolló en los años noventa y primera década del siglo XXI, inspirado por el trabajo de Dagneaux, Denness y Granger (1998) en el *Centre for English Corpus Linguistics*, dando origen al primer corpus de aprendientes reconocido internacionalmente, el ICLE (*International Corpus of Learner English*)¹. Desde entonces, el desarrollo de

This thematic section is circumscribed in the area of the Applied Linguistics called Corpus of Learners in computer format (from English, *Computer Learner Corpora*, CLC). The research focus of this line is interdisciplinary as a result of the interaction of two disciplines, until a few years ago unrelated: corpus linguistics (LC) and Second Language Acquisition (ASL) (Granger, 2002). Currently, research on the Corpus of Learners is focused on the analysis of the interlanguage and the authentic production of learners, achieving increasing benefit in the field of ASL (Granger, 2017).

The interest in the corpus of learners developed in the nineties and first decade of the 21st century, inspired by the work of Dagneaux, Denness and Granger (1998) at the *Centre for English Corpus Linguistics*, giving rise to the first internationally recognized corpus of learners, the ICLE (*International Corpus of Learner English*). Since then, the

los corpus de aprendientes ha permitido un avance significativo en los estudios de Adquisición de Segundas Lenguas (ASL) (Granger, 2012; 2015). Mientras que los trabajos anteriores -basados en procedimientos tradicionales de obtención de datos- eran limitados en cuanto al número de sujetos estudiados para poder controlar las variables que afectan a la producción del aprendiente, los nuevos corpus computarizados permiten trabajar con más y mejor calidad de datos de lengua natural. De este modo, el análisis de corpus de aprendientes proporciona una base empírica sólida para la descripción de la interlengua, esto es, del sistema lingüístico cognitivo específico de una lengua no materna, que posee un hablante no nativo.

Según la definición aportada por Granger (2017), los corpus de aprendientes son colecciones de textos orales o escritos digitales de muestras auténticas de interlengua de aprendientes de lenguas extranjeras o segundas, los cuales han sido recolectados acorde con un diseño de criterios y principios específicos para garantizar de cierta forma la calidad metodológica en su creación (Sinclair, 2005). Los corpus de aprendientes son implementados y procesados computacionalmente por medio de programas especializados para la recuperación de información acorde con determinados criterios de búsqueda. En este contexto, si bien uno de los objetivos principales en la ASL es construir modelos de las representaciones subyacentes al proceso de aprendizaje de un estadio en particular durante la adquisición de una L2, los métodos de recolección se han realizado a través de estudios experimentales, observacionales o métodos introspectivos. Así, como lo señala (Granger, 2004), la investigación en ASL ha favorecido los métodos con variables controladas y experimentales, relegando los datos lingüísticos del uso natural de la lengua. La principal motivación para el uso de este tipo

development of the corpus of learners has allowed significant progress in Second Language Acquisition (ASL) studies (Granger, 2012; 2015). While previous work -based on traditional data collection procedures- was limited in the number of subjects studied in order to control for variables affecting learner production, the new computerized corpus allows for more and better-quality natural language data. In this way, the analysis of the corpus of learners provides a solid empirical basis for the description of the interlanguage, that is, the specific cognitive linguistic system of a non-native language, which a non-native speaker possesses.

According to the definition provided by Granger (2017), corpora of learners are collections of oral or written digital texts of authentic interlanguage samples of foreign or second language learners, which have been collected according to a design of specific criteria and principles to guarantee in a certain way the methodological quality in their creation (Sinclair, 2005). The corpus of learners is implemented and processed computationally by means of specialized programs for the retrieval of information according to certain search criteria. In this context, although one of the main objectives in the ASL is to build models of the representations underlying the learning process of a particular stage during the acquisition of an L2, the collection methods have been carried out through experimental, observational or introspective methods. Thus, as noted (Granger, 2004), research in ASL has favored methods with controlled and experimental variables, relegating linguistic data to natural language use. The main motivation for the use of this type of procedure has to do with the fact that many times it is not possible to elicit the structures investigated in a natural

de procedimientos tiene que ver con el hecho de que muchas veces no es posible elicitar las estructuras investigadas de forma natural. Por esta razón, la producción debe prepararse acorde con el interés del estudio y en muchos casos forzarse para conseguir los datos necesarios. También existen dificultades para controlar las variables que afectan la producción oral o escrita en un contexto no experimental (Mackey y Gass 2015). No obstante, en el ámbito de la enseñanza de la lengua, la necesidad de observar cómo usan de manera real la lengua meta los sujetos, por ejemplo, en una situación comunicativa en que utilicen lo aprendido en clase, puede arrojar resultados que evidencien qué áreas presentan mayor dificultad, cuáles son persistentes o cuáles responden a diferentes fenómenos dentro del desarrollo de la interlengua. Por esta razón, la investigación en corpus de aprendientes contribuye a que la información obtenida de muestras a gran escala refleje el estado real y las tendencias de uso de la lengua objeto de estudio.

Las características de los corpus de aprendientes de lenguas son diversas según el tipo de investigación que se lleva a cabo. Tempranamente, Tono (2003) especifica como variables importantes los rasgos del texto, la recolección de los datos y las variables del aprendiente (ver Tabla 1). Los rasgos del texto se refieren al modo (escrito u oral), al género discursivo (carta, presentación oral, etc), el estilo o tipo de modalidad discursiva que presenta el texto (narrativo, argumentativo o expositivo) y finalmente, el tópico del cual trata el texto. Por otra parte, en la recolección de los datos, es importante señalar el modo de recolección de estos. Algunos métodos de recolección utilizados corresponden a tareas de escritura o grabación de conversaciones espontáneas. Otro modo de recolección es por medio del diseño de actividades realizadas a través del computador, ya sea de forma oral o escrita. Así también, las variables del aprendiente son muy

way. For this reason, the production must be prepared according to the interest of the study and in many cases forced to obtain the necessary data. There are also difficulties in controlling the variables that affect oral or written production in a non-experimental context (Mackey and Gass 2015). However, in the field of language teaching, the need to observe how subjects actually use the target language, for example, in a communicative situation where they use what they have learned in class, may yield results that show which areas present greater difficulty, which are persistent, or which respond to different phenomena within the development of the interlanguage. For this reason, research in the corpus of learners contributes to the fact that information obtained from large-scale samples reflects the actual state and trends of use of the language under study.

The characteristics of the corpus of language learners are diverse according to the type of research being carried out. Early on, Tono (2003) specifies text traits, data collection, and learner variables as important variables (see Table 1). The features of the text refer to the mode (written or oral), the discursive genre (letter, oral presentation, etc), the style or type of discursive modality that the text presents (narrative, argumentative or expository) and finally, the topic of which the text deals with. On the other hand, when collecting the data, it is important to point out the way in which they are collected. Some of the collection methods used correspond to writing tasks or recording spontaneous conversations. Another mode of collection is through the design of computer-based activities, either oral or written. Also, the variables of the learner are very relevant, since they must correspond to the research purposes, for example, if a description or contrast is to be made in

relevantes, ya que deben corresponderse con los propósitos investigativos, por ejemplo, si se pretende realizar una descripción o un contraste en relación con los niveles de competencia y/o la lengua materna, la selección de los sujetos debe realizarse acorde con dichos criterios.

Características para el diseño de un corpus de aprendientes					
Rasgos del texto		Recolección de los datos		Variables del aprendiente	
Modo	Escrito/oral	Tipo y modo de recolección	Transversal/longitudinal Manual/digital	Internas: Cognitiva afectiva	Edad/estilo de aprendizaje Motivación/actitud
Género	Diario/ensayos personales	Elicitación	Espontánea/ Por instrucción	L1	Lengua materna del aprendiente
Tipo	Narrativo/Descriptivo/Expositivo Argumentativo	Uso de referencias	Diccionario/texto fuente	L2 segunda extranjera	L2/ LE
Tópico	General/específico	Limitación de tiempo	Libre/controlado	Nivel de Competencia	Prueba para evaluar Competencia

Tabla 1. Adaptada y traducida de Tono (2003): Diseño y consideraciones para la construcción de un corpus de aprendientes. Fuente: traducción y adaptación propia

Además de dichas características, existen otras referidas al "tamaño", "representatividad" o "extensión del corpus", todas ellas dependen del propósito de la investigación y de la disponibilidad de la muestra (Granger, 2003, 2009; Parodi, 2008, 2010; Reppen, 2010; Sánchez, 2015; Graeser, 2017). Agudamente, Parodi (2007: 105) explica que:

El desafío de contar con un corpus representativo de una variedad determinada de lengua –incluso de un único registro específico de tal o cual lengua– es una cuestión compleja debido a la enorme diversidad y variedad inherente a cada lengua particular.

Recientemente, los estudios de corpus de aprendientes ponen de relieve una serie de aspectos metodológicos que deberían ser considerados en futuras investigaciones basadas en corpus, entre ellas, se recomienda que los estudios deben (1) estar teóricamente fundamentados y ser también explicativos (no sólo descriptivos), (2) usar sistemas de anotación más sofisticados (en

relation to the levels of competence and/or the mother tongue, the selection of the subjects must be made according to these criteria.

Características para el diseño de un corpus de aprendientes					
Rasgos del texto		Recolección de los datos		Variables del aprendiente	
Modo	Escrito/oral	Tipo y modo de recolección	Transversal/longitudinal Manual/digital	Internas: Cognitiva afectiva	Edad/estilo de aprendizaje Motivación/actitud
Género	Diario/ensayos personales	Elicitación	Espontánea/ Por instrucción	L1	Lengua materna del aprendiente
Tipo	Narrativo/Descriptivo/Expositivo Argumentativo	Uso de referencias	Diccionario/texto fuente	L2 segunda extranjera	L2/ LE
Tópico	General/específico	Limitación de tiempo	Libre/controlado	Nivel de Competencia	Prueba para evaluar Competencia

Tabla 1. Adaptada y traducida de Tono (2003): Diseño y consideraciones para la construcción de un corpus de aprendientes. Fuente: traducción y adaptación propia

In addition to these characteristics, there are others that refer to the "size", "representativeness" or "extension of the corpus", all of which depend on the purpose of the research and the availability of the sample (Granger, 2003, 2009; Parodi, 2008, 2010; Reppen, 2010; Sánchez, 2015; Graeser, 2017). Sharply, Parodi (2007: 105) explains that

The challenge of having a representative corpus of a given variety of language—even a single specific record of such or such language—is a complex issue due to the enormous diversity and variety inherent to each particular language.

Recently, studies of corpora of learners highlight a series of methodological aspects that should be considered in future corpus-based research. Among them, it is recommended that studies should (1) be theoretically based and also explanatory (not only descriptive), (2) use more sophisticated notation systems (instead of generic systems), and (3) build corpora of learners that are well

vez de sistemas genéricos) y (3) construir corpus de aprendientes que estén bien diseñados y que controlen las variables del aprendiente. El cambio en el procedimiento es importante desde el punto de vista cuantitativo y cualitativo, pues los análisis deben apoyarse en volúmenes de datos más elevados y con una mayor rigurosidad en la homogeneización y categorización de los errores.

Los artículos que integran esta sección temática están enfocados en los avances y resultados de proyectos recientes o en progreso de corpus de aprendientes de lenguas. Los estudios provienen de diversos contextos académicos, institucionales y geográficos y versan sobre diversas lenguas. Los corpus de aprendientes en que se basan los estudios corresponden a corpus de aprendientes en modalidad escrita, oral y multimodal y cubren diversos aspectos de las lenguas objeto de estudio desde aspectos gramaticales, discursivos hasta sociopragmáticos.

En el primer artículo "*El aprendizaje de la pragmática a través de un corpus multimodal de nativos y aprendientes de español*" la autora Marta Vacas del ACCENT / IES Abroad de Madrid presenta una investigación basada en un corpus multimodal de aprendientes nativos de español peninsular, y no nativos de niveles B1 y C1 de Español como Lengua Extranjera, ELE. El Corpus se denomina COREMAH (Corpus Español Multimodal de Actos de Habla). Se trata en este caso de explorar la evolución de los aspectos pragmáticos en estudiantes de una lengua extranjera. Los resultados muestran que a través de los vídeos y transcripciones anotadas y etiquetadas que integran el corpus se puede delimitar las tendencias de los hablantes nativos y no nativos.

En el segundo artículo "*Estudio comparativo sobre el uso de recursos metadiscursivos en corpus de hablantes nativos y no nativos grecófonos de español*" de Panagiota Salapata de la Universidad

designed and that control for learner variables. The change in procedure is important from the quantitative and qualitative point of view, since the analyses must be supported by higher volumes of data and with greater rigor in the homogenization and categorization of errors.

The articles that make up this thematic section focus on the progress and results of recent or ongoing language learning projects. The studies come from various academic, institutional and geographical contexts and deal with different languages. The corpus of learners on which the studies are based are written, oral and multimodal and cover various aspects of the languages under study, from grammatical, discursive to socio-pragmatic aspects.

In the first article "El aprendizaje de la pragmática a través de un corpus multimodal de nativos y aprendientes de español" the author Marta Vacas from ACCENT / IES Abroad in Madrid presents a research based on a multimodal corpus of native speakers of peninsular Spanish, and non-native speakers of levels B1 and C1 of Spanish as a Foreign Language, ELE. The corpus is called COREMAH (Multimodal Spanish Corpus of Acts of Speech). It is about exploring the evolution of pragmatic aspects in students of a foreign language. The results show that through the videos and annotated and labeled transcriptions that make up the corpus, it is possible to delimit the tendencies of native and non-native speakers.

In the second article "Comparative study on the use of metadiscursive resources in corpus of native and non-native Greek-speaking speakers of Spanish" by Panagiota Salapata from Antonio de Nebrija University and Angelica

Antonio de Nebrija y Angélica Alexopoulou de la Universidad Nacional y Kapodistriaca de Atenas se aborda un estudio sobre las estrategias metadiscursivas en el género carta al director en contextos formales. Para ello, las autoras se basan en un corpus de aprendientes constituido por 125 cartas (27.969 palabras) procedentes de hablantes nativos españoles y hablantes no nativos griegos. El estudio se sustenta en las aportaciones de la retórica intercultural, área de investigación en la adquisición de segundas lenguas, cuyo propósito consiste en el estudio del discurso entre individuos provenientes de distintas culturas. Los resultados generales evidencian que en las cartas al director las estrategias que apelan a las emociones del lector y las que proyectan la integridad del autor superan a las que apelan a la razón.

En la tercera contribución, "*La expresión de la actitud en un corpus oral de presentaciones académicas de estudiantes ELE y hablantes de herencia de español en contexto universitario en Estados Unidos*" de Eva Gómez García de la *Brown University*, se presenta un estudio que se enmarca dentro del área de Corpus de Aprendientes de Lenguas en formato computacional y del Análisis del Discurso. A partir de un corpus de 15 presentaciones académicas orales en español, producidas por 6 hablantes de herencia y 9 estudiantes de ELE, se analiza la expresión de la actitud a través de estrategias metadiscursivas para conectar con la audiencia. El corpus empleado en esta investigación recoge muestras orales de aprendientes ELE y hablantes de herencia con el objetivo de describir cómo los estudiantes se proyectan en sus presentaciones para conectar con su audiencia, por cuanto esto puede ayudar a entender las necesidades pedagógicas específicas de estos dos grupos, dado su diferente perfil lingüístico y sociocultural en el contexto de Estados Unidos.

En el cuarto artículo, "*Análisis de Errores en el Corpus CAELE: estudio sobre la*

Alexopoulou from the National and Kapodistrian University of Athens, a study on metadiscursive strategies in the letter to the editor genre in formal contexts is addressed. To this end, the authors rely on a corpus of learners consisting of 125 letters (27,969 words) from native Spanish speakers and non-native Greek speakers. The study is based on the contributions of intercultural rhetoric, an area of research in the acquisition of second languages, whose purpose is the study of discourse between individuals from different cultures. The general results show that in the letters to the editor the strategies that appeal to the reader's emotions and those that project the integrity of the author outweigh those that appeal to reason.

In the third contribution, "The expression of attitude in an oral corpus of academic presentations by ELE students and speakers of Spanish heritage in a university context in the United States" by Eva Gómez García from Brown University, a study is presented that is framed within the area of the Corpus of Language Learners in computer format and of the Analysis of the Discourse. Based on a corpus of 15 oral academic presentations in Spanish, produced by 6 heritage speakers and 9 ELE students, the expression of attitude is analyzed through metadiscursive strategies to connect with the audience. The corpus used in this research collects oral samples of ELE learners and heritage speakers in order to describe how students project in their presentations to connect with their audience, as this can help understand the specific pedagogical needs of these two groups, given their different linguistic and socio-cultural profile in the context of the United States.

In the fourth article, "Error Analysis in the Corpus CAELE: study on the grammatical

concordancia gramatical en el verbo SER en aprendientes francófonos y anglosajones" presentamos un estudio sobre Análisis de Errores Asistido por el Computador sustentado en el Corpus de Aprendientes de Español como Lengua Extranjera, CAELE. Dicho corpus se constituye en una colección de 1217 textos producidos por 210 aprendientes en un ambiente presencial a través de un aula virtual. El corpus ha sido procesado a través del software *Sketch Engine* y está conformado por 265.362 palabras, por 296.136 tokens y por 14.902 oraciones. Las temáticas de escritura de los textos se determinaron mediante un análisis de necesidades y consideraron temas del ámbito de español para fines académicos científicos y humanísticos. Los objetivos de la investigación que se presenta son determinar las frecuencias de errores de concordancia gramatical en el verbo *ser* y comparar las semejanzas y diferencias observadas entre aprendientes francófonos y anglosajones. Los resultados han arrojado que para los aprendientes anglosajones hay mayores dificultades en la selección del género gramatical y número gramatical en los adjetivos, especialmente en la concordancia de los atributos. En cambio, para los aprendientes del francés, si bien presentan un número de errores en dichos aspectos gramaticales, su frecuencia es notoriamente menor.

En el quinto artículo "*Dinámica del error de concordancia plural en un corpus de cuatro aprendientes italianos de español L2*" de Pablo Ezequiel Marafioti de la Universidad Nacional de Córdoba, Argentina se presenta un estudio basado en un análisis cuantitativo de patrones recurrentes para los tipos de errores de concordancia gramatical definidos, etiquetados y acotados en un corpus oral longitudinal producido por cuatro aprendientes italianos de español L2. La investigación hace un aporte interesante al ámbito en la medida que incorpora técnicas cuantitativas de análisis del área de los sistemas complejos a un corpus oral

agreement in the verb SER in French and English speaking learners" we present a study on Computer Aided Error Analysis based on the Corpus of Spanish as a Foreign Language Learners, CAELE. This corpus is a collection of 1217 texts produced by 210 learners in a face-to-face environment through a virtual classroom. The corpus has been processed through the Sketch Engine software and is made up of 265,362 words, 296,136 tokens and 14,902 sentences. The writing themes of the texts were determined through a needs analysis and considered topics from the field of Spanish for scientific and humanistic academic purposes. The objectives of the research presented are to determine the frequencies of errors of grammatical agreement in the verb *ser* and to compare the similarities and differences observed between French and English-speaking learners. The results have shown that for Anglo-Saxon learners there are greater difficulties in the selection of grammatical gender and grammatical number in adjectives, especially in the agreement of attributes. On the other hand, for French learners, although they present a number of errors in these grammatical aspects, their frequency is notoriously lower.

In the fifth article "Dynamics of the plural concordance error in a corpus of four Italian learners of Spanish L2" by Pablo Ezequiel Marafioti from the National University of Córdoba, Argentina, a study is presented based on a quantitative analysis of recurrent patterns for the types of grammatical concordance errors defined, labeled and delimited in a longitudinal oral corpus produced by four Italian learners of Spanish L2. The research makes an interesting contribution to the field as it incorporates quantitative techniques of analysis of the area of complex systems to an oral

longitudinal de errores. Se analiza la evolución de la producción de concordancia plural en cuatro aprendientes italianos de español L2. El trabajo resulta novedoso porque aporta evidencia empírica de producción oral longitudinal, obtenida en entrevistas guiadas de 4 sujetos italianófonos.

Cierra esta sección temática el artículo, "Estudio comparativo de métodos de transcripción para corpus orales: el caso del español" de Marimar Rufino Morales de la Universidad de Montreal. El estudio forma parte de una investigación mayor cuyo objetivo es describir el perfil *ad hoc* del recurso denominado rehablador *off-line* para optimizar la transcripción de grabaciones de corpus orales. La técnica de rehablado *off-line* ha permitido el uso de un programa automático de dictado en su estado actual como herramienta para potenciar la transcripción de entrevistas en menos tiempo y con menos errores. El estudio se basa en el Corpus oral de la lengua española en Montreal, COLEM, conformado por una serie de entrevistas semidirigidas grabadas en entorno natural, de una hora de duración aproximadamente y estructuradas en función de un protocolo de encuesta común. El estilo que recoge el COLEM es el conversacional y familiar (habla espontánea).

NOTAS

1 Granger desarrolló corpus escritos anotados con errores tanto para el inglés (ICLE) como para el francés (FRIDA), utilizando para ello un análisis computacional riguroso de errores, y creando programas de edición y análisis efectivos. ICLE es el corpus de aprendientes más extenso de todos los conocidos hasta el momento, Incluye dos millones de palabras procedentes de trabajos escritos de alumnos de nivel avanzado con distintas lenguas maternas.

longitudinal corpus of errors. The evolution of the production of plural concordance in four Italian L2 Spanish learners is analyzed. The work is novel because it contributes with empirical evidence of longitudinal oral production, obtained in guided interviews of 4 Italian-speaking subjects.

This thematic section closes with the article, "Comparative study of transcription methods for oral corpus: the case of Spanish" by Marimar Rufino Morales from the University of Montreal. The study is part of a larger research project whose objective is to describe the *ad hoc* profile of the resource called *off-line* speaker to optimize the transcription of oral corpus recordings. The *off-line* re-speaker technique has allowed the use of an automatic dictation program in its current state as a tool to enhance the transcription of interviews in less time and with fewer errors. The study is based on the oral corpus of the Spanish language in Montreal, COLEM, consisting of a series of semi-directed interviews recorded in a natural environment, lasting approximately one hour and structured according to a common survey protocol. The style used in COLEM is conversational and familiar (spontaneous speech).

NOTES

1 Granger developed annotated error corpus papers for both English (ICLE) and French (FRIDA), using rigorous computer error analysis, and creating effective editing and analysis programs. ICLE is the most extensive corpus of learners known to date. It includes two million words from written work of advanced students with different mother tongues.

REFERENCIAS BIBLIOGRÁFICAS

- Granger, S. (1998). *The computer learner corpus: a versatile new source of data for SLA research* (pp. 3-18). In Granger, S. (ed.) (1998). *Learner English on Computer*. Addison Wesley Longman: London and New York, 3-18.
- Granger, S. (2002). A bird's-eye view of learner corpus research. *Computer learner corpora, second language acquisition and foreign language teaching*, 6, 3-33.
- Granger, S. (2003). The international corpus of learner English: a new resource for foreign language learning and teaching and second language acquisition research. *Tesol Quarterly*, 37(3), 538-546.
- Granger, S. (2004). Computer learner corpus research: Current status and future prospects. *Applied Corpus Linguistics* (pp. 123-145). Brill Rodopi.
- Granger, S. (2009). The contribution of learner corpora to second language acquisition and foreign language teaching. *Corpora and language teaching*, 33, 13-32.
- Granger, S. (2012). Learner corpora. *The encyclopedia of applied linguistics*, 1- 8.
- Granger, S. (2015). Contrastive interlanguage analysis: A reappraisal. *International Journal of Learner Corpus Research*, 1(1), 7-24.
- Granger, S. (2017). Learner corpora in foreign language education. In S. Thorne and S. May (eds.) *Language and Technology. Encyclopedia of Language and Education*. 3rd edition. Springer International Publishing.
- Mackey, A., y Gass, S. (2015). *Second language research: Methodology and design*. Routledge: New York.
- Parodi, G. (2007). El discurso especializado escrito en el ámbito universitario y profesional: Constitución de un corpus de estudio. *Revista signos*, 40(63), 147-178.
- Parodi, G. (2008). Lingüística de corpus: una introducción al ámbito. *RLA. Revista de Lingüística Teórica y Aplicada*, 46(1), 93-119.
- Parodi, G. (2010). *Lingüística de corpus: de la teoría a la empiria*. Iberoamericana Vervuert.
- Reppen, R. (2010). *Using corpora in the language classroom*. Cambridge University Press.
- Sánchez, A. (2015). La Investigación de Corpus de Aprendientes y el desarrollo de los estudios de la interlengua del español. *Language Design* 17, pp., 57-84.
- Sinclair, J. (2005). How to build a corpus. En M. Wynne (ed.), *Developing Linguistic Corpora: A Guide to Good Practice*. Oxford: Oxbow books, 79-83.
- Tono, Y. (2003). Learner corpora: design, development and applications. *Proceedings of the 2003 Corpus Linguistics Conference*, 800-809.

El aprendizaje de la pragmática a través de un corpus multimodal de nativos y aprendientes de español

Learning Pragmatics through a Multimodal Corpus of Spanish Learners and Native Speakers

Marta Vacas Matos

ACCENT / IES Abroad

titavama@gmail.com

RESUMEN

Cuando los aprendientes de una lengua extranjera se enfrentan a ella, además de asimilar una nueva gramática y un vocabulario diferente, con el fin de evitar errores pragmalingüísticos y sociopragmáticos, deben aprender a ver el mundo a través de una nueva pragmática. La web del corpus que se presenta es un recurso avanzado pensado para enseñar pragmática en una clase ELE a diversos niveles. A su vez, analiza y compara en nativos y no nativos la realización de tres actos de habla conflictivos para los aprendientes de una lengua extranjera (los cumplidos, el rechazo y las disculpas). Además, dada su naturaleza multimodal, también se pueden analizar y comparar las diferencias en el lenguaje no verbal, en los gestos, la proxémica y el tono de voz. El formato descargable de las transcripciones permite tres versiones de las mismas (sin anotaciones, con anotaciones y con etiquetaciones) para su explotación.

Palabras clave: corpus, multimodal, pragmática, actos de habla, comunicación intercultural.

ABSTRACT

When foreign language learners face a new language, besides learning a new grammar and vocabulary, they should learn to see the world through new pragmatic rules, in order to avoid pragmalinguistic and sociopragmatic mistakes. The corpus website presented here is an advanced resource built to teach pragmatics in a SL/FL class at different levels. It analyzes and compares three different speech acts that are controversial and difficult for foreign language learners (compliments, refusals, and apologies). Moreover, given its multimodal nature, non-verbal language differences, gestures, as well as voice tone and proxemics are analyzed. The corpus allows to download the transcriptions in three different modes (with or without annotations and tags) for their study.

Keywords: corpus, multimodal, pragmatics, speech acts, intercultural communication.

1. INTRODUCCIÓN

Desde los años 90, varios lingüistas (Bouton, 1988, 1990, 1992, 1994; Kasper y Rose; 1999, Rose y Kasper, 2001; y Kasper 2001) han ido exponiendo en numerosos estudios publicados los beneficios de la instrucción de la pragmática en clase de una lengua extranjera. Sin embargo, no es hasta los años 2000 cuando aparecen los primeros trabajos importantes sobre la enseñanza de la pragmática, abogando por una consciencia pragmática necesaria para la adquisición (Bardovi-Harlig y Mahan-Taylor, 2003). Esta adquisición sería favorecida por tres pilares fundamentales, como son las actividades de consciencia pragmática, las muestras auténticas de lengua, o que el input preceda a estas actividades. Según estas autoras, un modo ideal de integrar la pragmática en la clase desde los primeros niveles es tener siempre presente la L1 de los estudiantes como herramienta indispensable para contrastar los ejemplos de la L2. La obra de Bardovi-Harlig y Mahan-Taylor, escrita por y para profesores de lenguas extranjeras, impulsó no solo la instrucción de la pragmática en el aula, sino también la investigación en este campo, buscando nuevas formas de llevar la pragmática al aula (Kasper, 2001; Rose y Kasper, 2001; Tateyama, 2001; Alcón Soler, 2005; Félix-Brasdefer, 2008; Louw, Derwing y Abbott, 2010)

En 2001 se publica el *Marco Común Europeo de Referencia para las lenguas* (MCER), el cual incluye la pragmática como una parte de la competencia comunicativa, lo que perpetúa su visibilidad en la adquisición de segundas lenguas e impulsa la publicación de nuevos manuales que incluyen ya esta parte de la lengua, frecuentemente ignorada en el pasado. Del mismo modo, el *Plan Curricular del Instituto Cervantes* (PCIC) en 2006 ya enfoca la pragmática desde un punto de vista intercultural y apela así a la competencia y a la identidad intercultural de los aprendientes para ser conscientes de referentes culturales, saberes y comportamientos socioculturales, y habilidades y actitudes interculturales. Estos conocimientos se dividen dentro del PCIC en tres fases para cada nivel del MCER (A1-C2): fase de aproximación, fase de profundización y fase de consolidación, reflejando así el hecho de que la pragmática es una competencia del estudiante independiente del nivel léxico-gramatical que tenga el alumno.

La pragmática aparece también en sistemas de acreditación como el DELE (*Diploma de Español Lengua Extranjera*), o el IDI (*Intercultural Development Inventory*), una herramienta que mide la adaptación intercultural de las personas. Sin embargo, en el caso de los exámenes DELE, la pragmática no se evalúa en profundidad, al menos hasta llegar a los niveles C, a pesar de ser en todos sus niveles un examen basado en el MCER y el PCIC. No obstante, numerosos autores han apoyado la instrucción de la pragmática desde niveles bajos (Takahashi y Dufon, 1989; Wildner-Basset, 1994; Tateyama et al. 1997; Bardovi-Harlig y Dornyei, 1998; Hassall, 2001; Tateyama, 2001; Félix-Brasdefer y Cohen, 2012), obteniendo muy buenos resultados incluso si los estudiantes no tienen bases sólidas de gramática y vocabulario (Kasper, 1997), o mostrando también que la enseñanza de la pragmática en niveles bajos sirve como catalizador de otras competencias y no solo de la competencia pragmática (Tateyama et al., 1997).

La pragmática, pues, es un componente de la lengua que necesita ser enseñado, pudiendo pasar desapercibido para el alumno como comportamiento cultural e incluso individual. Desde los años 80, algunos autores como Thomas (1983) ya empezaron a reflexionar sobre la necesidad de enseñar la pragmática en la clase de lenguas extranjeras, especialmente en el caso de los estudiantes que están en programas de estudio en el extranjero. De esta manera, el nuevo hablante podrá seguir convenciones pragmáticas para

poder expresarse y comunicarse mejor. Igualmente, por parte de los investigadores en pragmática y aprendizaje de lenguas, se han realizado grandes esfuerzos para encontrar las maneras más eficaces de enseñar este conocimiento inherente a los nativos. Sin embargo, el peso de la tradición gramatical en la enseñanza de idiomas ha podido más a la hora de componer los planes de estudios (Lörscher y Schulze, 1988) y los contenidos de los manuales, de modo que la competencia pragmática se ha visto afectada hasta tiempos recientes. A propósito de esto, dos de los investigadores más activos en el campo de la adquisición de la pragmática, Félix-Brasdefer y Cohen (2012), han reflexionado sobre cómo hacer llegar la pragmática a los alumnos a través de la gramática.

Como sabemos, la pragmática va más allá de la simple comunicación y una pragmática errónea puede afectar a la imagen que el hablante da a sus interlocutores en términos de cortesía y etiqueta (Riley, 1989; Kasper, 1990). Desafortunadamente, los profesores siguen sin enseñar la pragmática en el aula por diferentes razones, entre ellas, la falta de tiempo a la hora de cumplir los programas. Muchas veces esta falta de involucración por parte de algunos profesores está provocada por la estructura de algunos manuales, los cuales tradicionalmente se limitan a enseñar los saludos y los registros lingüísticos para cubrir la competencia pragmática sin ir mucho más allá. En otras ocasiones, ni siquiera los mismos profesores son conscientes del componente pragmático del lenguaje hasta que surge la pregunta o hasta que un estudiante comete un error pragmático grave. Por esta razón, urge diseñar herramientas que les faciliten este proceso.

Sin duda, una de estas herramientas son los manuales de español que podemos encontrar en los últimos años en el mercado y que recientemente están haciendo un buen trabajo al respecto. Algunas editoriales españolas como Difusión, con manuales como *Aula*, *Abanico* y *El ventilador* (que dedica una unidad en exclusiva a la pragmática y conocimientos culturales) ya introducen en su organigrama la pragmática de la L2 y la reflexión sobre la pragmática intercultural en el aula. *Destino Erasmus*, de la Universidad de Barcelona, es otro ejemplo de cómo un manual puede acercar la pragmática a los estudiantes¹.

Sin embargo, no existía hasta la fecha de su creación un corpus multimodal que reflejase, por una parte, cómo reaccionan y se comportan los hablantes nativos de español y, por otra, cómo evoluciona la competencia pragmática en la interlengua de los aprendientes. Fue por este motivo por el que se creó el COREMAH (Corpus Español Multimodal de Actos de Habla): por un lado, para investigar esta evolución de la competencia pragmática en estudiantes de una lengua extranjera (no solo en el ámbito pragmalingüístico —relacionado con qué estructuras usar en determinados contextos—, sino también en el sociopragmático —más relacionado con qué es socioculturalmente adecuado y qué no en un determinado contexto—); por otro, para observar, a través del vídeo y las transcripciones anotadas y etiquetadas, las tendencias en las manifestaciones pragmáticas de los hablantes: las de los nativos y las de los no nativos.

2. EL CORPUS

2.1. El contexto teórico-didáctico del COREMAH

Se ha observado a partir de numerosos estudios (Schumann, 1986; Olshtain y Bum-Kulka, 1985; Blum-Kulka y Olshtain, 1986; Bardovi-Harlig y Dornyei, 1998; Kasper y Rose, 1999; Dekeyser, 2007; Taguchi, 2008; Llanes y Muñoz, 2009; Alcón Soler, 2015) que una de las maneras más efectivas de adquirir la pragmática es el contacto con hablantes nativos, en

un proceso llamado *aculturación*², en el cual, inconscientemente, los estudiantes adquieren conocimientos pragmalingüísticos y sociopragmáticos, y van adaptando su manera de comportarse pragmáticamente hacia la cultura meta. Desafortunadamente, esta manera de *absorber* el conocimiento inconscientemente no es tan viable para la mayoría de los estudiantes de una lengua extranjera, que la aprenden en su propio país, y no en el país donde se habla esta lengua. Por este motivo, una de las metas del COREMAH es acercar la experiencia del contacto con la lengua extranjera a través del vídeo y así suplir de algún modo esa carencia que tienen la mayoría de los estudiantes de una lengua extranjera.

A su vez, el hecho de vivir en un país en el que se habla la lengua que se está estudiando no garantiza la adquisición de su pragmática si no se mantienen relaciones sociales con individuos de la cultura meta. Pero no solo se necesita eso para aprender la pragmática, a su vez el estudiante necesita ser consciente de que los comportamientos de los individuos de una sociedad se rigen por diferencias pragmáticas y no por meras diferencias personales e individuales. Es decir, el estudiante necesita ser consciente de que estas diferencias pragmático-culturales existen y que esto diferencia a grupos de personas, no solo desde un plano intercultural, sino también en un plano intracultural, donde se pueden encontrar diferencias de registro, de léxico o incluso gramaticales. Esto es, hay que partir de la activación de la consciencia pragmática para llegar a la adquisición de la pragmática de la lengua meta.

2.2. Otros recursos multimodales

Otros corpus multimodales, como el CID³ (*Corpus of Interactional Data*) para el francés, con ocho horas de conversación grabadas y en el que se analizan los gestos, la prosodia y la risa, entre otros componentes pragmáticos del discurso, o el M4 (*Multimodal Meeting Manager*) de la Comisión Europea⁴, en el cual se han grabado, gracias a habitaciones dotadas de sensores de movimiento, conversaciones en universidades y organismos europeos en seis países, analizando movimientos, gestos y distancias, para mejorar herramientas de detección de emociones en el campo de la inteligencia artificial. La mayoría de estos corpus se han desarrollado desde una perspectiva social, psicológica o emocional. Sin embargo, una pequeña parte ha sido creada para el análisis intercultural, como el *UTEP-ICT Cross-Cultural Multiparty Multimodal Dialog Corpus*⁵, de la Universidad del Sur de California, creado para analizar las percepciones de distintos grupos culturales a la hora de interpretar comportamientos humanos. Un caso especial es *The Spanish Corpus Proficiency Level Training*⁶, de la Universidad de Texas en Austin, dedicado a la enseñanza de idiomas. LAITS (*Liberal Arts Instructional Technology Services*), también de la Universidad de Texas en Austin, cuenta con una página web de ejercicios por niveles, con un compendio de vídeos clasificados por niveles de comprensión en los que un hablante nativo narra un evento o hace una descripción. Igualmente, el *Media Center* de la Universidad de Texas en Austin ha creado unos productos multimedia que, en su apartado de vídeos⁷ contienen narraciones, documentales y hasta entrevistas a hispanohablantes. Estas tres herramientas contienen transcripciones sin anotar de sus vídeos, aunque van acompañados de contenidos de apoyo de vocabulario y gramática para usarlos como materiales en las clases de español. Otro ejemplo de material multimodal, enfocado más en la fonética del mundo hispano, es *The Sounds of Spanish*⁸, de la Universidad de Ohio State.

Aparte de estos recursos, existe, basada en la Universidad de Minnesota, una página web llamada *Dancing with Words: Strategies for Learning Pragmatics in Spanish*⁹, que

contiene, entre otros, recursos para profesores y estudiantes para la enseñanza de la pragmática con una guía de niveles.

Sin embargo, hasta la creación del COREMAH (Vacas, 2017) no se había creado ningún corpus multimodal de español específico para investigación intercultural y la enseñanza de la pragmática de ciertos actos de habla.

2.3. Metodología de la creación del COREMAH

El COREMAH es un corpus de vídeo de hablantes nativos y estudiantes de español de dos niveles diferentes. En el corpus se ven, pues, diferencias interculturales entre hablantes nativos y no nativos a través de la transferencia pragmática que presentan estos últimos. Para ello, se seleccionaron tres grupos de hablantes: uno nativo de hablantes de español peninsular, y dos no nativos estadounidenses de niveles B1 y C1 con inglés como L1. Se les propuso desarrollar una serie de *role-plays* de manera dirigida, siguiendo el modelo de enseñanza de pragmática de Cohen (1995), pero de desarrollo libre, en los siguientes actos de habla: cumplimiento, rechazo y disculpas. Se eligieron estos tres actos de habla en concreto y no otros, aunque se grabaron 10, ya que estos eran los más conflictivos y más complejos en su realización, y en los que más diferencias se encontraban entre los hablantes nativos (en este caso españoles) y los no nativos (estadounidenses).

Se eligió el método de *role-play* para grabar los vídeos, ya que muestra de manera más natural que el DCT¹⁰ (Discourse Completion Test) o la entrevista *role-play* (una adaptación del DCT), y porque las respuestas son más variadas y elaboradas, incentivando la negociación entre hablantes (Bou-Franch y Lorenzo-Dus, 2008). La tipología de corpus, multimodal, fue elegida así para poder analizar también otros componentes de la comunicación que no son estrictamente el lenguaje verbal, pero que aportan significado al mismo. Esto convierte al COREMAH en una herramienta más completa, ya que muestra componentes paralingüísticos, y en los actos comunicativos reales hay una relación fundamental entre lengua, gestos y cognición (Adolphs, Knight y Carter, 2011). De este modo, se podrán analizar otros rasgos de la comunicación como son los gestos, la proxémica o la entonación, y también se podrá hacer a los estudiantes más conscientes de la naturaleza multimodal de la comunicación real nativa (Kress y van Leeuwen, 1996; Ackerley y Coccetta, 2007).

El corpus se grabó en dos escenarios diferentes: el campus estadounidense de Middlebury College en Vermont y el campus satélite que tiene la misma institución en Madrid. Las grabaciones se hicieron con una cámara *High Definition Camcorder Canon XA10 Flash Memory Professional*, para preservar una calidad tanto acústica como visual óptima y, una vez fragmentadas, se subieron a la plataforma Youtube en formato oculto para poder enlazarlas al corpus.

Para escoger a los estudiantes de español se optó por los niveles B1 y C1 del MCER, que se corresponden al nivel de los mismos cuando van a realizar cursos en el extranjero en su tercer año de universidad, y han cursado al menos tres semestres de español (B1); y al nivel de los alumnos no nativos que forman parte del máster de español de Middlebury (C1). Cada grupo no nativo está compuesto por 24 hablantes nativos de inglés: el grupo intermedio está compuesto por, 8 hombres y 16 mujeres, de edades entre los 19 y los 26 años (M=20,7); por otro lado, los sujetos del nivel avanzado son 5 hombres y 19 mujeres, de entre 22 y 55 años (M=29,79); por último, el grupo de control de nativos de español lo componen 7 hombres y 17 mujeres, de entre 18 y 45 años (M=30,08)¹¹.

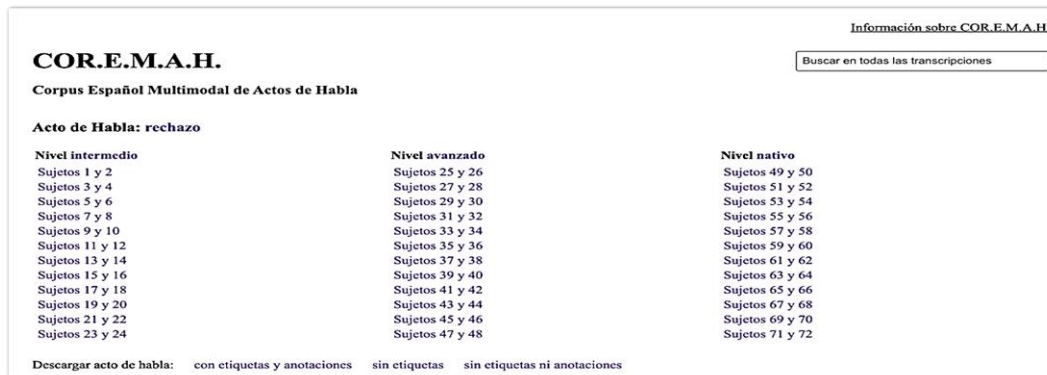


Figura 1. Página de inicio del COREMAH

Antes de empezar las grabaciones se le entregó a cada participante un cuestionario sobre sus conocimientos de español y su historia con el aprendizaje de la lengua¹², y antes de grabar cada acto de habla se les entregaron los contextos de cada situación a los dos participantes, A y B, como se puede ver en el siguiente ejemplo:

Hablante A	Hablante B
Tienes un amigo que habla sorprendentemente bien español. Díselo.	Tu amigo te dice lo bien que hablas español. Reacciona.

Figura 2. Ejemplo de instrucciones para el *role-play*

Se les dio tiempo a los hablantes para que prepararan su escena, y una vez que ambos estaban listos, se empezaba a desarrollar el *role-play*.

Los actos de habla se transcribieron minuciosamente, tanto en su parte de lenguaje verbal como no verbal. Las transcripciones tomaron como modelo la norma de transcripción del grupo Val.Es.Co (Valencia, Español Coloquial), de la Universidad de Valencia, adaptándolas al objeto y naturaleza del estudio, de modo que se eliminaron anotaciones y etiquetaciones que no eran pertinentes a la investigación que se hizo con el COREMAH, como algunas de tipo fonético, ya que el objeto de dicha investigación era averiguar si la competencia pragmática se adquiere al mismo tiempo y nivel que la competencia gramatical. A continuación, se puede ver la tabla de transcripciones que se ha utilizado:

Intervención de un interlocutor identificado como 1	#1:
Palabra truncada/falso inicio	per-, yo quie-
Alargamiento	bue::no, hola::
Solapamiento	inicio [/ final]
Mantenimiento del turno de un participante en solapamiento	= cuando quieras =
Pronunciación marcada o enfática	AHORA
Pronunciación silabeada	a-ho-ra
Exclamación	i!
Interrogación	¿?
Interrogaciones exclamativas	¿i !?
Pausa / silencio	[...]
Ininteligible	[xxx]
Anotaciones sobre la prosodia o gestos (aparecen al margen o dentro de los enunciados)	(muy despacio) (ríen) (se toca la cara)

Tabla 1. Tabla de transcripción del COREMAH

Las etiquetaciones son dependientes de cada acto de habla dada la naturaleza de estos, aunque algunas se comparten entre ellos, como las de intensificadores y mitigadores, destacándose en el corpus mediante diferentes colores, para facilitar la visualización de las tendencias pragmáticas de los hablantes. Igualmente se han respetado los errores lingüísticos, tanto de no nativos como de nativos, para preservar la naturaleza del *role-play*, y se ha hecho una transcripción fiel de cada acto de habla según aparece en el vídeo. Se transcribieron y etiquetaron un total de 108 vídeos, lo que equivale a un total aproximado de 2160 horas de trabajo de transcripción y etiquetación.

2.4. Descripción del COREMAH

El corpus está compuesto por 108 vídeos, 36 para cada acto de habla, de una duración total de 108 minutos, y sus correspondientes transcripciones, contando con un total de 18737 palabras. El corpus, dividido por actos de habla, también se puede organizar por grupo de hablantes, pulsando el enlace del grupo en la página principal.

Para volver a la página principal del corpus, basta con pulsar en el enlace a COREMAH, en cada página en la esquina superior izquierda, o en el enlace de inicio, situado en la misma posición en cada una de las grabaciones. Para acceder a una grabación en concreto, basta con pulsar en los sujetos que se quieran consultar, y este enlace lleva directamente a la página donde se puede ver el vídeo junto a sus transcripciones y los datos de los hablantes: género, nivel de español, tiempo estudiando español y estudios en el extranjero o procedencia de sus profesores de español, entre otros.

Inicio > rechazo > avanzado

	Sujeto 29 55	Sujeto 30 38
Edad		
Género	F	F
Estudios en el extranjero	España	España
Estudios en el extranjero (meses)	10	0
Estudios en el extranjero (años)	0	6
Estudios en el extranjero (meses totales)	10	72
Nivel de español	máster	máster
Autopercepción del nivel de español	alto	avanzado
Tiempo estudiando español (años)	20	24
Otras lenguas en casa	no	no
Otras lenguas que habla	no	no
Contacto con otras culturas	viajes	México, Cuba
Procedencia de sus profesores de español	España, Argentina, Puerto Rico	España

Buscar palabras y etiquetas:

- h-ayu
- h-int-g
- h-org
- o-dis
- o-ind
- o-mit
- h-ins
- h-mit
- h-pre
- o-exc
- o-int
- o-mit-g
- h-int
- h-mit-g
- o-alt
- o-gra
- o-int-g

Rechazo - Avanzado - 29 y 30



Etiquetas y Anotaciones
descargar

Ocultar Etiquetas

Ocultar Anotaciones

#30: ¿Sabes qué? Que-que esto no-no lo entiendo. No entiendo esto muy bien.

#29: =¿Cómo, qué?=-

#30: [Hay tantas] teorías y nombres tan =¿cómo vas a estudiar?=-

#29: [Sí] Bueno, primero [...] se estudia una tarea y que va siempramente s-siempre con un investigador. El nombre.

#30: =Claro=-

Figura 3. Ejemplo de grabación del COREMAH

A su vez, tanto en cada grabación como en cada página del COREMAH, se puede acceder a un buscador, el cual extrae las apariciones de cada palabra (o palabras) o etiqueta en la sección que se está buscando, por grupo, por acto de habla o en todo el corpus. Igualmente, en cada vídeo se puede pulsar cada uno de los botones de las etiquetas, que aparecen en la parte inferior izquierda de la página, para resaltar con colores en el texto las apariciones de esta estrategia, y se pueden destacar varias a la vez.

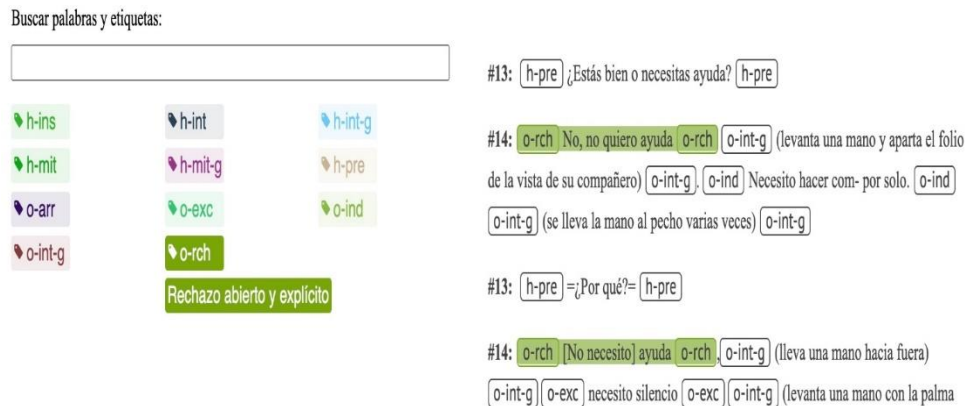


Figura 4. Ejemplo de etiqueta seleccionada con la estrategia resaltada

Las etiquetas que empiezan por “h” son aquellas en las que interviene el hablante que realiza el acto de habla, y las que comienzan por “o” son las pertenecientes a los oyentes, que reaccionan a este acto de habla en concreto. Al poner el cursor del ratón encima de cada una de las etiquetas, se expande el nombre completo para facilitar así la búsqueda de cada estrategia, como se puede ver en la siguiente imagen:

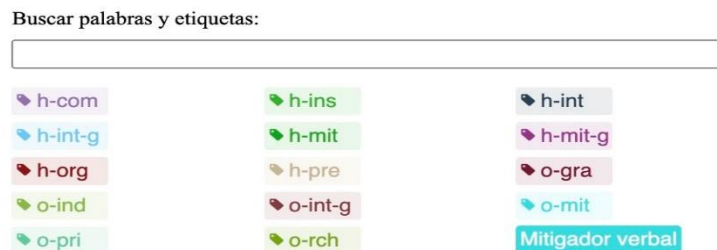


Figura 5. Muestra de cómo se expande el nombre en las etiquetas al poner el cursor del ratón encima

De esta manera, se puede consultar el corpus, pudiéndose descargar, a su vez, cada transcripción de tres modos distintos para su uso: completa, con etiquetaciones de estrategias y anotaciones del lenguaje no verbal; ocultando solo las etiquetas, de modo que se apreciaría el diálogo y se reflejaría el lenguaje no verbal a través de las anotaciones; o solo el texto que dicen los hablantes, en la pestaña de “ocultar anotaciones”. Igualmente se pueden descargar todas las transcripciones de cada acto de habla en estos tres modos, en la parte inferior de cada uno de ellos.

2.5. Los actos de habla elegidos

Se escogió el acto de habla del rechazo a recibir ayuda al observar una diferencia sustancial de respuesta entre los dos grupos para no aceptar ayuda de otros. Ambas culturas son muy distintas en su comportamiento a la hora de reconocer sus propios méritos públicamente: mientras la cultura estadounidense es más propensa a mostrar orgullo de sus méritos, la española tiende más a una respuesta basada en la modestia, más humilde, aunque esta humildad sea solo lingüística, por ser este comportamiento socialmente mejor aceptado que una respuesta más vanidosa, como es la de la cultura estadounidense, que muestra su orgullo ante el hecho de reconocer méritos propios. Tanto en este caso como en los otros dos actos de habla, veremos que los hablantes tratan de salvar su imagen, de acuerdo con los estándares de la cultura a la que pertenecen como nativos.

En el caso del cumplido, se seleccionó por la gran diferencia en frecuencia de aparición de este acto de habla en el comportamiento observado entre estadounidenses y españoles. Mientras en la primera cultura es tremendamente frecuente, en la segunda, es un acto de habla poco frecuente, aunque más complejo. Se escogió el cumplido de halagar la capacidad de hablar la lengua extranjera, ya que evitaría que la variable sexo/género afectase a los resultados a la hora de escoger las estrategias necesarias tanto para hacer un cumplido como para reaccionar a él. Esta decisión se tomó debido al desequilibrio del número de hombres y mujeres que participan en los grupos.

Finalmente, se escogió el acto de habla de la petición de disculpas por no presentarse a una cita para ir al gimnasio por la gran diferencia cronémica en el comportamiento de ambos grupos. Mientras los españoles son un grupo que tiende más al policronismo, este tipo de acto de habla refleja, a su vez, las tendencias más grupales de los españoles o individualistas entre los estadounidenses.

2.6. Resultados de la investigación

Los resultados de la investigación mostraron que los estudiantes de nivel intermedio y los de nivel avanzado respondían de manera similar en los tres actos de habla, y que, salvo en algunos casos muy excepcionales, la pragmática de los dos grupos de estudiantes era más parecida entre ellos (por lo tanto, más influida por su L1) que a la de los hablantes de la lengua que estaban estudiando. Esto corrobora la hipótesis de base del estudio que dio lugar al COREMAH: la pragmática debe ser enseñada explícitamente para poder ser asimilada, aprendida y adquirida. Aunque los estudiantes del grupo avanzado mostraron una mayor riqueza léxica y de estrategias para responder a cada acto de habla (es decir, su competencia pragmlingüística contaba con una mayor riqueza y variabilidad) su comportamiento todavía difería mucho del comportamiento de los nativos, su competencia sociopragmática, por lo tanto, era más débil y su comportamiento era más afín a la cultura estadounidense.

3. IMPLICACIONES PEDAGÓGICAS

3.1. Implicaciones generales

Un corpus multimodal como el COREMAH ayuda, a través de las imágenes, a analizar la adquisición de la cultura a través de otros rasgos de la comunicación que no son

estrictamente verbales, mostrando comportamientos de la vida cotidiana (Cohen 1995, Lorenzo-Dus 2008). Los gestos, la proxémica o la entonación, así como todo aquello que tiene que ver con la semiótica, hacen que los estudiantes se fijen en esta parte de la interacción, y que sean más conscientes de la naturaleza multimodal de la comunicación real nativa (Ackerley y Cocchetta 2007). Kress y van Leeuwen en 1996 ya afirmaban que un texto hablado no tiene solo una naturaleza verbal, sino también visual, y el habla se combina con otros rasgos paralingüísticos como son la expresión facial, la postura del cuerpo o los gestos. Esta multimodalidad que aporta el COREMAH, ayuda a los aprendientes a activar su conocimiento del mundo en un plano mucho más multidimensional que un texto, o un audio. Además, como *input* es más natural que cualquier material audiovisual creado o no *ad hoc* (una serie, un corto, una película, etc.) y guionizado. La mayoría de los estudiantes de una lengua extranjera no tienen la oportunidad de vivir en un país donde se habla por un periodo relativamente largo de tiempo, o de tener amistades nativas, con lo cual no tendrán muchas oportunidades de presenciar actos de habla o comunicación real en su día a día. Un corpus multimodal proporcionará, de este modo, a estudiantes y profesores, acceso directo a información concreta, además, sobre los participantes de la conversación, como son, por ejemplo, su actitud, su estado de ánimo, sus relaciones sociales, así como la distancia entre hablantes, las expresiones faciales, los gestos que utilizan, la entonación o las miradas (Ackerley y Cocchetta, 2007).

Debemos insistir sin duda en el papel de la consciencia lingüística en el campo de la adquisición de la pragmática. Schmidt (1990, 1993, 1995) describió a través de su *Hipótesis del Reconocimiento*, que hay algo estrictamente necesario para que los aprendientes de una lengua extranjera adquieran la competencia pragmática: la atención, la consciencia y el reconocimiento de los rasgos pragmáticos. Por lo tanto, para que el *input* pragmático se vuelva *intake*, según la hipótesis de Krashen (1985) del *input comprensible*, o *hipótesis del monitor*, los estudiantes deben ser conscientes de las diferencias interculturales y pragmáticas. Schmidt propone dos maneras básicas de enseñar la pragmática: de forma inconsciente, es decir, a través de la conversación; y de forma consciente, es decir, de manera explícita. A partir de esta propuesta, distintos autores han analizado los efectos de la enseñanza de la pragmática, obteniendo resultados más positivos en cuanto a la enseñanza explícita de la pragmática (House y Kasper, 1981; Blum-Kulka, 1982; Olshtain y Cohen, 1990; Bardovi-Harlig et al., 1991; Bouton, 1994; Rose, 1994; Schmidt, 1995; Cohen, 1996; House, 1996; Tateyama et al., 1997; Kasper y Rose, 1999; Kasper, 2001; Tateyama, 2001; Yoshimi, 2001; Alcón Soler, 2005; Félix-Brasdefer, 2008; Louw, Derwing y Abbott, 2010; Eslami y Liu, 2013).

El primer paso para que el estudiante pueda adquirir la pragmática de la lengua extranjera que está aprendiendo es activar la consciencia pragmática, es decir, llamar la atención de los rasgos pragmáticos de las dos culturas, la cultura materna y la cultura de la lengua meta, para así poder apreciar las diferencias. Por un lado, existirán rasgos pragmáticos comunes a las dos culturas y, por otro, rasgos diferentes. Serán estos últimos en los que tendremos que enfocarnos, porque son los que constituyen una puerta fácil de entrada para el error pragmático a través de la transferencia de la L1. De este modo, la pragmática intercultural juega, pues, uno de los papeles más importantes a la hora de enseñar y aprender la pragmática de la L2.

Por lo tanto, como hemos visto, hay dos modos básicos de enseñar la pragmática en la clase de ELE: de manera implícita y de manera explícita. De manera implícita, confiamos en las capacidades del estudiante para descifrar qué es relevante y qué no, en términos pragmáticos, y las diferencias que hay entre su cultura y la nuestra (Schmidt 1993). De

este modo, no estaremos cien por cien seguros de si el estudiante se ha dado cuenta de las diferencias interculturales y, más aún, si el estudiante está adquiriendo un patrón pragmático o no. Aunque sí es posible la adquisición inconsciente de algunos de estos patrones pragmáticos de manera implícita, lo cierto es que la mayoría de los estudios apuntan las ventajas de la enseñanza consciente y explícita de la pragmática, no solo para entender la actuación del nativo, sino también para comprender la organización de su sistema pragmalingüístico y sociopragmático.

3.2. Cómo usar el corpus

El corpus puede ser usado de diversas maneras. Por ejemplo, usando solo las transcripciones de los vídeos (con anotaciones) a modo de obra de teatro, como si fueran acotaciones, para analizar el lenguaje no verbal, la proxémica entre los hablantes, los movimientos de los mismos, o incluso el tono de voz de los hablantes. Así, se puede reflexionar primero, desde la propia lengua del estudiante, hacia la lengua meta.

Una actividad interesante que se podría llevar a cabo con este corpus es proponer a los estudiantes el juego del *role-play*, que interactúen ellos en parejas, y una vez han sido expuestos todos los intercambios, mostrar a los estudiantes un ejemplo de nativos de español, con una serie de preguntas previas a la muestra del vídeo, como, por ejemplo, "¿cómo crees que sería esta conversación entre dos personas nativas de español?" o "¿crees que en todas las culturas harían lo mismo?". De este modo, haremos que el estudiante y la clase reflexionen acerca de la variabilidad cultural incluso antes de ver un vídeo de nativos. Una vez mostrado el vídeo, podemos proponerles una serie de actividades de consciencia intercultural, como, por ejemplo, que analicen el tono de voz de los hablantes, sus gestos, cómo se ha desarrollado el acto de habla, las diferencias con lo que han visto en sus compañeros de clase, y lo que detectan que es o no adecuado en su cultura.

Otro modo de usar el corpus y, sobre todo, de aprovechar su multimodalidad, es visualizar cualquiera de los vídeos, pero sin sonido, hacer lo mismo con otros, y tratar de emparejar los vídeos con las transcripciones. Esto se puede hacer con la transcripción pura (sin anotaciones del lenguaje no verbal) o con anotaciones. Además de resultar una actividad con buena acogida entre los estudiantes, los ayuda a ver las grandes diferencias interculturales entre los grupos, si esto se hace con vídeos de nativos de español y vídeos de nativos de inglés. Este ejercicio también se puede desarrollar poniendo los vídeos sin sonido para que los estudiantes escriban lo que creen que dicen los hablantes. Luego, al ver el vídeo con sonido, pueden comparar los diálogos que ellos han creado con los que realmente se han producido y, de este modo, comparar culturas, la suya con la de los hablantes de inglés, o con la de los hablantes nativos de español.

También se pueden usar las transcripciones sin anotaciones para observar qué tipo de estructuras o vocabulario son más comunes en cada cultura, o los patrones de cada acto de habla. Una diferencia muy notable entre los hablantes españoles y estadounidenses fue el uso de palabras malsonantes, prácticamente ausente en estos últimos. Se puede usar para reflexionar, pues, sobre la variabilidad léxica y estructural y la riqueza de los hablantes no nativos del nivel avanzado, en comparación con los recursos limitados que tiene un hablante intermedio. Este tipo de actividad, igualmente, es idónea para introducir en la clase el lenguaje informal y coloquial. Así, se podrían ver en contexto locuciones y conjunciones que no se suelen estudiar en la clase de español, y que forman parte del día a día de las conversaciones entre nativos.

De la misma manera, se les pueden ofrecer las transcripciones a los estudiantes (sin anotaciones) para que las interpreten, como si fueran actores, e inmediatamente después de esta interpretación, mostrar el vídeo de cada transcripción y analizar las diferencias en el lenguaje no verbal. Una parte muy interesante de esta actividad sería que los mismos estudiantes se grabaran, para así poder analizar de una manera más detallada sus movimientos.

Igualmente se pueden analizar las impresiones de los alumnos al ver los vídeos de los nativos españoles: qué impresiones tienen de las personas que están actuando. De este modo, se puede apreciar cómo una reacción natural de un nativo podría ser malinterpretada por un no nativo, y viceversa. Se puede comparar, de la misma manera, cómo ha resuelto cada grupo el conflicto que se planteaba en cada conversación y qué elementos son comunes en cada cultura. Por ejemplo, en el caso del *role-play* de las disculpas es común entre los nativos de español terminar con besos y abrazos, algo que no se ve en las grabaciones de los estadounidenses. Es más, si los estudiantes están en un programa de estudios en el extranjero, se les puede proponer que observen estos actos de habla en la vida real y que lleven sus notas u observaciones a clase para que se puedan discutir en grupo.

Por último, en un ejercicio de análisis del discurso, también se pueden usar las transcripciones junto a los vídeos del grupo de los hablantes nativos de español para examinarlas en un plano mucho más profundo, incluso para grupos de profesores, analizando lingüísticamente el texto y viendo qué marcadores del discurso se han utilizado. También se podría ver qué mitigadores e intensificadores aparecen, las pausas que se han producido, el tono, los patrones seguidos, las estructuras, e incluso lo que no aparece, pero ellos esperaban que apareciera, o lo que les ha sorprendido que aparezca (como palabras malsonantes, o el reproche, que parece algo exclusivo del grupo nativo de españoles).

En este artículo solo se han expuesto algunas de las actividades que se pueden desarrollar con el corpus, sin embargo, al ser un recurso multimodal, se podrían desarrollar muchas más actividades en torno a la pragmática para su reflexión, explotación y futura adquisición.

4. CONCLUSIONES Y FUTURAS INVESTIGACIONES

A través de la construcción del corpus, se pudo observar que a no ser que se explique explícitamente, o se haya vivido por un gran periodo de tiempo en el país donde la L2 se habla, es muy difícil que los estudiantes hayan adquirido la pragmática de la lengua que estudian (Vacas, 2017). En la competencia sociopragmática, los estudiantes avanzados mostraban unos patrones de conducta muy similares a los de sus compañeros de intermedio, y más alejados de los que mostró el grupo de nativos de español. Sin embargo, en la competencia pragmalingüística, esto es, en cuanto a estructuras y léxico, los estudiantes avanzados tenían un abanico mucho más amplio de recursos que sus compañeros de nivel intermedio. Estos hallazgos en la investigación vinieron a ratificar la hipótesis básica del estudio en el que se basó el corpus: que la pragmática se debe enseñar de manera explícita.

El COREMAH es, por lo tanto, un recurso multimodal que muestra el registro informal y la pragmática de estos tres actos de habla complejos para que puedan ser analizados, se pueda reflexionar sobre ellos y, de este modo, adquirir la pragmática del español, ya que este tipo de recursos típicamente no aparecen en los manuales al uso de español, ni siquiera

en los de nivel avanzado. A pesar del reto que supuso transcribir y etiquetar el lenguaje no verbal, ya que no existía en español ningún corpus que hiciera esto en el momento en el que se creó el COREMAH, el hecho de haberlo creado abre la puerta a nuevos corpus y nuevos materiales para la clase de ELE que sí lo hagan en el futuro. Junto con estos tres actos de habla se grabaron otros siete más que también podrán ser transcritos y analizados en un futuro.

Hay otras muchas opciones que se podrían tener en cuenta para futuros trabajos con corpus multimodales de actos de habla, como es el analizar otros actos de habla, igualmente de diferentes variedades del español que no sea el peninsular (incluso en parejas mixtas, con hablantes de dos variedades distintas), analizar y etiquetar la prosodia, el tono, los silencios, o incluso la proxémica más detalladamente. Otras opciones que producirían sin duda resultados muy interesantes serían corpus en los que las parejas estuvieran formadas por un nativo y un no nativo, corpus bilingües y corpus de hablantes de herencia, para ver qué aspectos de la pragmática nativa de su lengua materna conservan estos hablantes y cuáles no.

Este corpus y el estudio lingüístico que lo hizo posible tenían como meta, además del análisis y los datos obtenidos, multiplicar las opciones y animar a futuros investigadores y profesores a construir más corpus y actividades basadas en la idea de los *role-plays* y el comportamiento natural, de modo que los futuros estudiantes de español tengan acceso a modelos pragmáticos para evitar así el fallo pragmático y promover la adquisición de la esta competencia a todos los niveles.

NOTAS

- 1 Para ver un análisis más exhaustivo de estos y otros manuales, recomiendo leer Vacas (2017), donde se comentan y analizan varias actividades.
- 2 Según el Instituto Cervantes: "El proceso de aculturación consiste, por un lado, en la incorporación de elementos de la nueva cultura y, por otro, en el reajuste de los patrones culturales del individuo o grupo, motivados ambos por la necesidad de reorientar sus pensamientos, sentimientos y formas de comunicación a las exigencias de las realidades externas". Extraído de: http://cvc.cervantes.es/ensenanza/biblioteca_ele/diccio_ele/diccionario/aculturacion.htm
- 3 <http://www2.lpl-aix.fr/~otim/ressources.html>
- 4 <https://cordis.europa.eu/project/id/IST-2001-34485>
- 5 <http://nld.ict.usc.edu/group/corpora/utep-ict-cross-cultural-multiparty-multimodal-dialog-corpu>
- 6 <https://www.laits.utexas.edu/spt/intro>
- 7 <http://laits.utexas.edu/spanvid/>
- 8 <http://dialectos.osu.edu/advanced-search>
- 9 http://carla.umn.edu/speechacts/sp_pragmatics/home.html
- 10 DCT, en sus siglas en inglés proviene de "Discourse Completion Test. una técnica usada en estudios de pragmática y lingüística en general por la cual se provoca una respuesta de los sujetos investigados ante una situación a través de un guion en el que se explica la situación concreta en la que se desarrolla el acto de habla, o el rasgo lingüístico concreto que se quiere investigar.
- 11 Para ver más detalles de los hablantes se puede consultar la página web del COREMAH, en su apartado de información.
- 12 Para ver más detalles sobre esto, consúltese Vacas (2017).

REFERENCIAS BIBLIOGRÁFICAS

- Ackerley, K., y Cocchetta, F. (2007). Enriching Language Learning through a Multimedia Corpus. *ReCALL*, 19(3), pp. 351-370.
- Adolphs, S., Knight, D., y Carter, R. (2011). Capturing context for heterogeneous corpus analysis: Some first steps. *International Journal of Corpus Linguistics*, 16 (3), pp. 305-324.
- Alcón Soler, E. (2005). Does Instruction Work for Learning Pragmatics in the EFL Context?. *System: An International Journal of Educational Technology and Applied Linguistics*, 33 (3), pp. 417-435.
- Alcón Soler, E. (2015). Pragmatic Learning and Study Abroad: Effects of Instruction and Length of Stay. *System: An International Journal of Educational Technology and Applied Linguistics* 48, pp. 62-74.
- Bardovi-Harlig, K, Hartford, B.A.S., Mahan-Taylor, R., Morgan M.J. y Reynolds, D.W. (1991). Developing Pragmatic Awareness: Closing the Conversation. *ELT Journal*, 45(1), pp. 4-15.
- Bardovi-Harlig, K, y Dornyei, Z. (1998). Do Language Learners Recognize Pragmatic Violations? Pragmatic Versus Grammatical Awareness in Instructed L2 Learning. *TESOL Quarterly*, 32 (2), pp. 233-262.
- Bardovi-Harlig, K. y Mahan-Taylor, R. (2003) *Teaching Pragmatics*. Washington, DC: United States Department of State. Consultado el 8 de septiembre de 2020 en la web: <https://www.indiana.edu/~dsls/publications/printtableofcontents.doc>
- Blum-Kulka, S. (1982). Learning to Say What You Mean in a Second Language: A Study of the Speech Act Performance of Learners of Hebrew as a Second Language. *Applied Linguistics* 3(1), pp. 29-59.
- Blum-Kulka, S. y Olshtain, E. (1986). Too Many Words: Length of Utterance and Pragmatic Failure. *Studies in Second Language Acquisition*, 8 (2), pp. 165-179.
- Bou-Franch, P. y Lorenzo-Dus, N. (2008). Natural versus elicited data in cross-cultural speech act realisation: The case of requests in Peninsular Spanish and British English. *Spanish in Context*, 5 (2), pp. 246-277.
- Bouton, L.F. (1990). The effective use of implicature in English: Why and how it should be taught in the ESL classroom". En: Bouton L. y Kachru Y. Eds. *Pragmatics and language learning*, 1, University of Illinois, Urbana-Champaign: Division of English as an International Language.
- Bouton, L.F. (1992). The interpretation of implicature in English by NNS: Does it come automatically -without being explicitly taught?. En: Bouton, L. y Kachru, Y. Eds. *Pragmatics and language learning*, 3, University of Illinois, Urbana-Champaign: Division of English as an International Language.
- Bouton, L.F. (1988) "A Cross-Cultural Study of Ability to Interpret Implicatures in English", *World Englishes*, 7 (2), pp. 183-196.
- Bouton, L.F. (1994). Conversational Implicature in a Second Language: Learned Slowly When Not Deliberately Taught. *Journal of Pragmatics*, 22 (2), pp. 157-167.
- Cervantes, Instituto (2002) *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación*. Madrid: Secretaría General Técnica del MECD, Subdirección General de Información y Publicaciones Anaya.
- Cervantes, Instituto (2006) *Plan curricular del Instituto Cervantes niveles de referencia para el español*. Madrid: Biblioteca Nueva.
- Chamorro Guerrero, M.D., Lozano López, G., Ríos Rojas, A., Rosales Varo, F., Ruiz Campillo, J.P., Ruiz Fajardo, G. (2010) *El ventilador: curso de español de nivel superior C1*. Barcelona: Difusión.
- Cohen, A.D. (1995). Investigating the production of speech act sets. Speech Acts across Cultures: Challenges to Communication in a Second Language. *Studies on Language Acquisition*, 11, pp. 21-43.
- Cohen, A.D. (1996). Developing the Ability to Perform Speech Acts. *Studies in Second Language*

Acquisition, 18(2), pp. 253-267.

- Corpas, J., García, E. y Gamendia, A., (2013) *Aula 1: curso de español*. Barcelona: Difusión.
- Corpas, J., Gamendia, A. y Soriano, C. (2013) *Aula internacional 2*. Barcelona: Difusión.
- Corpas, J., Gamendia, A. y Soriano, C. (2015) *Aula 3: curso de español*. Barcelona: Difusión.
- Chamorro, M.D., Lozano, G., Martínez, P., Muñoz, B., Rosales, F., Ruiz Campillo, J. P. y Ruiz G. (2010) *Abanico. Libro del alumno*. Barcelona: Difusión.
- DeKeyser, R. (2007) *Practice in a Second Language: Perspectives from Applied Linguistics and Cognitive Psychology*. Cambridge; New York: Cambridge University Press.
- Eslami, Z.R., y Liu, C.N. (2013). Learning Pragmatics through Computer-Mediated Communication in Taiwan. *International Journal of Society, Culture & Language*, 1(1), pp. 52-73.
- Félix-Brasdefer, J.C. (2008). Teaching Pragmatics in the Classroom: Instruction of Mitigation in Spanish as a Foreign Language. *Hispania: A Journal Devoted to the Teaching of Spanish and Portuguese*, 91 (2), pp. 479-494.
- Félix-Brasdefer, J.C. y Cohen, A.D. (2012). Teaching Pragmatics in the Foreign Language Classroom: Grammar as a Communicative Resource. *Hispania*, 95(4), pp. 650-669.
- Hassall, T. (2001). Modifying Requests in a Second Language. *International Review of Applied Linguistics in Language Teaching*, 39 (4), pp. 259-283.
- House, J. (1996). Developing Pragmatic Fluency in English as a Foreign Language: Routines and Metapragmatic Awareness. *Studies in Second Language Acquisition*, 18(2), pp. 225-252.
- House, J. y Kasper G. (1981). Politeness Markers in English and German. En: Columas, F. Ed. *Conversational Routine: Explorations in Standardized Communication Situations and Prepatterned Speech*. Mouton. M: The Hague, pp. 157-185.
- Kasper, G. (1990). Linguistic Politeness: Current Research Issues. *Journal of Pragmatics*, 14 (2), pp.193-218.
- Kasper, G. (1997). *Can pragmatic competence be taught?* (NetWork #6) [documento HTML]. Honolulu: University of Hawai'i, Second Language Teaching & Curriculum Center. Consultado el 8 de septiembre de 2020 en la Web: <http://www.nflrc.hawaii.edu/NetWorks/NW06/>
- Kasper, G. (2001). Classroom Research in Interlanguage Pragmatics. En: Kasper G., y Rose K.R., Eds. *Pragmatics in language teaching*. Cambridge; New York: Cambridge University Press. Cambridge applied linguistics series.
- Kasper, G. y Rose, K.R. (1999). Pragmatics and SLA. *Annual Review of Applied Linguistics*, 19, pp. 81-104.
- Krashen, S. (1985) *The Input Hypothesis: Issues and Implications*. London; New York: Longman.
- Kress, G. y van Leeuwen, T. (1996) *Reading Images: The Grammar of Visual Design*. London; New York: Routledge.
- Llanes, A. y Muñoz, C. (2009). A Short Stay Abroad: Does It Make a Difference?. *System: An International Journal of Educational Technology and Applied Linguistics*, 37(3), pp 353-365.
- Lörscher, W. y Schulze R. (1988). On Polite Speaking and Foreign Language Classroom Discourse. *International Review of Applied Linguistics in Language Teaching* 26 (3), pp. 183-199.
- Louw, K.J., Derwing, T. y Abbott M.L. (2010). Teaching Pragmatics to L2 Learners for the Workplace: The Job Interview. *Canadian Modern Language Review/La Revue Canadienne des Langues Vivantes* 66 (5), pp. 739-758.
- Olshtain, E. y Blum-Kulka, S. (1985). Degree of Approximation: Nonnative Reactions to Native Speech Act Behavior. En: Gass, S.M. y Madden, C.G. Ed. *Input in Second Language Acquisition*. Cambridge, Mass.: Newbury House, pp. 303-325.
- Olshtain, E. y Cohen, A.D. (1990). The Learning of Complex Speech Act Behaviour. *TESL Canada Journal/Revue TESL du Canada*, 7(2), pp. 45-65.
- Riley, P. (1989). "Well Don't Blame Me! On the Interpretation of Pragmatic Errors. En: Oleksy, W., Ed. *Contrastive Pragmatics*. Amsterdam: John Benjamins, pp. 231-249.
- Rose, K. R. (1994). Pragmatics consciousness-raising in an EFL context. *Pragmatics and language learning*, 5, pp. 52-63.
- Rose, K.R. y Kasper G. (2001) *Pragmatics in language teaching*. Cambridge; New York: Cambridge University Press. Cambridge applied linguistics series.

- Schmidt, R. (1990). *The Role of Consciousness in Second Language Learning*. *Applied Linguistics* 11 (2), pp. 129-158.
- Schmidt, R. (1993). Consciousness, Learning, and Interlanguage Pragmatics. En Kasper, G. y Blum-Kulka, S. Eds. *Interlanguage Pragmatics*. New York; Oxford: Oxford University Press, pp. 21-42.
- Schmidt, R. (1995). Consciousness and Foreign Language Learning: A Tutorial on the Role of Attention and Awareness in Learning. En: Schmidt R. Ed. *Attention and Awareness in Foreign Language Learning*. Honolulu: Second Language Teaching & Curriculum Center, University of Hawai'i at Manoa, cop., pp. 1-64.
- Schumann, J.H. (1986). Research on the Acculturation Model for Second Language Acquisition. *Journal of Multilingual and Multicultural Development*, 7(5), pp. 379-392.
- Taguchi, N. (2008). Cognition, Language Contact, and the Development of Pragmatic Comprehension in a Study-Abroad Context. *Language Learning: A Journal of Research in Language Studies*, 58(1), pp. 33-71.
- Takahashi, S., y DuFon M.A. (1989). *Cross-Linguistic Influence in Indirectness: The Case of English Directives Performed by Native Japanese Speakers*. Manuscrito no publicado, Departamento de Inglés como lengua extranjera, Universidad de Hawai en Manoa 28 sep. 1989, pp. 1-43.
- Tateyama, Y. (2001). Explicit and Implicit Teaching of Pragmatic Routines: Japanese sumimasen. En: Kasper G., y Rose K.R., Eds. *Pragmatics in language teaching*. Cambridge; New York: Cambridge University Press. Cambridge applied linguistics series.
- Tateyama, Y., Kasper, G., Mui, L.P., Tai, H. y Thananart, O. (1997). Explicit and implicit teaching of pragmatics routines. *Pragmatics and language learning*, 8, pp. 145-177.
- Thomas, J. (1983). Cross-Cultural Pragmatic Failure. *Applied Linguistics Applied Linguistics* 4 (2), pp. 91-112.
- Vacas, M. (2017) *Diseño y compilación de un corpus multimodal de análisis pragmático para la aplicación a la enseñanza del español L2/LE* (Tesis sin publicar). Universidad Autónoma de Madrid.
- Vacas, M. (2018). *Multimodal corpus of Spanish speech acts: Main features and potential pedagogical uses*. En: F. Rosell-Aguilar, T. Beaven, y M. Fuertes (Eds.). *Innovative language teaching and learning at university: Integrating informal learning into formal language education* (pp. 89-97). Voillans: Research-publishing.net. <https://doi.org/10.14705/rpnet.2018.22.779>
- Wildner-Bassett, M. (1994), "Intercultural Pragmatics and Proficiency: "Polite Noises" for Cultural Appropriateness. *IRAL: International Review of Applied Linguistics in Language Teaching* 32 (1), pp. 3-17.
- Yoshimi, D.R. (2001). Explicit Instruction and JFL Learners' Use of Interactional Discourse Markers. En: Kasper G., y Rose K.R., Eds. *Pragmatics in language teaching*. Cambridge; New York: Cambridge University Press. Cambridge applied linguistics series.

Estudio comparativo sobre el uso de recursos metadiscursivos en corpus de hablantes nativos y no nativos grecófonos de español

A comparative corpus-based study on the use of metadiscourse resources between native and Greek non-native speakers of Spanish

Panagiota Salapata

Universidad Antonio de Nebrija

psalapat@otenet.gr

Angélica Alexopoulou

Universidad Nacional y Kapodistriaca de Atenas

aalexop@spanll.uoa.gr

RESUMEN

El presente estudio tiene como objetivo estudiar las estrategias metadiscursivas en el género carta al director en contextos formales. Para ello se recoge una muestra de ciento veinticinco cartas (27.969 palabras) procedentes de hablantes nativos españoles y hablantes no nativos griegos. El estudio se sustenta en las aportaciones de la retórica intercultural, área de investigación en la adquisición de segundas lenguas, cuyo propósito consiste en el estudio del discurso entre individuos provenientes de distintas culturas. La categorización de los recursos retóricos se realiza a partir del modelo interpersonal de Hyland, que se adapta a las características del género carta al director. Los resultados obtenidos sugieren que existen diferencias significativas a nivel global y por categoría tratada en la mayoría de las comparaciones realizadas.

Palabras clave: marcadores metadiscursivos, carta al director, retórica intercultural, estudiantes novatos, competencia escritora.

ABSTRACT

The aim of the present study is to explore metadiscourse strategies in the genre letter to the editor produced in educational contexts. For this purpose, we collected a sample of one hundred twenty-five letters to the editor (27.969 words) from Greek novice students in L2 Spanish and L1 Spanish speakers. The study adopts an intercultural rhetoric approach in order to study discourse between individuals from different cultures. The classification of the rhetorical features is based on Hyland's typology (2017), adapted to meet the characteristics of the letter to the editor genre. The results suggest statistical significance in almost all global and per-category comparisons.

Keywords: metadiscourse markers, letter to the editor, intercultural rhetoric, novice students, writing skills

1. INTRODUCCIÓN

Las estrategias metadiscursivas han recibido muy poca atención en las investigaciones realizadas en el periodismo interactivo en lengua materna (L1), así como en los estudios contrastivos (Dafouz Milne, 2006; González Arias, 2014; Muñoz, 2016). La revisión de la literatura relevante indica que durante los últimos años el análisis de los marcadores metadiscursivos ha despertado gran interés entre los investigadores, sobre todo, en lengua inglesa. Gran parte de estos esfuerzos se ha concentrado en los géneros artículo de investigación y ensayo argumentativo (Ädel, 2006), tanto en L1 como en los estudios contrastivos. De igual forma, se observa gran interés por comparar los marcadores metadiscursivos en textos escritos por autores expertos y novatos. En lo que se refiere al griego moderno, destaca el análisis de Koutsantoni (2003;2005) cuyos resultados, aunque se centran en el discurso académico especializado, podrían servir como referencia para la interpretación de otros géneros discursivos. En lengua española destacan los trabajos sobre diversos géneros pertenecientes al discurso periodístico, que cimientan las bases para futuras indagaciones en el campo. Especialmente, los estudios de González Arias (2014) y Moya Muñoz (2016), que analizan el comentario de lectores en los foros de la prensa chilena y española, abren una nueva línea de investigaciones en la esfera pública de la actividad discursiva, esto es, el periodismo participativo. Los estudios comparativos entre distintas L1 son escasos, aunque últimamente se observa cierto interés en español por estudiar los patrones retóricos entre español e inglés en el discurso profesional (Dafouz Milne, 2006; Suau Jiménez, 2011a; 2011b).

De igual forma, los estudios realizados con aprendientes griegos en lengua extranjera (L2/LE) son prácticamente nulos, con excepción del trabajo de Mayor (2006) sobre los examinados en el diploma IELTS, que puede ofrecer valiosas informaciones sobre las preferencias retóricas de los aprendientes griegos, aunque sea en inglés L2. Si bien los estudios en lengua extranjera son menos numerosos, el análisis de las estrategias metadiscursivas entre L1 y L2 es más profuso, aunque se basan siempre en inglés como punto de referencia de la actuación lingüística nativa (Hyland y Milton, 1997; Hinkel, 2005). Entre las investigaciones contrastivas cabe señalar, por una parte, la de Neff-van Aertselaer y Dafouz (2008) con informantes ingleses y españoles y, por otra, las de Koutsantoni (2003; 2005) y de Hatzitheodorou y Mattheoudakis (2011) con grecófonos y anglófonos. En particular, las investigaciones de Koutsantoni y Neff-van Aertselaer y Dafouz son las únicas que en vez de contrastar HNN con HN inexpertos, se complementan con el análisis de la actuación lingüística de hablantes nativos expertos en el campo.

No obstante, la mayoría de los estudios contrastivos se basa en grandes corpus que corresponden a diversos géneros, tópicos y actividades. Según Hinkel (2002: 59), los estudios comparativos tienen enorme importancia y utilidad cada vez que se realizan comparaciones detalladas en contextos próximos y actividades. En caso contrario, los resultados obtenidos pueden ser erróneos al no ser representativos de la actuación lingüística de una comunidad discursiva. Además, la pluralidad de los modelos metadiscursivos y la falta de acuerdo entre los teóricos sobre la categorización de los ítems que los componen dificultan la comparación de los resultados obtenidos. De igual forma, estos estudios, raras veces, ofrecen un análisis holístico de las estrategias metadiscursivas, limitándose al análisis de subcategorías muy específicas. Ello impide obtener una visión global sobre la interrelación de las funciones retóricas y cuál

predomina en un género específico. Teniendo en cuenta todo lo anterior, en el presente trabajo se presentan los resultados de un estudio contrastivo entre HNN grecófonos y HN hispanófonos a la hora de escribir una *carta al director*, género del que disponemos de muy poca información sobre su estructura o patrones recurrentes.

De hecho, las cartas al director constituyen un género discursivo que ha recibido muy poca atención en los estudios textuales y, especialmente, en los contrastivos por su tendencia a priorizar en el discurso académico y/o profesional, dejando de lado el análisis del discurso propio del ámbito público o privado, igual de esencial para la comunicación oral o escrita en una lengua extranjera. El discurso como acción e interacción en la sociedad implica un gran número de usuarios que utilizan activamente la escritura y el habla no simplemente como escritores, hablantes, mediadores, oyentes o lectores, sino también como miembros de una o más comunidades discursivas, interactuando a diario. Una de estas comunidades –virtual o real– es la que se forma por los autores y los lectores de un diario quienes interactúan y dialogan entre sí en un nuevo ecosistema comunicativo. Ante este fenómeno incluso los medios más tradicionales se han visto obligados a ceder, en mayor o menor grado, la palabra a los lectores en sus ediciones digitales, con la creación de blogs, foros y una sección dedicada a las cartas de los lectores que permita mayor interacción entre un público que traspasa las fronteras nacionales para acoger en sus senos distintas comunidades lingüísticas y culturales. Por lo tanto, el tradicional lector se convierte en «prosumidor», es decir, en productor y consumidor de la información al mismo tiempo, desarrollando múltiples actividades interactivas con otros usuarios (Martínez Rodríguez, 2005: 272).

Ahora bien, todos estos cambios en la forma en que comunican e interactúan las personas en contextos naturales tienen impacto directo sobre la didáctica de lenguas, tanto maternas como extranjeras, cuyo propósito y materiales educativos se tienen que ajustar a la aparición de nuevos géneros o la hibridación de los ya existentes, como es el caso del género carta al director. Especialmente, en contextos de español como lengua extranjera (E/LE) la interacción constituye uno de los aspectos clave junto con la comprensión, la expresión y la mediación a lo largo de las distintas etapas de enseñanza-aprendizaje. Ello se hace evidente por el papel central que se le otorga a las estrategias y actividades de interacción en el *Marco común europeo de referencia para las lenguas* (MCER) (Consejo de Europa, 2002), donde se insiste sobre su importancia para facilitar la comunicación –en el ámbito público, profesional, educativo y personal– entre personas que tienen distintas lenguas maternas. No obstante, en la literatura circulante sobre los géneros periodísticos, así como en los manuales de E/LE, escasean las referencias tanto a la estructura como a las características del género carta al director, puesto que se centran sobre todo en las de índole comercial y laboral. Por lo tanto, ¿cuáles son las convenciones del género carta al director que el estudiante de E/LE debe conocer?, ¿con qué estrategias retóricas tiene que estar familiarizado para lograr persuadir a su audiencia?, ¿su nivel de competencia escritora influye en la cantidad y tipo de recursos retóricos que emplea? Por último, estos modos de persuasión ¿siguen un patrón retórico común o difieren de cultura a cultura y de lengua a lengua?

A todas estas cuestiones ha intentado responder, aunque fuera parcialmente, la retórica contrastiva (RC), área de investigación en segundas lenguas iniciada por Kaplan (1966) a finales de los años sesenta en los Estados Unidos. Dicho estudioso, partiendo de la premisa de que la lógica no tiene carácter universal, sino que varía de cultura a cultura, busca ofrecer un modelo contrastivo destinado al análisis de textos escritos en L1 y L2 con el fin de hallar aquellas variables retóricas que difieren en las diversas comunidades de

habla. Este primer intento de Kaplan ha sido fuertemente criticado, entre otras razones, por su carácter etnocentrista, es decir, por privilegiar la escritura en inglés y por enfocarse solo en el producto de la escritura y, en particular, en la estructura del párrafo. No obstante, estas críticas han contribuido al desarrollo y la mejora de los paradigmas teóricos en los que se sustenta la RC, así como a su evolución de un enfoque restringido al análisis de textos en L1 y L2 a una visión mucho más amplia de análisis contrastivos entre comunidades lingüísticas tanto a nivel transcultural como intercultural. Ello se hace evidente en la propuesta de Connor (2008) de sustituir el binomio retórica contrastiva por el término *retórica intercultural* (RI) para incorporar un amplio abanico de teorías y modelos de análisis del discurso y designar que ninguna tradición retórica es pura, sino que todo existe entre culturas.

Por lo tanto, el objetivo general que guía el presente trabajo consiste en analizar el empleo de los patrones metadiscursivos en cartas al director, escritas por HN españoles y HNN griegos en contextos educativos. En cuanto a los objetivos específicos que nos hemos planteado alcanzar, estos se desglosan en lo siguiente: primero nos proponemos adaptar el modelo interpersonal de Hyland (2005; 2017) a las características del género carta al director y a las particularidades de la lengua española y, a continuación, identificar y clasificar los patrones metadiscursivos en los textos producidos por parte de los estudiantes novatos.

2. NOCIÓN Y TAXONOMÍAS DEL METADISCURSO

El término *metadiscurso* aparece por primera vez en 1958 en un artículo de Zellig Harris titulado «Linguistic transformation for information retrieval». El lingüista estadounidense (1970: 464-466), desde un enfoque estructuralista, emplea la expresión «núcleos metadiscursivos» (*metadiscourse kernels*) para indicar aquellas partes del discurso que hacen referencia al «material principal» (*main material*) –como en aquellos casos en los que el investigador realiza la discusión de los resultados y utiliza expresiones del tipo «hemos comprobado que...»– pero sin definirlo con más precisión o profundizar sobre sus distintos tipos y usos. A partir de la década de los ochenta, el concepto se retoma en las obras pioneras de teorizadores inscritos, en su mayoría, en el campo de la composición y retórica como Williams (1981; 1990; 2007), Vande Kopple (1985; 2002; 2012), Crismore (1985), Crismore y Farnsworth (1989) o Crismore, Markkanen y Steffensen (1993), quienes intentan definir y categorizar estos segmentos textuales de una forma más sistemática. De modo indicativo, Williams (1981: 211-212) califica el metadiscurso como «escribiendo sobre escribir, todo lo que no se refiere al tema tratado», Vande Kopple (1985: 83), al igual que Crismore y Farnsworth (1989: 92), hace referencia al «discurso acerca del discurso» (*discourse on discourse*), mientras que Crismore, Markkanen y Steffensen (1993: 40) lo definen como: «el material lingüístico en textos, orales o escritos, que no añade nada al contenido proposicional, sino que está destinado a ayudar al oyente o lector a organizar, interpretar y evaluar la información proporcionada».

El primer intento de cartografiar los marcadores metadiscursivos lo encontramos en el modelo de Vande Kopple quien, basándose en la distinción tripartita de las metafunciones del lenguaje propuesta por Halliday (1973; 1978; 1985), considera que el contenido proposicional o discurso primario corresponde a la metafunción ideacional, mientras que el no proposicional o metadiscursivo a las metafunciones interpersonal y textual (Vande Kopple, 1985: 86). Partiendo de esta idea, el autor propone una tipología

que se distingue en siete categorías: conectores textuales, glosas de código, marcadores ilocucionarios, marcadores de validez, narradores, marcadores de actitud y comentarios. De acuerdo con este teórico, los marcadores interpersonales facilitan a los autores revelar sus evaluaciones y actitudes en cuanto al contenido ideacional, a reseñar cuál es su papel en la situación comunicativa y a predecir la reacción de los lectores, mientras que los textuales les permiten obviar cómo se estructura la información proposicional del texto, de acuerdo con el acto comunicativo (2012: 38). Sin embargo, según advierte el propio autor, algunas palabras o grupos de palabras pueden cumplir más de una función metadiscursiva como, por ejemplo, en el caso de la expresión «mi hipótesis es» que puede actuar tanto como marcador ilocucionario, así como de validez (1985, p. 85). La categorización de Vande Kopple ha sido criticada por su ambigüedad en las categorías tratadas y sus frecuentes solapamientos funcionales, algo que según Hyland (2005: 32-33), dificulta su aplicación. Uno de los problemas que destacan es la dificultad en distinguir los narradores de los atribuidores, especialmente en el caso del discurso académico donde las citas cumplen intenciones retóricas variadas. No obstante, en un modelo posterior Vande Kopple (2002) excluye la categoría narradores de su tipología y el término marcadores de validez se substituye por el de marcadores epistemológicos (*epistemology markers*).

Crismore, Markkanen y Steffensen (1993), por su parte, en una revisión crítica de la primera categorización de Vande Kopple, eliminan la categoría narradores y algunas de sus sub-funciones se trasladan a una nueva categoría, la de los marcadores textuales, mientras que las glosas de código y los marcadores ilocucionarios forman parte de la categoría de los marcadores interpretativos. Las dos nuevas categorías creadas pretenden dar cuenta del papel textual del metadiscurso, con los marcadores textuales refiriéndose a aquellos elementos que facilitan la organización del discurso, frente a los interpretativos que ayudan al lector a descifrar y entender mejor lo que quiere comunicar el autor, así como las estrategias que emplea.

Ahora bien, las propuestas que más repercusión han tenido, sobre todo en las investigaciones relacionadas con el discurso académico, son las de Hyland (1998a; 1998b; 2005; 2013; 2015; 2017) quien ofrecerá un marco teórico más sólido sobre el metadiscurso como respuesta a la ambigüedad de los trabajos anteriores. Según comenta el propio autor, aunque la mayoría de los analistas adoptan un enfoque funcional, frecuentemente, malinterpretan la distinción tripartita de Halliday y confunden los criterios sintácticos con los funcionales (Hyland, 2005: 24). Como alternativa, Hyland (2017:20) presenta un modelo denominado interpersonal, en el que se suprimen las categorías formales incluidas en el modelo de Crismore y otros (1993), y se intentan controlar los solapamientos funcionales. Bajo la categoría interactiva, que permite al autor organizar su discurso teniendo en cuenta las necesidades del lector, se incluyen: (i) los marcadores de transición (*transition markers*), que expresan relaciones semánticas entre las oraciones principales del texto, (ii) los marcadores de marco (*frame markers*), que hacen referencia a actos discursivos, secuencias o etapas, (iii) los marcadores endofóricos (*endophoric markers*), que remiten a otras partes del texto, (iv) los evidenciales (*evidentials*), que remiten a otros textos y (v) las glosas de código (*code glosses*), que reformulan, ejemplifican o clarifican lo dicho. En la categoría interaccional, mediante la cual se expresa la voz del autor, su evaluación y grado de complicidad frente al contenido textual y a su audiencia, cuentan: (i) los atenuadores (*hedges*), que retienen el compromiso del autor frente a su proposición, los intensificadores (*boosters*), que enfatizan la certeza con la que se presenta el autor en el texto, (iii) los marcadores de actitud (*attitude markers*), que

reflejan la actitud afectiva del autor frente a su proposición, (iv) los marcadores de compromiso (*engagement markers*), que hacen referencias explícitas al lector o fomentan una relación entre autor-lector y (v) las autorreferencias (*self-mentions*), que se refieren de forma explícita al autor o autores.

3. MÉTODO

Teniendo en cuenta las investigaciones reseñadas con anterioridad, así como la naturaleza de nuestro estudio, la pregunta que se plantea en este punto es la siguiente:

- ¿Influye el nivel de competencia escritora (bajo, medio y avanzado) de los HNN y los HN en el uso de los marcadores metadiscursivos (interactivos e interaccionales)? Más concretamente, ¿existe diferencia significativa a nivel inter-grupal en la proporción y tipo de marcadores metadiscursivos empleados en los textos que consiguen una baja media o alta puntuación?

Las hipótesis formuladas a partir de la pregunta de investigación son las siguientes:

- H_0 : No habrá diferencia significativa en la cantidad y tipo de marcadores metadiscursivos empleados en los textos que consiguen una baja, media o alta puntuación.
- H_1 : Se espera una diferencia significativa en la cantidad y tipo de marcadores metadiscursivos empleados en los textos que consiguen una baja, media o alta puntuación.

3.1. Diseño

El estudio, de tipo transversal, se basó en un diseño descriptivo - correlacional para la identificación, categorización y comparación de los patrones metadiscursivos recurrentes en un texto de tipo argumentativo, perteneciente al género *carta al director*. Las muestras en las que se basó la investigación son de tipo intencional o de conveniencia provenientes de HN españoles y HNN que cursan educación secundaria y superior, respectivamente. Nuestra decisión de comparar informantes novatos que pertenecen a distintos niveles educativos se basó en la afinidad temática y los objetivos comunes de la asignatura de lengua que cursan ambos grupos

Con el fin de comparar los datos obtenidos a nivel inter-grupal, se tuvo que medir el nivel de la competencia comunicativa de los sujetos en relación con una actividad de expresión e interacción escritas común para ambos grupos. El objetivo de la actividad consiste en que los alumnos produzcan una *carta al director*, de entre 200 a 250 palabras, basándose en un texto sacado del periódico *El País*. Los escritos producidos por los HNN se evaluaron de acuerdo con la escala graduada, del examen IELTS (International English Language Testing System) Formación General, diseñada para medir la competencia comunicativa, a través de actividades que reflejan situaciones laborales y sociales, en un rango de 9 a 0 puntos, es decir de mayor a menor competencia comunicativa. Por otro lado, la evaluación de la expresión escrita de los HN se realizó mediante la parrilla analítica del Instituto Vasco de Evaluación e Investigación Educativa (ISEI-IVEI, 2013: 21), destinada a la valoración diagnóstica de la competencia en comunicación lingüística

de alumnos de secundaria a la hora de escribir un texto argumentativo. Con el fin de validar los instrumentos de medición se aplicó un Protocolo de Calibración basado en una línea de investigaciones realizadas con anterioridad en este campo (Muñoz, et al. 2006; Shaw & Falvey 2008).

Para la identificación de los ítems metadiscursivos se empleó el programa de concordancias AntConc (versión 3.4.4.) y las variables se analizaron mediante los programas estadísticos SPSS (versión 22), Excel y el calculador del cociente de verosimilitud G^2 (*log-likelihood ratio*) del Departamento Computacional de la Universidad de Lancaster.

3.2. Participantes

En la investigación participaron ciento veinticinco informantes en total (27.969 palabras), de los cuales setenta y ocho pertenecen al grupo griego (16.275 palabras), mientras que los cuarenta y siete restantes forman parte del grupo español (11.694 palabras).

Los informantes griegos que participaron en el estudio están matriculados en la asignatura *Lengua III* que forma parte del bloque curricular obligatorio del Departamento de Lengua y Literaturas Hispánicas de la Universidad Nacional y Kapodistriaca de Atenas. Cincuenta y seis de ellos provienen del año académico 2012-2013 y veintidós del año académico 2014-2015. La muestra griega cuenta con una representación femenina mayoritaria de 84,6% (66 frecuencias) frente a la masculina que alcanza el 15,4% (12 frecuencias). La edad media del grupo se sitúa en los 29,62 años con una desviación estándar de 9,792 y una variación de los 18 a los 58 años. La categoría más numerosa es la que comprende las edades entre 17 y 22 años (24), sigue la que abarca las edades entre los 29 y 34 (19), en tercera posición está la de los 23 a los 28 años (15) y en cuarta posición, con una igual distribución se encuentra, por un lado, la categoría que incluye a las personas que están entre 35 y 40 años (10) y, por otro, la de las personas que tienen más de 40.

Los HN son alumnos preuniversitarios hispanohablantes que cursan segundo de Bachillerato, Modalidad de Ciencias y Tecnología, en la autonomía de Murcia. Treinta y seis de ellos provienen del IES Mediterráneo de Cartagena de dos años escolares consecutivos, a saber, de 2012- 2013 y de 2013-2014, mientras que los once restantes están cursando en el IES San Juan Bosco de Lorca durante el año escolar 2014-2015. La recolección de los textos se realizó en el período entre abril y mayo, época durante la cual los alumnos estaban finalizando sus estudios. La distribución de la muestra en función del sexo es bastante homogénea con el 55,3% (26) siendo varones y el 44,7% (21) restante mujeres. En lo relacionado a la edad, treinta de ellos tienen 17 años, dieciséis tienen 18 años y solo una alumna ha cumplido los 20 años.

3.3. Variables

Para cumplir con los objetivos planteados en el presente estudio se han tomado en cuenta tres variables, esto es, una dependiente y dos independientes. Como variable dependiente se ha considerado la frecuencia porcentual de los marcadores metadiscursivos. Para su operativización se ha adaptado la taxonomía de Hyland (2017) a las características de la carta al director y también a las particularidades de la lengua española (Salapata 2017). Como se puede ver en la figura 1, la sub-categoría de los marcadores de transición, así

como la de los marcadores de marco de la categoría interactiva se han adaptado al castellano, según las categorizaciones de Calsamiglia y Tusón (2012).

Items metadiscursivos interactivos		
Categoría	Sub-categoría	Ejemplo
Interactiva		
Marcadores de transición	Aditivos, contrastivos (oposición, sustitución, restricción, concesión), de base causal	además, encima, y/pero, sin embargo, sino, excepto si, a pesar de/por eso, de ahí que...
Marcadores de marco	Ilocucionarios, iniciadores, distribuidores, ordenadores, continuativos, digresivos, recapitulativos-conclusivos	el objetivo de esta carta es/para empezar/por un lado/primero, en segundo lugar/entonces/por cierto/en conclusión...
Marcadores endofóricos	Anáfora y catafora	lo anteriormente mencionado /véase más abajo...
Evidenciales	Citas directas, indirectas	(fecha/nombre) /según la autora del texto...
Glosas de código	Reformuladores	a saber, p.ej., en otras palabras, paréntesis, guiones, dos puntos...

Figura 1. Variables metadiscursivas interactivas

Asimismo, en lo que atañe a la categoría interaccional, hemos introducido una nueva categoría bajo los marcadores de compromiso (fig. 2), a saber, las estructuras de cortesía que aparecen en la sección de apertura y cierre de una carta.

Items metadiscursivos interaccionales		
Interaccional		
Atenuadores	Verbos de juicio, verbos modales, adverbios epistémicos y locuciones adverbiales, estructuras adjetivales, adjetivos, diminutivos	indicar, calcular, dudar/poder, parece que, deber de/probablemente/es probable, -ito, illo, ico...
Intensificadores	Marcadores de certeza, verbos cognitivos	está claro, sin duda/saber, conocer...
Marcadores de actitud	Formas adverbiales, adjetivos y formas adjetivales, verbos de actitud	lamentablemente/ bueno, malo, es+adjetivo calificativo/coincidir, estar de acuerdo...
Autorreferencias	Expresiones de opinión personal, Formas tónicas o atonas de los pronombres	A nivel personal / yo, mi, a mí, me, nosotros, nuestro, nos...

	personales, posesivos, flexión verbal en 1ª persona, nosotros exclusivo	
Marcadores de compromiso	Pronombres referidos al lector, directivos (imperativos, verbos modales deónticos, verbo copulativo+sintagma adjetival o adjetivo), preguntas, conocimiento compartido, apartes, estructuras de cortesía	tú, usted, vosotros/imagina, deber+infinitivo/es necesario/es bien sabido/ Estimado director, atentamente

Figura 2. Variables metadiscursivas interaccionales

Asimismo, todos los rasgos lingüísticos se normalizaron a partir de la propuesta de Biber (1988:75-76), es decir, número de rasgos lingüísticos dividido por la extensión de cada subcorpus y multiplicado por 100.

Las variables independientes que serán contrastadas con la frecuencia porcentual de los marcadores metadiscursivos son, por un lado, la lengua materna de los informantes, es decir el español (L1), así como el español como lengua extranjera (L2) en el caso de los aprendientes griegos. Por otro lado, el nivel de competencia escritora, medido a través de las dos escalas de evaluación antes mencionadas constituye la segunda variable independiente que interviene en el presente estudio.

4. RESULTADOS

Bajo este epígrafe se presentan los resultados de la comparación de los recursos metadiscursivos entre los estudiantes griegos universitarios y los españoles preuniversitarios. A través del análisis contrastivo (tabla 1), se pudo comprobar que existen diferencias significativas a nivel global y por categoría metadiscursiva. En concreto, el número total de marcadores metadiscursivos del subcorpus griego (2644 ocurrencias) frente a aquel del español (1695 ocurrencias) hace patente una diferencia significativa de $G^2= 14,49$ ($p < 0,001$) a favor de los HNN. De igual modo, se observan diferencias significativas sobre el total de los ítems interactivos ($G^2= 7,28$, $p < 0,01$) a favor de los HN y de los interaccionales sobreutilizados por los HNN ($G^2=40,93$, $p < 0,0001$). Un análisis pormenorizado de los datos por categoría metadiscursiva evidencia diferencias significativas en dos de las cinco sub-categorías interactivas, es decir, en los marcadores de transición que presentan unas frecuencias más bajas en los textos de los HNN ($G^2=13,13$, $p < 0.001$) y en los marcadores de marco que son más numerosos en los textos de estos últimos ($G^2=8,59$, $p < 0.01$). Por lo que se refiere al metadiscurso interaccional, el análisis del cociente de verosimilitud revela diferencias significativas a favor de los HNN en los marcadores de actitud ($G^2=23,35$, $p < 0,0001$), mientras que en los marcadores de compromiso ($G^2=8,45$) la significación es de 0,01.

Marcadores metadiscursivos					
Categoría	HNN		HN		G ²
	01	%1	02	%2	
Interactiva					
Marcadores de transición	514	3,16	466	3,98 -	13,13
Marcadores de marco	120	0,74	54	0,46 +	8,59
Endofóricos	3	0,02	5	0,04 -	1,38
Evidenciales	69	0,42	60	0,51 -	1,16
Glosas de código	66	0,41	56	0,48 -	0,83
Sub-total	772	4,74	641	5,48 -	7,28
Interaccional					
Atenuadores	69	0,42	65	0,56 -	2,44
Intensificadores	108	0,66	68	0,58 +	0,73
Marcadores de actitud	196	1,20	75	0,64 +	23,35
Autorreferencias	406	2,49	273	2,33 +	0,72
Marcadores de compromiso	1093	6,72	682	5,83 +	8,45
Sub-total	1872	11,50	1054	9,01 +	40,93

Tabla 1. Resultados globales de G² entre HNN y HN

A continuación, se ofrece un análisis contrastivo por pares, entre HNN de bajo, medio o nivel alto y HN a fin de descubrir en qué categorías y sub-categorías se asemejan o se diferencian sus producciones.

4.1. Resultados entre HNN de nivel bajo y HN

Si se comparan las producciones escritas de los HN con las de los HNN de nivel bajo (tabla 2), se nota que las categorías en las que la significación estadística es fuerte ($p < 0,0001$) a favor de los estudiantes griegos son los marcadores de compromiso ($G^2=27,40$), los marcadores de actitud ($G^2=26,34$), las autorreferencias ($G^2=38,47$), así como los marcadores de marco ($G^2=8,46$) con una significación de $p < 0,01$.

Marcadores metadiscursivos					
Categoría	HNN nivel bajo		HN		G ²
	01	%1	02	%2	
Interactiva					
Marcadores de transición	187	3,39	466	3,98 -	3,51
Marcadores de marco	46	0,83	54	0,46 +	8,46
Endofóricos	1	0,02	5	0,04 -	0,73
Evidenciales	20	0,36	60	0,51 -	1,90
Glosas de código	25	0,45	56	0,48 -	0,05
Sub-total	279	5,06	641	5,48 -	1,23
Interaccional					
Atenuadores	28	0,51	65	0,56 -	0,16
Intensificadores	34	0,62	68	0,58 +	0,08
Marcadores de actitud	81	1,47	75	0,64 +	26,34
Autorreferencias	156	2,83	164	1,40 +	38,47
Marcadores de compromiso	444	8,06	682	5,83 +	27,40
Sub-total	743	13,49	1054	9,01 +	68,74

Tabla 2. Resultados globales de G² entre HNN de nivel bajo y HN

En un análisis más detallado, como se puede observar en la tabla 3, se nota que los marcadores de marco en los que se detecta una diferencia significativa son, por un lado, los ilocucionarios ($G^2=27,44$, $p < 0,0001$) y, por otro, los ordenadores ($G^2=4,52$, $p < 0,05$).

Subtipo	Marcadores de Marco				G^2
	HNN nivel bajo		HN		
	O1	%1	O2	%2	
Ilocucionarios	23	0,42	6	0,05 +	27,44
Iniciadores	2	0,04	11	0,09 -	1,88
Distribuidores	6	0,11	10	0,09 +	0,21
Ordenadores	6	0,11	3	0,03 +	4,52
Continuativos	1	0,02	2	0,02 +	0,00
Recapitulativos- conclusivos	8	0,15	22	0,19 -	0,41

Tabla 3. Marcadores de marco: resultados de G^2 entre HNN de nivel bajo y HN

En cuanto a los marcadores de actitud, las diferencias entre los dos grupos se detectan en el empleo de los adjetivos, ($G^2=20,35$, $p < 0,0001$) y en los exclamativos (tabla 4), ambos sobreutilizados ($G^2=27,33$, $p < 0,0001$) en los textos de los informantes griegos.

Tipo	Marcadores de actitud				G^2
	HNN nivel bajo		HN		
	O1	%1	O2	%2	
Adverbios	4	0,07	6	0,05 +	0,28
Adjetivos	47	0,85	37	0,32 +	20,35
Verbos de actitud	17	0,31	32	0,27 +	0,16
Exclamativos	12	0,22	0	0,00+	27,33

Tabla 4. Marcadores de actitud: resultados de G^2 entre HNN de nivel bajo y HN

Tal y como se presenta en la tabla 5, el empleo de las autorreferencias difiere en el pronombre personal *yo*, escasamente utilizado por el grupo de los HNN ($G^2=4,49$, $p < 0,05$), los determinativos posesivos ($G^2=11,53$, $p < 0,001$) y los verbos en primera persona del singular ($G^2= 17,07$, $p < 0,0001$), sobreutilizados por estos últimos, así como la sección de la firma ($G^2=40,74$, $p < 0,0001$) que tiene una representación muy restringida en las cartas de los alumnos preuniversitarios.

Tipo	Autorreferencias				
	HNN nivel bajo		HN		G ²
	O1	%1	O2	%2	
Pronombres Personales					
Formas tónicas					
Yo	1	0,02	12	0,10 -	4,49
Mí	1	0,02	4	0,03 -	0,36
Formas átonas					
Me	11	0,20	18	0,15 +	0,45
Nos (exclusivo)	0	0,00	1	0,01 -	0,77
Determinativos posesivos					
Mi(s), mío, a(s)	37	0,67	35	0,30 +	11,53
Flexión verbal					
Flexión verbal 1ª pers. sing.	78	1,42	86	0,74 +	17,07
Flexión verbal 1ª pers. pl.	0	0,00	2	0,02 -	1,54
Expresiones de opinión	1	0,02	2	0,02 +	0,00
Firma	27	0,49	4	0,03 +	40,74

Tabla 5. Autorreferencias: resultados de G² entre HNN de nivel bajo y HN

Por último, en la tabla 6 se puede observar que las diferencias en la categoría de los marcadores de compromiso se centran en las referencias al lector (G²=10,62, p < 0,01), los argumentos que apelan al conocimiento compartido (G²=4,98 p < 0,05) y las estructuras de cortesía (G²=65,90, p < 0,0001).

Tipo	Marcadores de compromiso				
	HNN nivel bajo		HN		G ²
	O1	%1	O2	%2	
Referencias al lector	314	5,70	527	4,51 +	10,62
Directivos	47	0,85	96	0,82 +	0,05
Preguntas	9	0,16	24	0,21 -	0,35
Conocimiento compartido	18	0,33	18	0,15 +	4,98
Apartes	0	0,00	2	0,02 -	1,54
Estructuras de cortesía	56	1,02	15	0,13 +	65,90

Tabla 6. Marcadores de compromiso: resultados de G² entre HNN de nivel bajo y HN

Si se mira de forma más detenida en las referencias al lector (tabla 7), se sugiere que en los textos de los HN tanto el empleo de la forma tónica del pronombre personal *nosotros*, como la átona *nos* muestran una diferencia significativa frente a los de los HNN (G²= 10,81, p < 0,01 y G²= 17,55, p < 0,0001, respectivamente). Por otro lado, la forma tónica *usted* y la átona *le* para referirse al director del periódico o la autora del texto tienen una presencia fuertemente relevante en los textos de los HNN (G²= 25,31, p < 0,0001 y G²= 12,82, p < 0,001, respectivamente), al igual que el determinativo posesivo *su* (G²= 25,10, p < 0,0001). De igual modo, los estudiantes griegos hacen uso extenso de los verbos en segunda persona del singular y del plural para dirigirse abiertamente al lector (G²= 19,37, p < 0,0001 y G²= 5,74, p < 0,05 respectivamente). Por último, los HN

emplean con más frecuencia la primera persona del plural para implicar a su audiencia ($G^2= 6,07$, $p < 0,05$).

Subtipo	Referencias al lector				G^2
	HNN nivel bajo		HN		
	O1	%1	O2	%2	
Pronombres Personales					
Formas tónicas					
Tú	0	0,00	1	0,01 -	0,77
Nosotros	0	0,00	14	0,12 -	10,81
Usted(es)	14	0,25	1	0,01 +	25,31
Formas átonas					
Te	3	0,05	12	0,10 -	1,08
Nos	31	0,56	142	1,21 -	17,55
Le(s)	11	0,20	3	0,03 +	12,82
Determinativos posesivos					
Tu(s)	1	0,02	2	0,02 +	0,00
Su(s)	19	0,34	4	0,03 +	25,10
Nuestro	56	1,02	58	0,50 +	14,31
Flexión verbal					
Flexión verbal 2ª pers. sing.	13	0,24	2	0,02 +	19,37
Flexión verbal 1ª pers. pl.	155	2,81	274	2,34 +	3,26
Flexión verbal 2ª pers. pl.	1	0,02	2	0,02 +	0,00

Tabla 7. Referencias al lector: resultados de G^2 entre HNN de nivel bajo y HN

Los argumentos que apelan al conocimiento común son más numerosos en los textos de los HNN de nivel bajo (tabla 8). La mayor diferencia se detecta en las falacias *ad populum*, sobreutilizadas por los estudiantes griegos de nivel bajo y, especialmente la forma *todos* ($G^2=14,14$, $p < 0,001$).

Subtipo	Argumentos <i>ad populum</i>				G^2
	HNN nivel bajo		HN		
	O1	%1	O2	%2	
La mayoría	4	0,07	6	0,05 +	0,28
Todos	13	0,24	4	0,03 +	14,14

Tabla 8. Argumentos *ad populum*: resultados de G^2 entre HNN de nivel bajo y HN

Por último, las estructuras de cortesía, como se puede también observar en la tabla 9, hacen patente una diferencia significativa en casi todas las secciones de la carta y, sobre todo, en la línea de saludo.

Subtipo	Estructuras de cortesía				G^2
	HNN nivel bajo		HN		
	O1	%1	O2	%2	
Línea de saludo	27	0,49	8	0,07 +	30,04
Conclusión	11	0,20	3	0,03 +	12,82
Despedida	17	0,31	4	0,03 +	21,35
Sección de posdatos	1	0,02	0	0,00 +	2,28

Tabla 9. Estructuras de cortesía: resultados de G^2 entre HNN de nivel bajo y HN

4.2. Resultados de HNN de nivel medio y HN

El análisis contrastivo de los HN con los HNN de nivel medio permite observar diferencias en la mitad de las categorías metadiscursivas (tabla 10), a saber, los marcadores de transición, los de marco, los endofóricos, los marcadores de actitud y las autorreferencias. El cociente de verosimilitud sobre el total de los ítems interactivos alcanza el $G^2= 8,70$ ($p < 0,01$) a favor de los HN, mientras que en el caso del conjunto de los interaccionales es de $G^2=11,76$ ($p < 0,001$) en pro de los HNN.

Categoría	Marcadores metadiscursivos				G^2
	HNN nivel medio		HN		
	01	%1	02	%2	
Interactiva					
Marcadores de transición	239	2,96	466	3,98 -	14,28
Marcadores de marco	58	0,72	54	0,46 +	5,47
Endofóricos	0	0,00	5	0,04 -	5,25
Evidenciales	38	0,47	60	0,51 -	0,17
Glosas de código	30	0,37	56	0,48 -	1,28
Sub-total	365	4,52	641	5,48 -	8,70
Interaccional					
Atenuadores	33	0,41	65	0,56 -	2,12
Intensificadores	58	0,72	68	0,58 +	1,40
Marcadores de actitud	78	0,97	75	0,64 +	6,42
Autorreferencias	182	2,26	164	1,40 +	19,48
Marcadores de compromiso	501	6,21	682	5,83 +	1,14
Sub-total	852	10,56	1054	9,01 +	11,76

Tabla 10. Resultados globales de G^2 entre HNN de nivel medio y HN

Si se indaga más en las sub-categorías en las que se nota una diferencia significativa se comprueba que en lo referido a los marcadores de transición los HNN de medio rendimiento infrutilizan los marcadores de base causal ($G^2=11,84$, $p < 0,001$) y los contraargumentativos ($G^2=6,73$, $p < 0,01$), incluidos casi todos los subtipos que les corresponden, salvo los marcadores de sustitución, como se puede observar también de forma resumida en la tabla 11.

Tipo	Marcadores de transición				G^2
	HNN nivel medio		HN		
	01	%1	02	%2	
Aditivos	143	1,77	237	2,03 -	1,62
Contraargumentativos	57	0,71	124	1,06 -	6,73
<i>Oposición</i>	40	0,50	86	0,74 -	4,43
<i>Restricción</i>	1	0,01	9	0,08 -	4,73
<i>Sustitución</i>	15	0,19	18	0,15 +	0,29
<i>Concesión</i>	1	0,01	11	0,09 -	6,45
De base causal	39	0,48	105	0,90 -	11,84
<i>Causativos</i>	29	0,36	76	0,65 -	7,95
<i>Consecutivos</i>	10	0,12	29	0,25 -	3,94

Tabla 11. Marcadores de transición: resultados de G^2 entre HNN de nivel medio y HN

En cuanto a los marcadores de marco (tabla 12), los dos grupos divergen en el empleo de los ilocucionarios, sobreutilizados por los HNN ($G^2=19,28$, $p < 0,0001$), los iniciadores que tienen una presencia nula en sus cartas en comparación con las de los HN ($G^2=11,54$, $p < 0,001$) y, por último, con una significación de 0,05 los ordenadores, que son escasos en los escritos de estos últimos ($G^2= 4,59$).

Marcadores de marco					
Tipo	HNN nivel medio		HN		G²
	01	%1	02	%2	
Ilocucionarios	24	0,30	6	0,05 +	19,28
Iniciadores	0	0,00	11	0,09 -	11,54
Distribuidores	5	0,06	10	0,09 -	0,36
Ordenadores	8	0,10	3	0,03 +	4,59
Continuativos	3	0,04	2	0,02 +	0,74
Recapitulativos-Conclusivos	18	0,22	22	0,19 +	0,20

Tabla 12. Marcadores de marco: resultados de G^2 entre HNN de nivel medio y HN

En el caso de los marcadores de actitud si se desagregan los datos por tipo metadiscursivo se percatan diferencias relevantes en el empleo de los exclamativos, tipo metadiscursivo que tiene una nula presencia en el corpus de los HN españoles ($G^2= 5,38$, $p < 0,05$) como se puede también observar en la siguiente tabla (tabla 13).

Marcadores de actitud					
Subtipos	HNN nivel medio		HN		G²
	01	%1	02	%2	
Adverbios	8	0,10	6	0,05 +	1,51
Adjetivos	32	0,40	37	0,32 +	0,87
Verbos de actitud	35	0,43	32	0,27 +	3,54
Exclamativos	3	0,04	0	0,01 +	5,38

Tabla 13. Marcadores de actitud: resultados de G^2 entre HNN de nivel medio y HN

Por último, en la categoría de las autorreferencias los rasgos lingüísticos en los que se nota una diferencia significativa, tal y como se puede ver en la tabla 14, son el pronombre *yo*, menos utilizado por los aprendientes griegos de nivel medio ($G^2= 7,33$, $p < 0,01$), los verbos en primera persona del singular ($G^2= 5,08$, $p < 0,05$), así como la sección de la firma que es más frecuente en las cartas de estos últimos ($G^2= 28,74$, $p < 0,0001$).

Autorreferencias					
Subtipo	HNN nivel medio		HN		G²
	01	%1	02	%2	
Pronombres Personales Formas tónicas					
Yo	1	0,01	12	0,10 -	7,33
Mí	1	0,01	4	0,03 -	0,98
Formas átonas					
Me	19	0,24	18	0,15 +	1,67
Nos (exclusivo)	0	0,00	1	0,01 -	1,05

**Determinativos
posesivos**

Mi(s), mía,a(s)	35	0,43	35	0,30 +	2,40
Nuestro	1	0,01	0	0,00 +	1,79

Flexión verbal

Flexión verbal 1ª pers. sing	79	0,98	86	0,74 +	3,35
Flexión verbal 1ª pers. pl.	0	0,00	2	0,02 -	2,10
Expresiones de opinión	0	0,00	2	0,02 -	2,10
Firma	27	0,33	4	0,03 +	28,74

Tabla 14. Autorreferencias: resultados de G^2 entre HNN de nivel medio y HN

4.3. Resultados de HNN de nivel alto y HN

La comparación entre los HN y los HNN de nivel alto permite constatar más similitudes que diferencias, como se puede también observar en la tabla 15. En particular, a nivel global el empleo de los marcadores interactivos e interaccionales no presenta diferencias estadísticamente significativas. Las únicas sub-categorías de índole interaccional en las que se registra un empleo más alto por parte de los HNN son los marcadores de actitud ($G^2=12,90$, $p < 0,001$) y las autorreferencias ($G^2=15,08$, $p < 0,001$).

Categoría	Marcadores metadiscursivos				G^2
	HNN nivel alto		HN		
	01	%1	02	%2	
Interactiva					
Marcadores de transición	88	3,26	466	3,98 -	3,12
Marcadores de marco	16	0,59	54	0,46 +	0,74
Endofóricos	2	0,07	5	0,04 +	0,40
Evidenciales	11	0,41	60	0,51 -	0,52
Glosas de código	11	0,41	56	0,48 -	0,25
Sub-total	128	4,74	641	5,48-	2,31
Interaccional					
Atenuadores	8	0,30	65	0,56 -	3,31
Intensificadores	16	0,59	68	0,58 +	0,00
Marcadores de actitud	37	1,37	75	0,64 +	12,90
Autorreferencias	68	2,52	164	1,40 +	15,08
Marcadores de compromiso	148	5,48	682	5,83 -	0,43
Sub-total	277	10,26	1054	9,01+	3,61

Tabla 15. Resultados globales de G^2 entre HNN de nivel alto y HN

En lo referido a los marcadores de actitud, los adjetivos constituyen el único elemento lingüístico que tiene una presencia más alta en los textos de los HNN griegos ($G^2=18,60$, $p < 0,0001$), tal y como se puede constatar en la tabla 16.

Tipo	Marcadores de actitud				
	HNN nivel alto		HN		G ²
	01	%1	02	%2	
Adverbios	4	0,15	6	0,05 +	2,42
Adjetivos	27	1,00	37	0,32 +	18,60
Verbos de actitud	5	0,19	32	0,27 -	0,72
Exclamativos	1	0,04	0	0,01 +	3,35

Tabla 16. Marcadores de actitud: resultados de G² entre HNN de nivel alto y HN

La segunda categoría interaccional en la que se detectaron diferencias significativas son las autorreferencias. Como se puede ver en la tabla 17, las formas lingüísticas en las que se advierte una diferencia significativa son, por un lado, el pronombre personal *mí*, que es más frecuente en los textos de los HNN de nivel alto (G²= 3,96, p < 0,05) y, por otro, la sección de la firma que es escasa en las cartas redactadas por los HN (G²=18,39, p < 0,0001).

Tipo	Autorreferencias				
	HNN nivel alto		HN		G ²
	01	%1	02	%2	
Pronombres personales					
Formas tónicas					
Yo	1	0,04	12	0,10 -	1,28
Mí	4	0,15	4	0,03 +	3,96
Formas átonas					
Me	9	0,33	18	0,15 +	3,23
Nos (exclusivo)	0	0,00	1	0,01 -	0,42
Determinativos posesivos					
Mi(s), mía(s)	14	0,52	35	0,30 +	2,77
Flexión verbal					
Flexión verbal 1ª pers. sing	28	1,04	86	0,74 +	2,35
Flexión verbal 1ª pers. pl.	0	0,00	2	0,02 -	0,83
Expresiones de opinión	2	0,07	2	0,02 +	1,98
Firma	10	0,37	4	0,03 +	18,39

Tabla 17. Autorreferencias: resultados de G² entre HNN de nivel alto y HN

Resumiendo, se podría sugerir que el nivel en el que se presenta la mayor variabilidad entre las producciones escritas de los HNN frente a las de los HN, es el medio, sigue el bajo, mientras que el alto presenta más similitudes que diferencias en el empleo de las estrategias metadisucursivas.

5. DISCUSIÓN DE LOS RESULTADOS

Los resultados obtenidos, confirman parcialmente la hipótesis nula presentada en la sección metodológica. En concreto, el análisis contrastivo se centró en un primer

momento en el conjunto de los HNN frente a los HN. Los resultados obtenidos revelaron diferencias significativas a nivel global y por sub-categoría tratada. En la categoría interactiva, la sub-categoría en la que se evidenciaron diferencias significativas fueron los marcadores de transición, seguidos por los de marco, mientras que en las demás categorías no se detectaron diferencias relevantes. En la categoría interaccional se mostró una diferencia altamente relevante a nivel global, así como en la mayoría de las sub-categorías que la componen, con excepción de los atenuadores y los intensificadores. Por orden jerárquico, las categorías en las que se notó un sobreuso significativo por parte de los aprendientes griegos fueron las autorreferencias, seguidas por los marcadores de actitud y los de compromiso. En un análisis más profundo por pares se pudo comprobar que en los textos de los HNN de nivel bajo se infrutilizaban los marcadores de transición, mientras que en los de los aprendientes avanzados los endofóricos. En cuanto a los marcadores de marco se observó una sobreutilización por los aprendientes de nivel bajo y medio frente a los alumnos españoles. En lo referido a la categoría interaccional, se pudo evidenciar el sobreuso de los marcadores de actitud y las autorreferencias en los tres niveles de competencia escritora, mientras que los marcadores de compromiso presentaron una diferencia altamente significativa solo en las cartas que se calificaron de bajo rendimiento. Al comparar estos resultados con estudios previos, en su mayoría enfocados en géneros propios del ámbito académico, se observan coincidencias y divergencias, tal y como se verá más adelante.

Un análisis pormenorizado por nivel de competencia escritora en la categoría interactiva reveló que los HNN de nivel medio infrutilizaban los marcadores de transición frente a los HN. Por otro lado, los marcadores de marco tenían una presencia más alta en las cartas de los estudiantes universitarios griegos de nivel bajo y medio en comparación con los alumnos preuniversitarios. Estas conclusiones están en desacuerdo con los resultados obtenidos por Neff-van Aertselaer y Dafouz (2008), quienes descubrieron que los estudiantes novatos anglófonos y los hispanófonos en L2 inglés hacían un uso parecido de los conectores lógicos y los secuenciadores en el género ensayo argumentativo. En cuanto a la frecuencia más alta de los endofóricos en los textos de los HN, esta podría ser justificada por el hecho de que muchos de los alumnos preuniversitarios responden a la consigna como si fuera un comentario de texto, género en el que se suelen usar mecanismos deícticos para guiar al lector.

En cuanto a la categoría de los atenuadores e intensificadores que, según Fuentes Rodríguez (2016: 163), constituyen «dos polos en un contínuum que el hablante maneja para su fin persuasivo», es decir, para mitigar una eventual crítica por parte del lector o imponer su propia propuesta frente a alternativas, aunque no se observaron diferencias significativas entre los HNN y los HN a nivel global, sí que existieron variaciones en los tipos y sub-tipos que los componen. Por un lado, la atenuación, que en el presente trabajo se ha distinguido en los verbos de juicio, los modales, las expresiones epistémicas y los diminutivos, presentaron una distribución variada entre los dos corpus. En particular, los verbos de juicio junto con las expresiones modales presentaron una frecuencia más alta en las cartas redactadas por los españoles frente a las redactadas por los griegos, quienes hacían un uso más elevado de los verbos modales.

Por otro lado, la intensificación que, según resalta Albelda Marco (2005: 96), opera en distintos niveles de la lengua de acuerdo con la finalidad hacia la que se orienta, en el presente estudio comprende cuatro tipos, esto es, verbos cognitivos, adjetivos, expresiones de certeza y adverbios. Como se pudo comprobar, mientras que la frecuencia de las expresiones de certeza fue bastante parecida entre los dos grupos (0,28% los HNN

y 0,26% los HN), en los demás tipos se notó un empleo bastante variado, con los españoles utilizando más del doble de adverbios (0,17%) que los griegos (0,08%) y estos últimos más del doble de adjetivos (0,19% los HNN frente a 0,06% los HN) y casi el doble de verbos cognitivos (0,13% los HNN frente a 0,07% los HN). No obstante, incluso en las expresiones epistémicas en que se evidenciaron frecuencias parecidas entre los dos grupos, el análisis cualitativo reveló diferencias en los tipos empleados. En este sentido, los griegos mostraron una preferencia por las estructuras *es verdad que* (0,08%), *está claro que* (0,07%), frente a los españoles que emplearon con más frecuencia las formas *es cierto que* (0,11%) y *de verdad* (0,05%).

Estos hallazgos contradicen los resultados obtenidos por estudios contrastivos previos entre L1 y L2 sobre géneros propios del discurso académico y, en particular, sobre el ensayo argumentativo. En este sentido, Hyland y Milton (1997) al realizar un análisis contrastivo entre HNN chinos y HN británicos a partir de ensayos provenientes de la prueba del bachillerato internacional encontraron diferencias significativas entre los dos grupos. Según los resultados obtenidos, los HNN chinos utilizaban con más frecuencia los marcadores de certeza frente a los HN cuyos ensayos comprendían un alto porcentaje de marcadores de probabilidad. En la misma línea, Hinkel (2005) realizó un análisis contrastivo entre HNN de diversa procedencia (chinos, japoneses, coreanos, indonesios, vietnamitas y árabes) y HN anglosajones, todos novatos. La investigadora llegó a la conclusión, por un lado, que los ensayos de los HNN se caracterizaban por un tono coloquial y, por otro, que prevalecían los intensificadores frente a los atenuadores. En el ámbito griego, Hatzitheodorou y Mattheoudakis (2011) partiendo de un análisis contrastivo en inglés L1 y L2 observaron que la persuasión en el grupo griego se logró principalmente mediante los intensificadores, siendo el marcador *of course* el más común entre todos, mientras que los grupos de los nativos –estudiantes británicos y estadounidenses– mostraban una nítida preferencia por los atenuadores y, en especial, por los ítems léxicos *probably* y *possibly*. Ädel (2006: 193), por su parte, tras analizar el género ensayo argumentativo encontró que los HNN suecos exhibían un nivel alto de lo que se denomina como «modestia retórica» (*rhetoric of modesty*), puesto que recurrían más a las estrategias atenuadoras en comparación con los estudiantes británicos y estadounidenses.

En lo referente a los marcadores de actitud, estos revelaron diferencias en todos los niveles de competencia escritora de los HNN frente a los alumnos españoles. Una vez desagregados los datos, se pudo comprobar que en el nivel medio no hubo diferencias en ningún tipo actitudinal, mientras que en el nivel bajo la diferencia más alta se detectó en los exclamativos, seguidos por los adjetivos. En el nivel alto las diferencias se observaron solo en las formas adjetivales. Si se comparan estos resultados con los de Hatzitheodorou y Mattheoudakis (2011) con HN anglófonos y HNN griegos en L2 inglés, se observa que hay diferencias en los resultados obtenidos. En concreto, estas investigadoras llegaron a la conclusión que los marcadores actitudinales eran escasos en ambos grupos con excepción del adverbio *unfortunatley*, sobre todo, en el corpus GRICLE en que aparecía con más frecuencia conjuntamente con el nexos adversativo *but*.

En cuanto a las autorreferencias, como ya se ha comentado anteriormente, las diferencias se observaron en todos los niveles. En este contexto, en el nivel bajo se notó la infrautilización de las formas *nosotros* con valor exclusivo, así como del pronombre personal *yo* por parte de los HNN, mostrando una preferencia por los verbos en primera persona del singular. De igual modo, sus cartas iban en su mayoría firmadas frente a las de los HN que eran anónimas. En cuanto a los textos situados en el nivel medio, estos

presentaron un uso más elevado de los verbos en primera persona del singular y de expresiones de opinión, frente a los de los nativos en los que se registró una frecuencia más relevante del pronombre personal *yo* al igual que en la comparación con los aprendientes de nivel bajo. En el nivel alto, las diferencias se centraron en la sección de la firma y en el pronombre personal *mí*, ambos más frecuentes en los textos de los HNN. Dado que no existen trabajos que comparen el empleo de los marcadores metadiscursivos entre escritores novatos y por nivel de competencia, nuestros resultados no pueden ser contrastados con los obtenidos por estudios previos. Asimismo, mediante la revisión de la literatura relevante, se pudo comprobar que son muy pocos los trabajos que hacen un análisis de todas las categorías del modelo de Hyland, centrándose en la mayoría de los casos en la intensificación y la atenuación. Aparte de ello, en muchos trabajos no se hace una distinción entre el *nosotros* con valor inclusivo y exclusivo, llevando a resultados contradictorios incluso sobre el mismo género, como se verá más adelante.

Por último, el empleo de los marcadores de compromiso mostró variaciones solo en el análisis contrastivo de los HNN de nivel bajo. Los tipos compromisivos en los que se observaron variaciones entre los dos subcorpus fueron las referencias al lector, los argumentos que apelan al conocimiento compartido, así como las estructuras de cortesía, todos sobre utilizados por los estudiantes griegos. En un análisis más detallado se pudo comprobar que en las referencias al lector, los HN hacían un uso más elevado de las formas exclusivas *nosotros* y *nos*, mientras que los HNN empleaban con más frecuencia las fórmulas de cortesía *usted* y *le* para dirigirse al director del medio o la autora del texto estímulo. Moya Muñoz (2016: 271) en el análisis que realizó en el género comentario de lectores en la prensa chilena, encontró que estas fórmulas eran muy escasas, concluyendo que «en los discursos que se construyen por medio de los ordenadores prima una cercanía comunicativa que deriva en rasgos de coloquialidad».

En cuanto a los argumentos que apelan al sentido común, la diferencia entre los dos grupos se centró en el uso de argumentos *ad populum* (*Afortunadamente todos creen que...*), sobre usados por los HNN para presentar premisas «universalizables», es decir, aceptables, en principio, por todos los miembros del auditorio universal» (Perelman 1997: 18). Esta conclusión contradice los resultados de Koutsantoni (2003) quien, en el marco del discurso académico en griego e inglés, llegó a la conclusión que el uso de estos marcadores era limitado en los ensayos escritos por ambos estudiantes novatos anglófonos en L1 y grecófonos en inglés L2. La investigadora atribuyó las tendencias observadas a la asimétrica relación de poder entre los estudiantes y sus examinadores y la creencia general que la solidaridad solo puede operar en casos de ausencia de poder.

Por último, las estructuras de cortesía presentaron variaciones en los distintos movimientos que se suelen incluir en el género epistolar como, por ejemplo, la línea de saludo, la conclusión y la despedida, todos de una frecuencia más elevada en los escritos de los estudiantes griegos frente a los estudiantes españoles que apenas utilizan este tipo de formalismos en sus cartas. Si se tiene en cuenta lo que señala Pastor Pérez (2010), es decir, que «los convencionalismos que encabezan y despiden las cartas al director se han limitado hasta la eliminación en algunas publicaciones», se podría sugerir que los HN son conscientes de las convenciones genéricas de este tipo de cartas en la era digital. No obstante, su actuación deja margen a la interpretación. Si se considera que un buen número de textos fue calificado de banda 3 en la realización de la tarea por no cumplir con los requisitos de este género, se puede entender que no es posible saber si estos autores están familiarizados con lo que describe Pastor Pérez o si simplemente reproducen el género requerido en las Pruebas de Selectividad, que suele ser el comentario de texto

A modo de conclusión, se podría deducir, que las diferencias de nuestros datos con los estudios realizados en el campo se deben a dos factores: primero, estas investigaciones se centran en el discurso académico y, en concreto, al género ensayo argumentativo cuyo propósito comunicativo es distinto al de la carta al director y segundo los grupos comparados en estas investigaciones provienen de otros contextos socioculturales. Aunque carecemos de investigaciones contrastivas en griego y español, las comparaciones realizadas de estos idiomas con otras lenguas como, por ejemplo, el inglés (Koutsantoni 2005; Neff-van Aertselaer & Dafouz Milne 2008; Hatzithodorou & Mattheoudakis 2011) revelan ciertos patrones similares que podrían explicar las similitudes en el modo de emplear los atenuadores y los intensificadores.

6. CONCLUSIONES

Se ha llevado a cabo un estudio contrastivo a partir de un diseño descriptivo-correlacional con el fin de analizar las estrategias metadiscursivas recurrentes en el género carta al director. Para ello, se ha recogido una muestra de cartas escritas en contextos formales por estudiantes novatos grecófonos e hispanófonos. Los resultados globales nos han revelado un claro predominio de los marcadores interaccionales sobre los interactivos, lo que nos indica que, a diferencia del discurso académico, en las cartas al director las estrategias que apelan a las emociones del lector y las que proyectan la integridad del autor superan a las que apelan a la razón.

En este punto sería oportuno hacer unas reflexiones generales sobre las dificultades que hemos afrontado a lo largo de este estudio. En primer lugar, como ya hemos comentado, dada la falta de acuerdo entre los distintos teóricos en proporcionar una tipología común sobre el metadiscurso que facilite la identificación y descripción de los marcadores metadiscursivos, la comparación con las investigaciones realizadas en el campo resulta una tarea difícil. Asimismo, el hecho de que este tipo de modelos estén pensados para el inglés crea problemas y solapamientos. Más concretamente, el inglés es una lengua de sujeto obligatorio, lo que significa que se puede «fragmentar» fácilmente y catalogar el pronombre personal *I* en las autorreferencias y el verbo que lo acompaña como, por ejemplo, *think* en la sub-categoría de los atenuadores. No obstante, en español esta información se transmite mediante la flexión verbal.

Sin embargo, a pesar de las dificultades que conlleva un trabajo de investigación de este tipo, con todas las debilidades anteriormente mencionadas, no deja de ser una valiosa herramienta para la mejora de la metodología de la enseñanza-aprendizaje de L2. En este sentido, el metadiscurso constituye un elemento esencial para la comunicación, puesto que solo cuando se hayan evaluado correctamente los recursos de los que dispone el lector para interpretar cualquier texto, así como su probable reacción, el autor es capaz de construir sus argumentos de forma adecuada. Hasta ahora en la didáctica de L2 la expresión escrita suele limitarse al contenido, es decir, al cómo el autor transmite sus ideas, invirtiendo mucha energía en las reglas y en un exhaustivo inventario de conectores discursivos, dejando de lado las estrategias interpersonales que tienen un papel primordial en la producción de un texto. Mientras que el conocimiento explícito de las reglas puede contribuir parcialmente a aprender cómo escribir, ello constituye solamente una parte del proceso didáctico, lo cual se tiene que complementar con el

conocimiento de las funciones metadiscursivas que adaptan el texto al contexto y la audiencia a la que se dirige.

No obstante, el metadiscurso no debería ser tratado en los manuales didácticos como un recurso estilístico que puede variar a voluntad del autor. La importancia de este modelo descansa en la dinámica subyacente de la retórica que lo contextualiza de acuerdo con el género y el propósito comunicativo. En este sentido, la conciencia de género y de registro ayuda a mejorar las destrezas de escritura de los estudiantes en los distintos niveles educativos, puesto que les guía a escoger de un variado repertorio de opciones aquellas que son apropiadas a la situación comunicativa.

ADSCRIPCIÓN DE LA INVESTIGACIÓN

El corpus utilizado en el presente trabajo ha sido extraído de la tesis doctoral *Patrones retóricos en el género carta al director: análisis contrastivo en contextos naturales y educativos en español y griego* defendida por Panagiota Salapata en 2017 y dirigida por las Dras María Cecilia Ainciburu y Angélica Alexopoulou.

REFERENCIAS BIBLIOGRAFICAS

- Ädel, A. (2006). *Metadiscourse in L1 and L2 English*. Amsterdam: John Benjamins.
- Albelda Marco, M. (2005). El refuerzo de la imagen social en conversaciones coloquiales en español peninsular: la intensificación como categoría pragmática. En: *Estudios de la (des)cortesía en español: categorías conceptuales y aplicaciones a corpora orales y escritos*. Estocolmo: Dunken, pp. 93-118.
- Calsamiglia Blancafort, E. & Tusón Valls, A. (2012). *Las cosas del decir: manual de análisis del discurso*. 3a ed. Barcelona: Ariel.
- Connor, U. (2008). Mapping multidimensional aspects of research: reaching to intercultural rhetoric. En: U. Connor, E. Nagelhout & W. V. Rozycki, edits. *Contrastive rhetoric reaching to intercultural rhetoric*. Amsterdam: John Benjamin, pp. 299-315.
- Consejo de Europa (2002). *Marco común europeo de referencia para las lenguas: aprendizaje, enseñanza, evaluación*. Madrid: Secretaría General Técnica del MEDCD y Grupo Anaya.
- Crismore, A. (1985). *Metadiscourse in social studies texts*. Urbana-Champaign: University of Illinois.
- Crismore, A. & Farnsworth, R. (1989). Mr. Darwin and his readers: exploring interpersonal discourse as a dimension of ethos. *Rhetoric Review*, 8(1), pp. 91-112.
- Crismore, A., Markkanen, R. & Steffensen, M. S. (1993). Metadiscourse in persuasive writing: a study of texts written by American and Finnish university students. *Written Communication*, 10(1), pp. 39-71.
- Dafouz Milne, E. (2006). Estudio de los marcadores interpersonales en el comentario periodístico: estrategias para la identificación autor- lector en el texto. *RESLA*, Volumen 19, pp. 67-82.
- Fuentes Rodríguez, C. (2016). Atenuación e intensificación estratégica. En: *Estrategias argumentativas y discurso político*. Madrid: Arco Libros, pp. 163-221.

- González Arias, C. (2014). El metadiscursos en columnas de opinión y en los comentarios de lectores en un ambiente virtual y público. *Spanish in Context*, 11(2), p. 155 -174.
- Halliday, M. A. K. (1973). *Explorations in the Functions of Language*. London: Edward Arnold.
- Halliday, M. A. K. (1978). *Language as a social semiotic: the social interpretation of language and meaning*. Edward Arnold ed. London: University Park Press.
- Halliday, M. A. K. (1985). *An Introduction to Functional Grammar*. 2ª ed. Londres: Edward Arnold.
- Harris, Z. S. (1970). *Papers in structural and transformational linguistics*. Dordrecht: Reidel Publishing Company.
- Hatzithodorou, A. M. y Mattheoudakis, M. (2011). The impact of culture on the use of stance exponents as persuasive devices: the case of GRCLE and English native speaker corpora. En: A. Frankenberg Garcia, L. Flowerdew & G. Aston, edits. *New Trends in corpora and language learning*. London: Continuum, pp. 229-246.
- Hyland, K. (1998a). *Hedging in Scientific Research Articles*. Amsterdam: John Benjamins Publishing Company.
- Hyland, K. (1998b). Persuasion and context: the pragmatics of academic metadiscourse. *Journal of Pragmatics*, Volumen 30, pp. 437-455.
- Hyland, K. (2005). *Metadiscourse: Exploring Interaction in Writing*. London: Continuum.
- Hyland, K. (2013). Metadiscourse. En: K. Hyland, ed. *Discourse Studies Reader: Essential Excerpts*. London: Bloomsbury Academic, pp. 65-88.
- Hyland, K. (2015). Metadiscourse. En: K. Tracy, ed. *The International encyclopedia of language and social interaction: Volume II*. Oxford: Wiley-Blackwell, pp. 996-1007.
- Hyland, K. (2017). Metadiscourse: What is it and where is it going? *Journal of Pragmatics*, Volumen 113, pp. 16-29.
- Hyland, K. & Milton, J. (1997). Qualification and Certainty in L1 and L2 Students' Writing. *Journal of Second Language Writing*, 6(2), pp. 183-205.
- ISEI-IVEI. (2013). *Evaluación Diagnóstica*. [En línea] Available at: http://ediagnostikoak.net/edweb/cas/materiales-informativos/ED_09_10_11/Crit_correc_%20ESO_2009_10_11.pdf [Último acceso: 10 febrero 2014].
- Kaplan, R. B. (1966). Cultural Thought Patterns in Inter-Cultural Education. *Language Learning*, 16(1-2), pp. 1-20.
- Koutsantoni, D.(2003). *Rhetoric and culture in published and unpublished scientific communication: a comparative study of texts produced by Greek and native English speaking engineers*. Tesis doctoral (PhD). The University of Birmingham.
- Koutsantoni, D. (2005). Greek cultural characteristics and academic writing. *Journal of Modern Greek Studies*, Volumen 23, pp. 97-138.
- Martínez Rodríguez, L. (2005). La participación de los usuarios en los contenidos periodísticos de la Red. En: G. López García, ed. *El ecosistema digital: modelos de comunicación, nuevos medios y público en internet*. Valencia: Servei de Publicacions de la Universitat de Valencia, pp. 269-332.
- Mayor, B. M. (2006). Dialogic and Hortatory Features in the Writing of Chinese Candidates for the IELTS Test. *Language, Culture and Curriculum*, 19(1), pp. 104-121.
- Moya Muñoz, P. (2016). *Análisis comparativo de las estrategias metadiscursivas en los comentarios de noticias en la prensa digital española y chilena*, Valencia: Universitat Politecnica de Valencia.

- Muñoz, A., Mueller, J., Álvarez, M. & Gaviria, S. (2006). Developing a Coherent System for the Assessment of Writing Abilities: Tasks and Tools. *Íkala, revista de lenguaje y cultura*, 11(17), pp. 265-296.
- Neff-van Aertselaer, J. & Dafouz Milne, E. (2008). Argumentation patterns in different languages: An analysis of metadiscourse markers in English and Spanish texts. En: M. Pütz & J. Neff van Aertselaer, edits. *Developing Contrastive Pragmatics: Interlanguage and Cross-Cultural Perspectives*. Berlin: Mouton de Gruyter, pp. 87-102.
- Pastor Pérez, L. (2010). *Teoría de las cartas al director: la gestión periodística del público I*. Barcelona, UOCpress.
- Perelman, C. (1997). *El imperio retórico: retórica y argumentación*. Santafé de Bogota: Grupo Editorial Norma.
- Salapata, P. (2017). *Patrones retóricos en el género carta al director: análisis contrastivo en contextos naturales y educativos en español y griego*. Tesis Doctoral inédita. Universidad Antonio de Nebrija.
- Shaw, S. & Falvey, P. (2008). The IELTS Writing Assessment Revision Project: Towards a revised rating scale.
- Suau Jiménez, F. (2011a). El metadiscurso interpersonal y la persuasión en géneros periodísticos: aplicación a la didáctica de la traducción. *Quaderns de Filologia. Estudis Linguistics*, 16(13), pp. 249-268.
- Suau Jiménez, F. (2011b). La persuasión a través del metadiscurso interpersonal en el género Página Web Institucional de Promoción Turística en inglés y español. En: M. V. Calvi & G. Mapelli, edits. *La lengua del turismo: géneros discursivos y terminología*. Bern: Peter Lang, pp. 177-200.
- Vande Kopple, W. (1985). Some exploratory discourse on metadiscourse. *College Composition and Communication*, 26(1), pp. 82-93.
- Vande Kopple, W. (2002). Metadiscourse, discourse and issues in composition and rhetoric. En: E. Barton & G. Stygall, edits. *Discourse studies in composition*. New Jersey: Hampton Press, pp. 91-113.
- Vande Kopple, W. (2012). The importance of studying metadiscourse. *Applied Research in English*, 1(2), pp. 37-44.
- Williams, J. (1981). *Style: ten lessons in clarity and grace*. Chicago: University of Chicago Press.
- Williams, J. M. (1990). *Style: toward clarity and grace*. Chicago: The University of Chicago Press.
- Williams, J. M. (2007). *Style: lessons in clarity and grace*. 9a ed. New York: Pearson-Longman.

La expresión de la actitud en un corpus oral de presentaciones académicas de estudiantes ELE y hablantes de herencia de español en contexto universitario en Estados Unidos

The expression of attitude in an oral corpus of academic presentations of L2 Students and Heritage Speakers of Spanish at the university context in the United States

Eva Gómez García

Brown University

eva_gomez_garcia@brown.edu

RESUMEN

En este artículo se presentan los resultados del análisis de un corpus (n=15) de presentaciones académicas orales (PAO) en español realizadas por estudiantes universitarios hablantes de herencia y de ELE estadounidenses. El objetivo es determinar la manera en la que se manifiesta la postura a través del metadiscurso. Con este fin se identificaron las marcas metadiscursivas de actitud, su frecuencia de uso y función retórica según las categorías en la taxonomía de Hyland. Los resultados muestran que el grupo de herencia hace mayor uso de estos recursos y que los adjetivos evaluativos en ambos grupos es la forma lingüística más empleada para codificar las muestras de actitud.

Palabras clave: análisis de corpus, postura, marcadores de actitud, metadiscurso

ABSTRACT

This article reports on the results of the analysis of a corpus (n=15) of oral presentations in Spanish given by undergraduate Heritage Speakers and L2 students in the US. The goal is to determine how students express their stance through metadiscourse. The analysis focused on attitude markers, their frequency and rhetorical function as proposed by Hyland. The results reveal that Heritage Speakers use these attitude markers the most and that evaluative adjectives, in both groups, is the most common linguistic realization to codify stance.

Keywords: corpus analysis, stance, attitude markers, metadiscourse

1. INTRODUCCIÓN

Este estudio se enmarca dentro del área de Corpus de Aprendices de Lenguas en formato computacional y de Análisis del Discurso. A partir de un corpus de 15 presentaciones académicas orales (PAO) en español, producidas por hablantes de herencia (n=6) y estudiantes de ELE (n=9), se analiza la expresión de la actitud para conectar con la audiencia a través de estrategias metadiscursivas. Las presentaciones académicas se tratan de un género formativo esencial y el conocimiento de sus características contribuye al aprendizaje de las prácticas discursivas en el ámbito de educación superior y, por lo tanto, es conducente a la adquisición de la competencia académica. Es debido a esto por lo que el trabajo con corpus electrónicos de aprendices se revela fundamental para poder llevar a cabo un análisis del funcionamiento de la lengua del que poder seguir aprendiendo. De entre las múltiples oportunidades de análisis que nos ofrece el discurso a través del uso de corpus, nos centramos en el metadiscurso, cuya investigación siempre ha estado estrechamente vinculada con la Lingüística de Corpus. El objetivo principal de este estudio es entender la forma en que estos dos grupos de estudiantes utilizan el español en una presentación oral y ver qué revela el análisis del metadiscurso sobre los mecanismos lingüísticos de su discurso oral en conexión con la manifestación de su postura y su expresión personal.

2. CONTRIBUCIONES DE LOS CORPUS PAO A LA DIDÁCTICA DE ELE: RELEVANCIA DE LA PRODUCCIÓN DE APRENDICES DE ESPAÑOL

El interés por entender cómo funciona la expresión personal en el discurso académico ha llevado a su investigación a partir de diferentes metodologías complementarias, desde descripciones de una única muestra de texto a investigaciones empíricas de los patrones generales con corpus electrónicos gracias a las nuevas tecnologías (Biber 2006). Es por ello, por lo que, en el campo de la adquisición de lenguas, la creación de los corpus de aprendices de lenguas se convierte en esencial y su utilidad es evidente a la hora de extraer información sobre un género textual o cuando se busca entender un fenómeno lingüístico.

El proyecto MiCASE (*Michigan Corpus of Academic Spoken English*) es uno de los esfuerzos más productivos para describir los registros universitarios orales en inglés. Sin embargo, en el campo del español académico, como apunta Ainciburu (2012), "es difícil disponer de material integrado en corpus" y "esa dificultad se potencia aún más cuando el texto meta es de producción oral". A pesar de intentos más articulados como el Corpus Académico y Profesional del Español de la Pontificia Universidad Católica de Valparaíso (PUCV-2006) y el Proyecto europeo ADIEU (Discurso Académico en la Unión Europea, 2001), sigue existiendo una ausencia de corpus representativos de géneros académicos de producción ELE no nativa.

Asimismo, si se considera que la PAO se trata de un género que todavía no se ha estudiado de forma exhaustiva en español (Villar, 2014), cabe esperar que la existencia de corpus de aprendices sea todavía reducida. Sin embargo, hay que señalar que en años recientes se han producido intentos individuales de elaboración de corpus de presentaciones académicas producidas por estudiantes de español.

Una de las primeras contribuciones al corpus de PAO de aprendices de ELE es la de Villar (2011, 2014). Esta investigadora se interesa por la producción oral de estudiantes alemanes de ELE y para llevar a cabo un estudio empírico crea el Corpus PRESAC (Presentaciones Académicas de fin de curso) en 2011/2013. Este corpus está formado por

tres subcorpus de presentaciones académicas comparables y un total de 57 PAO individuales: 20 de estudiantes hispanohablantes, 18 de estudiantes alemanes y 19 de estudiantes alemanes de ELE. Con este corpus realiza un análisis macrotextual, discursivo y contextual de la PAO y propone un modelo de estructura global textual. Ainciburu (2012/2013) contribuye también a los corpus de aprendices de ELE en el campo de EFE (Español con Fines Específicos) con el corpus BANKIT (Presentación de Bancos y productos financieros). Este corpus se compone de presentaciones orales de evaluación de fin de curso de estudiantes italianos. Ainciburu y Villar (2013) profundizan en la caracterización del género PAO con muestras de ambos corpus y concluyen que mientras que las muestras italianas presentan una estructura monológica, la estructura en el grupo alemán es dialógica por la participación del docente.

Posteriormente, Vergara Padilla (2017) elabora un corpus (2012-2016) para profundizar en la descripción macroestructural y discursiva de este género realizado por aprendices de ELE. El corpus Nebrija-CAO (Corpus Académico Oral) está disponible en TalkBank en formato CHAT (Programas CLAN; McWhinney 2000) y se compone de 67 presentaciones orales grabadas en vídeo de estudiantes estadounidenses ELE. De este corpus se emplearon 20 presentaciones monologadas en la investigación de Vergara Padilla y Mavrou (2019) en la que parten del modelo de Villar (2014) para proponer una estructura global de la PAO dirigida al ámbito formativo para aprendices de ELE.

Todos estos corpus, a pesar de no ser consistentes en número y naturaleza de las muestras, suponen una contribución valiosa que permite seguir profundizando en la producción textual de estudiantes no nativos en lengua española en el género discursivo de la presentación oral. Ainciburu (2018) recoge un panorama de los rasgos textuales estudiados en los últimos años en el ámbito del español con fines académicos dirigido a estudiantes no nativos. Esta investigadora considera que no es metodológicamente necesario comparar las muestras en ELE con un modelo nativo puesto que:

los estudios sobre producciones no nativas tienen un valor por la descripción del fenómeno que realizan y no porque agreguen un peso a la hipótesis de transferencia de la macroestructura o de cualquier otro elemento lingüístico típico de los textos de la lengua materna en la extranjera (142).

Por lo tanto, no cabe duda del valor de los corpus de aprendices de español y su compilación, pese a las dificultades metodológicas existentes, supone un impulso importante para el avance de esta disciplina.

El corpus empleado en esta investigación también recoge muestras en español de aprendices ELE y hablantes de herencia estadounidenses. Estos últimos son hablantes bilingües que han adquirido su español en el hogar o en su comunidad en Estados Unidos y en algunos casos ha realizado cursos de español en un contexto educativo. En definitiva, el objetivo no es señalar cuánto se aleja su actuación de la producción nativa, sino más bien describir cómo los estudiantes se proyectan en sus presentaciones mediante marcadores de actitud para conectar con su audiencia, por cuanto esto nos puede enseñar de las necesidades pedagógicas específicas de estos dos grupos, dado su diferente perfil lingüístico y sociocultural en el contexto de Estados Unidos.

3. EL METADISCURSO Y LA EXPRESIÓN DE LA ACTITUD

Las investigaciones interesadas por la presencia y el funcionamiento del metadiscurso en los textos han estado fundadas y muy estrechamente ligadas al campo de la Lingüística de

Corpus (Hyland, 2009). En este estudio se emplea el metadiscursos como marco teórico y la taxonomía propuesta por Hyland (2005a).

El metadiscursos revela que cualquier tipo de discurso va más allá de la mera transmisión de información, más bien los textos son una manifestación de que sus autores son conscientes tanto de los lectores como de las consecuencias de su propio texto (Hyland, 2005b). Hyland afirma que, en la construcción de un argumento convincente en el ámbito académico, el control del nivel de personalidad en el texto se convierte en central, el autor busca presentarse a sí mismo y su trabajo de forma creíble, y para ello apela a la solidaridad de los lectores, evalúa el contenido que comparte y reconoce puntos de vista alternativos. En definitiva, para Hyland la interacción en la escritura académica implica adoptar un punto de vista con respecto a los temas analizados y respecto a los otros participantes en la comunicación, que a su vez tienen sus propias perspectivas.

Al igual que sucede con la escritura académica, la presentación oral también constituye un proceso y un producto socialmente construido. Así, de la misma manera, en el caso del género PAO es importante que el expositor sea consciente de la medida en que debe manifestarse en su discurso, de forma que pueda establecer un vínculo efectivo con su audiencia.

Hyland (2005a, 2005b) ofrece un marco formado por dos dimensiones (interactiva e interaccional) para el análisis de los recursos lingüísticos del posicionamiento intersubjetivo. Este estudio se centra en los marcadores de actitud con los que se manifiesta la postura dentro de la dimensión interaccional, la cual “se refiere a la forma en que los escritores llevan a cabo la interacción, colándose en el discurso y haciendo comentarios sobre su propio mensaje” (Hyland 2005a: 49). La siguiente tabla recoge las categorías a través de las que se manifiesta la postura dentro de esta dimensión:

Dimensión interaccional	Postura	-Atenuadores -Intensificadores -Marcadores de actitud -Automención
-------------------------	---------	---

Tabla 1. Recursos clave de la manifestación de la postura (Hyland, 2005b)

La postura se puede entender como una dimensión actitudinal e “incluye características que se refieren a la manera en que el autor se presenta y transmite sus juicios, opiniones y compromisos” (Hyland, 2005b: 176). Según la propuesta de Hyland (2005b), la *postura* comprende tres componentes principales: la *evidencialidad*, el *afecto* y la *presencia*. Se refiere, por tanto, a las maneras en que los autores a través de diferentes recursos marcan sus textos para comentar la precisión o credibilidad de una hipótesis, en qué grado se comprometen con esa información o la actitud que quieren expresar al respecto.

De estos componentes de la postura, nos interesa el *afecto* puesto que alude a la expresión de la actitud y como afirma Hyland “involucra una amplia gama de actitudes personales y profesionales hacia lo que se dice, incluidas las emociones, las perspectivas y las creencias” (2005b: 178). El *afecto* se manifiesta a través de los *marcadores de actitud* que son recursos que indican “la actitud afectiva, más que epistémica, del autor hacia las propuestas, transmitiendo sorpresa, acuerdo, importancia, frustración, etc., más que compromiso” (2005a: 53).

Existen diferentes mecanismos para manifestar actitud en el discurso, pero se expresa más explícitamente mediante *verbos de actitud* (por ejemplo, *acordar, preferir*), *adverbios oracionales* (lamentablemente) y *adjetivos* (*apropiado, lógico, notable*). Cuando se emplean estos recursos, no solamente se expresa una postura, sino que se asume que el lector o la audiencia comparte estas actitudes, valores y reacciones. La presencia de estos elementos en una presentación oral refleja la actitud del expositor sobre el contenido que comparte.

4. MAPA DE LA PRESENCIA DEL METADISCURSO INTERACCIONAL EN DIFERENTES GÉNEROS Y TEXTOS

En esta sección se realiza un recorrido por diferentes géneros académicos, tanto escritos como orales, recogidos en corpus para dar cuenta del papel que desempeñan los marcadores de actitud en cada uno de ellos.

4.1. El metadiscursos en géneros académicos escritos

Las investigaciones de Hyland (2005a, 2005b, 2009) y Hyland y Jiang (2016) son reflejo de la importancia de los elementos metadiscursivos, tanto interactivos como interaccionales, dentro del discurso académico escrito en inglés. Los corpus con los que han trabajado estos investigadores para llevar a cabo sus investigaciones están formados por artículos de investigación publicados en diferentes disciplinas académicas. Dentro de la dimensión interaccional, Hyland se ha centrado sobre todo en el estudio de recursos como los *atenuadores, intensificadores y marcadores de actitud*. A su vez, el trabajo de Hyland ha tenido un enorme impacto en numerosas investigaciones que se han interesado no solo por el discurso escrito y oral académico en inglés, sino también por otros géneros y otras lenguas.

Con el fin de examinar cómo se logra la interacción en el discurso académico y las preferencias de las comunidades disciplinarias en la construcción de los autores y sus lectores, Hyland (2005b) aborda la expresión de la postura en 240 artículos de investigación pertenecientes a ocho disciplinas representativas de la práctica académica. El análisis del corpus revela que la expresión de la postura constituye una característica relevante de la escritura académica. Los marcadores de postura son muy comunes (30,9 por mil palabras, en adelante, pmp) y dentro de esta categoría los *marcadores de actitud* es el segundo recurso más empleado con un 20,5% del total o 6,4 ítems pmp. En cuanto a la distribución por disciplinas, esta investigación revela que los marcadores de actitud están presentes en todo el corpus, pero también destacan en las humanidades y ciencias sociales, lo que indica que los autores asumen explícitamente posiciones más personales y comprometidas.

En una investigación más reciente, Hyland y Jiang (2016) realizan un estudio diacrónico de la manifestación de la postura en tres periodos de tiempo diferentes (1965, 1985 y 2015) en los últimos 50 años. Para ello, crean tres corpus (2,2 millones de palabras) de artículos de investigación publicados en cinco revistas y pertenecientes a cuatro disciplinas diferentes (lingüística aplicada, biología, ingeniería y sociología). Los resultados revelan que se está produciendo un cambio en la expresión de la postura. En concreto, en el ámbito de la lingüística aplicada se observa una tendencia a emplear estos recursos menos. Por su parte, en los artículos de ciencias, en especial el caso de ingeniería eléctrica, presentan un aumento en la presencia de sus autores. En cuanto a la expresión del afecto a través de marcadores de actitud, en general se produce una reducción en su manifestación

explícita en lingüística aplicada (-26%) y en sociología (-22%), un aumento en artículos de ingeniería eléctrica (+73%) y apenas cambios en biología. En definitiva, Hyland y Jiang (2016) concluyen que se ha producido un movimiento hacia la homogeneización. Los marcadores de actitud constituyen una opción marcada con la que siempre se crea un mayor impacto. En cuanto a su realización en el corpus, se materializan a través de formas con fuerte significado positivo o negativo (*sorprendente, fascinante*) para manifestar juicio. En todas las disciplinas las opciones más empleadas fueron las formas *important* (importante) y *even* (incluso), mientras que para comentar los datos o los resultados presentados destaca el uso de *expected* (esperado) e *interesting* (interesante).

Los marcadores de actitud también se han estudiado en artículos de investigación en español. Beke (2005) analiza quince artículos (89.135 palabras) de una Revista de Pedagogía con el fin de determinar la manera en que los investigadores en educación comunican su posición y su compromiso con la disciplina. Entre las marcas metadiscursivas identificadas, los marcadores actitudinales ocupan el tercer lugar (21,59%/2,68 pmp) y su presencia sugiere que los investigadores en educación muestran cierta necesidad de evaluar sus planteamientos. Para marcar su actitud los autores recurren al verbo *deber* y expresiones afines (*hay que/de, tener que, es preciso, es necesario*), que suman un total de 101 casos (42%). También Beke (2005) identificó una variedad de adjetivos (*inevitable, alentador, curioso, etc.*), adverbios (*afortunadamente, contradictoriamente, etc.*) o combinaciones de ambos (*altamente compleja, harto difícil, muy curioso, etc.*).

Mur Dueñas (2010) se interesa en estudiar el papel del contexto cultural en la expresión de valores actitudinales a través de marcadores de actitud (Hyland, 2005b) en un corpus de 24 artículos de investigación en Dirección y Organización de Empresas en el contexto norteamericano (12) y español (12). Los resultados indican una frecuencia de uso (8,1 pmp inglés y 7,7 pmp español) y tendencias en los propósitos retóricos de los marcadores actitudinales similares, lo que sugiere que existen valores disciplinares compartidos. En cuanto a las categorías léxico-gramaticales empleadas, se identificaron adjetivos, verbos, sustantivos, adverbios y otras expresiones. Los adjetivos son la forma más frecuente con un 61,6% en inglés y un 64,4% (5,0 pmp) en español. En el corpus español se identificaron 67 adjetivos actitudinales diferentes, los más recurrentes fueron: *principal, importante, relevante, gran, necesario, bueno, adecuado, válido y escaso*. El segundo lugar lo ocuparon los sustantivos (1,5 pmp/ 19,7%) como *importancia y limitación*, seguidos por los adverbios (0,6 pmp/ 7,8%) como *solo, solamente o únicamente* y los verbos (0,4 pmp/ 5,1%) como *aportar y contribuir*, que son muy escasos. Mur Dueñas (2010) clasifica también los marcadores según su valor, siguiendo la taxonomía de Swales y Burke (2003), y los marcadores de actitud centralizados superan los polarizados.

Por último, cabe destacar la relevancia que adquieren los marcadores de actitud en el género académico de reseña de artículos de investigación por su papel esencial a la hora de formular una evaluación. Por lo tanto, este género contribuye enormemente a entender el funcionamiento de estos marcadores, ya que estos textos se convierten en espacios para la expresión de ideas, actitudes y evaluaciones. Azar y Hashim (2019) examinaron las estrategias de evaluación a través de marcadores de actitud en un corpus de 32 reseñas de artículos de investigación en inglés, publicados entre 2000-2007, en una revista sobre lingüística aplicada (ARAL, *Annual Review of Applied Linguistics*). Los resultados muestran que los marcadores de actitud (7,66 pmp del total del corpus) aparecen en las diferentes secciones. La mayoría de los marcadores eran adjetivos actitudinales (casi el 69%), seguidos de adverbios actitudinales, sustantivos y, por último, verbos. La presencia de adverbios, verbos y sustantivos fue escasa. Los marcadores más frecuentes en todas las

secciones fueron *only*, *important* y *issue*. Los marcadores de actitud permiten a los autores hablar de la importancia del contenido y justificar una investigación (*crucial*, *major*, *noteworthy*, *importance*, *significant*), compartir la emoción positiva/negativa que suscitan (*amazing*, *interestingly*, *surprisingly*, *unfortunately*), comentar las limitaciones y vacíos existentes (*issue*, *lack*, *limited*, *short of*, *neglect*) y realizar una evaluación (*adequate*, *caution*, *dilemma*, *marginal*, *obvious*).

4.2. El metadiscurso en géneros académicos orales: clases magistrales

El análisis metadiscursivo también se ha aplicado a géneros académicos orales. Biber (2006) se interesa por la manifestación de la postura en registros universitarios tanto escritos (libros de texto y programas de curso) como orales (clases magistrales y situaciones interactivas de gestión de clase). El sub-corpus con el que realiza su investigación se extrae del corpus T2K-SWAL (*TOEFL 2000 Spoken and Written Academic Language*) y resulta representativo de los registros orales y escritos que los estudiantes universitarios pueden encontrar en Estados Unidos. En términos generales, la postura expresa tanto significados epistémicos como actitudinales, pero en relación a la actitud en esta investigación, cabe señalar que se observa especialmente a través de verbos en la modalidad deóntica, adverbios de actitud (*amazingly*, *importantly*, *surprisingly*) y cláusulas de complemento: mediante verbos de actitud en la estructura *stance verb + that-clause* (*expect*, *hope*, *worry*); adjetivos en la estructura *stance adjective + that-clause/to-clause* de emoción/actitud (*amazed*, *shocked*, *surprised*) o de evaluación (*essential*, *interesting*, *noteworthy*); sustantivos de actitud o perspectiva en la estructura *stance noun + that-clause/to-clause* (*hope*, *view/ failure*, *obligation*, *tendency*). Los resultados de esta investigación ponen de manifiesto que el discurso oral, incluso en contextos académicos, está más cargado de elementos de postura de manera más explícita.

Otras investigaciones sobre el metadiscurso en el discurso oral en contexto académico se han centrado exclusivamente en su uso en clases magistrales. Lee y Subtirelu (2015) emplean el modelo de Hyland (2005a) para analizar dos corpus, uno de 18 lecciones de inglés para fines académicos (L2C2) y otro de 18 clases magistrales universitarias (MiCASE). Los resultados revelan que el metadiscurso interaccional se usó más que el interactivo en un 60% y que los instructores buscan mostrar su postura hacia el contenido e interactuar con los estudiantes. Sin embargo, los marcadores de actitud fueron los recursos menos usados (L2C2: 1,69 pmp / MiCASE: 2,10 pmp). Posteriormente, Kahkesh y Alipour (2017), a partir de un corpus de 40 clases magistrales en inglés (20) y en persa (20), realizan un estudio comparativo de los marcadores metadiscursivos usando el modelo de Hyland. El marcador menos usado fueron los marcadores de actitud con 0,36 pmp en inglés y 0,57 pmp en persa.

4.3. El metadiscurso en discursos públicos orales

Las investigaciones que han abordado discursos públicos orales presentan resultados interesantes respecto al uso de marcadores actitudinales, que pueden ser de interés en el estudio del comportamiento metadiscursivo en PAO. Sukma (2017) emplea ocho discursos de la campaña política de Barack Obama de 2012 usando el modelo propuesto por Dafouz-Milne (2008), concebido para escritura persuasiva, para el análisis de marcadores metadiscursivos interpersonales. La macrocategoría de los marcadores de actitud expresa los valores afectivos de los escritores hacia el texto y los lectores por lo que se orienta más

a la relación del orador/escritor con los lectores/audiencia. Según esta taxonomía, los marcadores de actitud se componen de las siguientes subcategorías: verbos deónticos (*have to, we must understand, needs to*), adverbios actitudinales (*unfortunately, remarkably, pathetically*), adjetivos actitudinales (*it is absurd, it is surprising*) y verbos cognitivos (*I feel, I think, I believe*). Los resultados revelan que los marcadores de actitud representan un 42,10% con 40 ítems y constituyen la categoría con la frecuencia de ocurrencias más alta. En el corpus de discursos se identificaron todas las subcategorías. Se concluye que la estrategia persuasiva de los discursos de Obama consiste en expresar afecto para establecer una relación emocional con sus potenciales votantes.

Tashi y Suksawas (2018) se interesan en la presencia de metadiscurso interaccional, según el modelo de Hyland, en cuatro discursos (5816 palabras) en inglés del primer ministro de Bhutan en los que aborda diferentes temas. Los resultados muestran que los marcadores de actitud (386,8 pmp) fueron los más empleados. Además, esta investigación indaga sobre la influencia que los temas de los discursos pueden tener en el uso de marcadores metadiscursivos interaccionales del orador, aunque la diferencia identificada no fue significativa. Por último, resulta interesante la clasificación que se establece de los marcadores de actitud según la *Appraisal Theory* de Martin y White (2005). Estos recursos pueden tomar la forma de afecto (*affect*), reacciones emocionales; juicio (*judgement*), valoración del comportamiento de las personas; y apreciación (*appreciation*), evaluación del valor de personas/objetos.

4.4. El metadiscurso en presentaciones académicas orales

Los resultados observados en las investigaciones que se ocupan de estudiar el metadiscurso en discursos públicos contribuyen a elaborar un mapa de cómo funciona el metadiscurso en otros géneros orales con características contextuales compartidas. A pesar de la abundancia de estudios sobre el metadiscurso en diferentes géneros y textos, todavía escasean las investigaciones interesadas por su presencia en el género académico de las presentaciones orales. Sin embargo, existen trabajos relativamente recientes que se han centrado en el papel del metadiscurso en el género PAO siguiendo la taxonomía de Hyland (Magnuczné Godó, 2012; Mameghani & Ebrahimi 2017; Mazidah *et al.* 2019; Gómez García 2020). Los estudios de Magnuczné Godó (2012) y Gómez García (2020), a partir de un corpus de aprendices universitarios de inglés (10) y español (15), respectivamente, se han interesado por la interacción con la audiencia a través de marcadores de compromiso en sus presentaciones orales en contexto de clase y concluyen que el uso de este tipo de estrategias interactivas influye en la efectividad y éxito de una presentación.

Mameghani y Ebrahimi (2017) centran su investigación en marcadores de actitud y de compromiso. El corpus sobre el que realizan su análisis lo extraen de MiCASE y se compone de 11 corpus de presentaciones de estudiantes nativos de inglés con un total de 143.373 palabras. Los resultados revelan que la presencia de los marcadores actitudinales (685 ítems/ 4,77 pmp) es menor. Dentro de los marcadores actitudinales la realización lingüística más frecuente son los adjetivos evaluativos (620 ítems/ 90%) como *interesting* e *important* seguidos de adverbios evaluativos (65 ítems/ 10%). Los resultados obtenidos sobre los marcadores de actitud indican que el uso de estas formas contribuye a que las presentaciones sean más subjetivas. Los investigadores concluyen que la naturaleza del género PAO influye en el hecho de que los estudiantes no hagan uso de estas formas al considerar que están simplemente presentando información previamente validada y que se considera un hecho. Otra hipótesis es que el conocimiento limitado de los estudiantes sobre

el tema de la presentación influya en cómo se manifiestan los estudiantes a la hora de presentar y no se sientan capacitados para evaluar el contenido de su presentación, por lo tanto, el uso de marcadores de actitud es más reducido.

Por último, Mazidah *et al.* (2019) estudiaron la presencia de todas las categorías interaccionales, coincidiendo con los resultados obtenidos por Mameghani y Ebrahimi (2017) respecto a los marcadores de actitud. Para ello, estos investigadores compilaron un corpus formado por 19 presentaciones académicas en un curso de un programa universitario de enseñanza de inglés. Los resultados revelan que los marcadores de actitud ocuparon el último lugar (4 ítems/ 0,95%). Además, se manifestaron por igual con dos ítems en forma de verbos en la modalidad deóntica (*look, we can look*) y dos ítems en la forma de adjetivos (*important*). Mazidah *et al.* (2019) concluyen que los estudiantes no manifiestan su actitud hacia las proposiciones en sus presentaciones orales.

La escasez de investigaciones que se han ocupado de los marcadores de actitud en presentaciones orales de estudiantes no nativos pone de relieve la necesidad de que se siga estudiando este género.

5. METODOLOGÍA

En este apartado se presentan las preguntas de las que se parte y que dan lugar a la investigación empírica de este estudio. Después, se describe el corpus, los informantes y el procedimiento seguido.

5.1. Preguntas de investigación

Los estudios empíricos existentes en diferentes géneros sobre marcadores de actitud han identificado que este recurso se manifiesta de manera explícita a través de determinadas categorías léxico-gramaticales de naturaleza actitudinal, en concreto sustantivos, adjetivos, verbos y adverbios. Sin embargo, no existen estudios empíricos que hayan caracterizado el patrón metadiscursivo relativo a la actitud de presentaciones orales en español de estudiantes hablantes de herencia y ELE.

Por lo tanto, a partir del marco conceptual de este estudio y los resultados en estudios previos, con el fin de contribuir a la escasez de investigaciones sobre el metadiscurso en presentaciones orales se formulan las siguientes preguntas.

Dados dos grupos de estudiantes universitarios, hablantes de herencia y ELE, que realizan presentaciones académicas en español:

1. ¿Emplean los dos grupos un número similar de marcadores de actitud?
2. ¿Qué tipo de patrón metadiscursivo (categoría léxico-gramatical, frecuencia y función retórica) presentan los marcadores de actitud?

5.2. Informantes y corpus

Con el fin de llevar a cabo la caracterización del patrón retórico de actitud en las PAO de los dos grupos, se procedió a la elaboración de un corpus. Los participantes son estudiantes en un curso sobre la diáspora latinoamericana en una universidad en el noreste de Estados Unidos. Se requiere un nivel de español de B2-C1, según el Marco Común Europeo de Referencia, para matricularse en este curso. Este requisito se constató a través de

puntuaciones obtenidas a través de exámenes oficiales (puntuación de más de 750 en *SAT II* y 5 en *AP Literature*) o la superación del curso más avanzado del programa de lengua.

Las muestras orales se obtuvieron a partir de la grabación de las presentaciones orales realizadas por los estudiantes. Los temas de las presentaciones se establecían a partir del contenido del curso que estaba siempre relacionado con diferentes aspectos de la diáspora latinoamericana en Estados Unidos como autores destacados, el papel del español y su pérdida, análisis de películas que abordan los efectos de la migración a este país, etc. Los estudiantes podían decidir sobre el tema de la presentación en función del calendario establecido por el profesor del curso.

El corpus de hablantes de herencia (corpus HH, en adelante) se compone de seis presentaciones orales con un número total de 7.574 palabras y el corpus de estudiantes de español como lengua extranjera (corpus ELE, en adelante) comprende nueve presentaciones orales con un total de 8.363 palabras. El corpus para este estudio está compuesto por 15 PAO (15.937 palabras) y se puede consultar en TalkBank (MacWhinney, 2000) dentro de la base de datos SLABank en español dedicada a la adquisición de segundas lenguas.

5.3. Procedimiento

El primer paso fue la elaboración de los corpus mediante la grabación en audio de las PAO. Los estudiantes no recibieron ningún tipo de instrucción previa a la realización de la presentación y desconocían el objeto de estudio de esta investigación.

Después de realizarse la transcripción de los registros orales, se inició una fase exploratoria con el fin de identificar elementos actitudinales explícitos a partir del metadiscurso como marco teórico (Hyland, 2005a) para estudiar este fenómeno. El análisis se centró en las categorías léxico-gramaticales que pueden adoptar los marcadores de actitud: adjetivos, sustantivos, verbos y adverbios. En esta primera fase exploratoria, se realizó una lectura de las transcripciones para extraer todas las categorías de marcadores de actitud presentes en el corpus.

Posteriormente, una vez se determinaron las realizaciones lingüísticas que eran marcadores de actitud en el corpus, se llevó a cabo una fase comprobatoria/ contrastiva en la que se realizaron búsquedas automáticas de las categorías léxico-gramaticales identificadas en la fase previa, a través de la herramienta AntConc (v. 3.4.4, Anthony, 2014) de análisis textual y de concordancias. En esta fase se comprobó también el contexto en el que aparecían para confirmar su función como marcador actitudinal y que eran una manifestación de la postura del expositor. Por último, el corpus se transcribió también en formato CHAT (Programas CLAN; MacWhinney, 2000).

Una vez cuantificados los datos de las diferentes categorías analizadas, se normalizaron sus frecuencias a ocurrencias por 1.000 palabras (pmp) para facilitar la comparación entre los dos corpus. También se aplicó un test estadístico Chi-cuadrado para determinar si las diferencias observadas eran o no significativas (se estableció el nivel de significancia en $<0,05$).

6. RESULTADOS Y DISCUSIÓN

En esta sección se presentan los resultados del análisis metadiscursivo de la categoría de marcadores de actitud del corpus HH y ELE. En primer lugar, se exponen los datos relativos

a su presencia en los corpus y después se recogen las categorías léxico-gramaticales que adoptan. Por último, se comparten ejemplos extraídos de los corpus para ilustrar su función retórica.

6.1. Presencia de marcadores de actitud en el corpus

Los resultados del estudio empírico confirman que en los dos corpus analizados se hace uso de marcadores de actitud. Como se observa en la Tabla 2, con respecto a la totalidad del corpus, en el grupo HH se emplea en 175 ítems (23,10 pmp/ 2,31%) y en el grupo ELE se manifiesta en 127 ítems (14,11 pmp/ 1,52%). Por lo tanto, el uso de marcadores de actitud está más presente en la producción de hablantes de herencia.

Categoría	Corpus HH (n=6)	Corpus ELE (n=9)
n° total palabras corpus	7574	8363
Frecuencia ítems	175	127
Frecuencia ítems pmp	23,10	15,18
Media	29,17	14,11
Desviación estándar	10,21	7,42

Tabla 2. Distribución global de marcadores de actitud

En la Figura 1, se presenta la distribución de los marcadores de actitud mediante porcentajes para facilitar el contraste. Se observa que los marcadores de actitud se emplean más en el corpus HH con un 2,31% que en el ELE con un 1,52%.

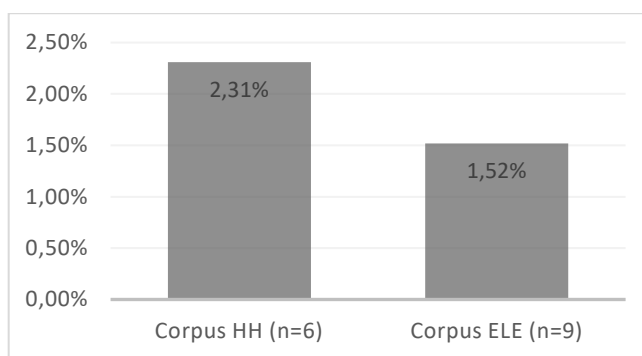


Figura 1. Porcentaje de marcadores de actitud

Además de los datos presentados, se ha aplicado un test estadístico de Chi-cuadrado que confirma que las diferencias existentes entre los valores observados en el uso de marcadores de actitud son estadísticamente significativas, $\chi^2 (1, N = 15) = 12.90, p = .0003$.

Los resultados obtenidos revelan que en los dos corpus del género PAO analizados se emplean marcadores de actitud, aunque en el corpus HH su presencia es mayor y la diferencia es significativa con respecto al corpus ELE. Estos datos contrastan con los obtenidos con artículos de investigación por Hyland (2005b) en los que los marcadores de actitud eran el segundo recurso más empleado con 6,4 ítems pmp y por Mur Dueñas (2010) con una frecuencia de 8,1 pmp en inglés y 7,7 pmp en español. Los datos de Azar y Hashim

(2019) con 7,66 pmp en su corpus de reseñas de artículos concuerdan más con estos resultados.

En los corpus PAO su frecuencia es superior con 23,10 pmp HH y 14,11 pmp ELE. En cierta manera, este dato confirma las conclusiones a las que llega Biber (2006) al afirmar que el discurso oral en contexto académico está más cargado con postura. Cabe destacar que los datos obtenidos están en la línea de las investigaciones sobre el discurso oral público (Sukma 2017, Tashi y Suksawas 2018) en las que los marcadores de actitud constituyen la categoría con la frecuencia de ocurrencias más alta. Sin embargo, contrasta también con los resultados de Mameghani & Ebrahimi (2017) en PAO que indican que los marcadores actitudinales tienen una ocurrencia de 4,77 pmp.

6.2. Patrón metadiscursivo de la expresión de la actitud

Del análisis del corpus, además del alcance de la manifestación de los marcadores de actitud, también se desprende el tipo de recursos usados. En esta sección se profundiza en el metadiscurso para expresar la actitud a través del tipo de palabras empleado, ejemplos de su realización extraídos de los corpus y las funciones retóricas que desempeñan.

Los marcadores de actitud adquieren la forma de verbos, sustantivos, adjetivos y adverbios. En la Tabla 3 se comparte la distribución de los marcadores identificados según la forma lingüística que adoptan en el corpus.

Categoría léxico-gramatical	Corpus HH			Corpus ELE		
	Total	pmp	%	Total	pmp	%
Verbos	32	4,22	0,42	20	2,39	0,24
Sustantivos	42	5,55	0,55	21	2,51	0,25
Adjetivos	81	10,69	1,07	84	10,04	1,00
Adverbios	20	2,64	0,26	2	0,24	0,02

Tabla 3. Frecuencia y porcentaje de las subcategorías de los marcadores de actitud

La Figura 2 recoge estos resultados en forma de porcentajes para mostrar el contraste entre ambos corpus.

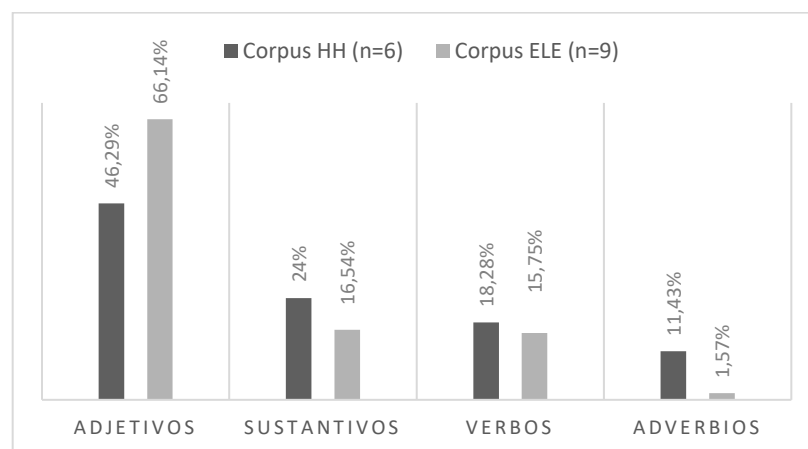


Figura 2. Contraste de la distribución de las formas de los marcadores de actitud

En ambos corpus se presenta un patrón metadiscursivo similar en cuanto a la frecuencia de uso de cada una de las categorías. Los adjetivos son la forma lingüística más empleada para manifestar actitud (46,29% HH y 66,14% ELE), seguidos de sustantivos (24% HH y 16,54% ELE), verbos (18,28% HH y 15,75% ELE) y, por último, adverbios (11,43% HH y 1,57% ELE).

Estos resultados coinciden con otras investigaciones en cuanto a que los adjetivos son el tipo de palabra más frecuente, pero parcialmente respecto al patrón que presentan el resto de las categorías. En el análisis de Mur Dueñas (2010) con artículos de investigación en español también los adjetivos actitudinales fueron la forma más frecuente (67 ítems/64,4%), seguida de los sustantivos (19,7%), mientras que los adverbios se situaron en tercer lugar (7,8%) y los verbos ocuparon el último puesto (5,1%). El estudio de Azar y Hashim (2019) con reseñas de artículos revela también la mayor presencia de adjetivos actitudinales (69%) seguidos de adverbios, sustantivos y verbos. Aunque la investigación de Mameghani y Ebrahimi (2017) con PAO solo considera las categorías de adjetivos y adverbios, de nuevo son los adjetivos (90%) los que superan la presencia de adverbios (10%). Todos estos datos ponen de relieve que son los adjetivos la categoría léxico-gramatical con la que los autores marcan su actitud en el discurso.

6.3. Ejemplos de la realización lingüística de los marcadores de actitud

En esta sección se recogen ejemplos de las formas lingüísticas identificadas en la sección anterior, con las que se manifiestan los marcadores de actitud.

6.3.1 Adjetivos

En la Tabla 4 se comparten algunos fragmentos para ilustrar el contexto en el que se identificaron ejemplos de adjetivos evaluativos:

Corpus HH	Corpus ELE
Adjetivos	
(1) y a través de cruzar la frontera solamente fue vagabundo, periodista, no podía ser todo porque esto es <i>incongruente</i> con lo que el Proyecto anti-inmigrante estadounidense toca que los inmigrantes sean [...]	(7) y vivir en Nueva York no fue <i>fácil</i> para los migrantes puertorriqueños.
(2) convertir en ser un laborador, alguien trabajando, un obrero, un bracero, algo <i>sub-humano</i> .	(8) pero también es <i>interesante</i> porque esto es súper <i>irónico</i> porque los mujeres que van a la clínica quieren tener un aspecto más como muy <i>northern lady</i> .
(3) Es muy <i>interesante</i> que los hispanos particularmente sientan esta extranjería.	(9) y eso es un ejemplo del razón porque hay estereotipos súper <i>fuertes</i> en las mujeres Latinas.
(4) y también esto es muy <i>relevante</i> a lo que estábamos diciendo hace un rato que el inmigrante es <i>limitado</i> en su identidad.	(10) eso es <i>importante</i> en el contexto de los inmigrantes porque los inmigrantes muchas veces pueden ser <i>vulnerables</i> y pueden llevar esa máscara para ser <i>invencibles</i> en vez de ser <i>vulnerables</i> .
(5) y en esta insuperabilidad de diferencias culturales que a la vez hacen más <i>difícil</i> que estas minoridades compartan las injusticias en contra de ellos.	(11) Pienso que esta cita es muy <i>rara</i> pero también su palabras son muy <i>fuertes</i> y demuestran la dedicación a la causa y la lucha.
(6) está otra vez marcado por violencia hacia los inmigrantes y a las mujeres en que son las más <i>indefensas</i> .	(12) a causa del simbolismo de la urna como una representación de la muerte me hizo pensar de era un comentario en el peligro y el peligro <i>fatal</i> de nuestras fantasías raciales en una cultura dominante.

Tabla 4. Ejemplos de uso de adjetivos en corpus HH y ELE

Los adjetivos que más ocurrencias presentan en ambos corpus son *importante* (32) e *interesante* (13), como en los fragmentos 3 y 10. Se puede observar en todos los ejemplos que, a través del uso de los adjetivos, se expresan los sentimientos y actitudes afectivas de

los estudiantes hacia el contenido que presentan, como en 6 al calificar a las mujeres de indefensas y en 11 cuando el expositor manifiesta su postura sobre la cita que comparte. Una razón de este resultado es que los estudiantes se sientan muy vinculados al tema de su presentación, especialmente el grupo de hablantes de herencia.

Esta función retórica en las PAO, que justifica el tipo de adjetivos identificados, parece que se aleja del papel que cumplen los adjetivos en los artículos de investigación y reseñas (Mur Dueñas, 2010; Hyland y Jiang 2016; Azar y Hashim, 2019), en los que se emplean, por ejemplo, para realizar una evaluación de las propuestas y teorías de otros investigadores, para manifestar juicio sobre datos y resultados propios o ajenos, para comentar limitaciones o vacíos existentes en una disciplina. Hyland y Jiang (2016) mencionan que *important* es el más empleado y se hace uso de *expected/interesting* en alusión a datos o resultados presentados. También obtienen similares resultados Azar y Hashim (2019) con *important* como el más empleado y adjetivos positivos para hablar de la importancia del contenido (*major, crucial, significant*) o para realizar una evaluación (*adequate, obvious*). Los más usados en su investigación fueron: *important, appropriate, complex, critical, better, significant, major, best, useful* y *main*. Los identificados por Mur Dueñas (2010) en español están en la misma línea y los más recurrentes fueron: *importante, principal, relevante, gran, necesario, bueno, adecuado, válido* y *escaso*.

Los resultados obtenidos por otras investigaciones dentro del discurso oral también presentan características compartidas. Mameghani y Ebrahimi (2017) y Mazidah *et al.* (2019) identifican también como más frecuentes los adjetivos *interesting* e *important* en presentaciones orales, aunque no comparten datos sobre otros adjetivos empleados. Resulta interesante que las conclusiones de ambos trabajos coincidan en que los estudiantes no hacen un uso destacado de este recurso y, en consecuencia, no manifiestan su actitud de manera tan explícita hacia el contenido de sus presentaciones. La hipótesis planteada, en concreto, por Mameghani y Ebrahimi es que es posible que los estudiantes, al no dominar el tema, no se sientan cómodos a la hora de usar marcadores de actitud y produzcan presentaciones menos marcadas.

Sin embargo, los resultados obtenidos en el presente estudio permiten que se establezcan conexiones con los resultados de Sukma (2017) que identifica adjetivos actitudinales como *difficult* en "it was more difficult to save, more difficult to retire" y otros, con los que Obama logra expresar de forma efectiva afecto y emoción en sus discursos para establecer un vínculo emocional con sus votantes.

6.3.2 Sustantivos

A continuación, en la Tabla 5 se recogen algunos fragmentos con ejemplos de sustantivos evaluativos en los corpus:

Corpus HH	Corpus ELE
Sustantivos	
(1) la <i>humanización</i> de la experiencia en ambos lados de la frontera.	(5) y enfoca la experiencia de los militantes en Nueva York y las dificultades de ellos, la discriminación contra ellos
(2) El <i>problema</i> es que muchos latinos, aunque esos que sean ciudadanos, no están viviendo la experiencia de un ciudadano americano porque no son vistos como americanos.	(6) vemos en Vega y también en todos los inmigrantes que hay un desafío de ser un ciudadano en los Estados Unidos durante este tiempo
(3) <i>So esta esta contradicción</i> de qué debe ser la mujer [...]	(7) Y también esta doble conciencia es muy importante (...) porque combina la conciencia del inmigrante y del ciudadano al mismo tiempo, , no solo demuestra las complejidades de la historia sociopolítica con los Estados Unidos [...]

(4) Y ahora en este mundo ser bilingüe se ve como *ventaja*, entonces, hay muchas personas, no solo hispanos pero de todos nacionalidades que están aprendiendo español para esta *ventaja*.

(8) El **problema** con eso es que Junot Díaz piensa que alguien tiene que ser vulnerable para ser víctima y para entender a otras personas.

Tabla 5. Ejemplos de uso de sustantivos en corpus HH y ELE

En el corpus se observan diferentes tipos de sustantivos, pero el que se usa de forma más repetida es *problema* (6 ítems). Los sustantivos identificados permiten a los estudiantes mostrar su actitud positiva (1, 4) o negativa (2, 3, 5, 6, 7) hacia el contenido sobre el que presentan y están muy conectados con temas del curso sobre la diáspora latinoamericana en Estados Unidos.

La función retórica identificada en el corpus PAO analizado difiere de la que caracteriza otros géneros académicos. Mur Dueñas (2010) en español identifica como sustantivos actitudinales más recurrentes *importancia* (18) y *limitación* (12), seguidos en orden de frecuencia por *problema* (8), *fiabilidad* (7) y *aportación* (5). La función de estos sustantivos en artículos de investigación es la de evaluar la propia investigación del autor o el contexto del estudio con la finalidad de tener credibilidad ante sus colegas al reconocer sus avances o limitaciones. Azar y Hashim (2019) también identifican sustantivos actitudinales que permiten realizar una evaluación efectiva en una reseña de artículos de investigación y los más frecuentes fueron *issue, need, support, problem, value, importance, insight, lack, difficulty y constraint*.

En las investigaciones con géneros orales mencionados previamente, como discursos y presentaciones orales, no se recogen datos sobre los sustantivos por lo que no se pueden contrastar con los resultados obtenidos.

6.3.3 Verbos

En este apartado dedicado a los verbos se muestran fragmentos (Tabla 6) con ejemplos de verbos evaluativos identificados en los corpus:

Corpus HH	Corpus ELE
Verbos	
(1) A ver, acá son algunas estrategias que <i>me parecieron</i> más salientes.	(7) yo les recomiendo estos dos libros porque a mí me gustaron muchísimo, mucho más que <i>Drown</i> .
(2) Entonces para pensar en dónde quedan los latinos, los hispanos, <i>es importante pensar</i> en los cinco grupos étnicos	(8) El primero cosa que me llama la atención es el nombre de la clínica, la clínica se llama <i>Northern Ladies</i>
(3) esa idea de doble ciudadanía <i>se tiene que vivir</i> políticamente y es una experiencia cultural.	(9) También tenemos que dar agencia de la creación de masculinidad a los hombres [...]
(4) Entonces, es por eso que <i>no es suficiente luchar</i> por ciudadanía política, pero luchar en hacerse miembros de esa sociedad y ciudadanos en todos los aspectos de la palabra.	(10) También en sus propias palabras Guillermo piensa que su arte es una forma de democracia radical y me gusta esta cita es más larga
(5) yo <i>siento</i> que va a ser más lento por varias razones.	(11) me hizo pensar de nuevo en las prácticas acerca de la frontera y me hizo sentir más perturbada.
(6) Entonces <i>no es completamente necesario</i> poder ser, poder hablar inglés perfectamente.	(12) me interesa mucho porque estos solo son lugares con poblaciones.

Tabla 6. Ejemplos de uso de verbos en corpus HH y ELE

Los estudiantes para marcar su actitud recurren a verbos actitudinales como (*hacer*) *sentir, llamar la atención, gustar, interesar, parecer* + adjetivo (1, 5, 7, 8, 10, 11, 12). También se manifiesta la actitud haciendo uso de la modalidad deóntica representada por el verbo *deber* y expresiones equivalentes como *tener que* (3, 9), *es necesario* (6), *es importante* (2). Con el uso de estas formas los estudiantes no solo comparten su postura

de forma explícita, sino que al hacerlo a través de la modalidad deóntica transmiten la idea de que existe una obligación colectiva de actuar en un determinado sentido.

Los resultados en los estudios con artículos de investigación reflejan una función diferente, ya que se busca evaluar acciones relacionadas con el ámbito académico. Biber (2006) identifica estructuras similares en inglés para los verbos de actitud como *expect*, *hope*, *worry*. Azar y Hashim (2019), aunque los verbos actitudinales fueron los menos frecuentes en su corpus, entre los más empleados se encontraron: *contribute*, *extend*, *fail*, *expect*, *prefer*, *think*, *believe*, *ensure*, *support* y *feel*. En español, Beke (2005) concluye que los autores para marcar su actitud recurren al verbo *deber* y expresiones afines y el corpus de Mur Dueñas (2010) revela que su presencia es escasa e identifica los siguientes ejemplos: *convenir*, *aportar*, *garantizar*, *contribuir*, *limitar*.

Los resultados en las investigaciones sobre PAO, en concreto la de Mazidah *et al.* (2019), se limitan a dos casos de verbos en la modalidad deóntica (*look*, *we can look*) con los que se dan instrucciones a la audiencia (mirar una diapositiva) para gestionar la presentación. Estos datos difieren de los obtenidos en nuestro trabajo.

Por el contrario, se observan más paralelismos con la investigación de Sukma (2017) sobre el discurso de Obama. En este estudio los marcadores de actitud se componen de verbos deónticos como *need to* como en el fragmento "That`s what we need to change" con el que Obama recuerda a la audiencia la obligación colectiva de cambiar el comportamiento negativo de algunos políticos. La función que desempeñan los verbos actitudinales es coincidente aquí con la observada en el corpus PAO.

6.3.4 Adverbios

En esta sección se recogen los adverbios evaluativos empleados en los corpus para facilitar su contraste. En la Tabla 7 se incluyen algunos ejemplos:

Corpus HH	Corpus ELE
Adverbios	
(1) este borrador sistemáticamente destruyó su realidad hacia sí mismo	(4) Y ahora la pérdida intencional que podemos ver en <i>Pollito Chicken</i> con Suzie que tiene una vergüenza de su identidad como hispana, como puertorriqueña y no solamente de su cultura pero también su cuerpo.
(2) la idea de American es reservada solamente para el estadounidense o para el, este, el anglo que viene de Inglaterra, ¿no?	(5) Y también estos inmigrantes participaron activamente en las políticas locales de Nueva York y también las políticas de Puerto Rico.
(3) aunque muchos no quieren incluir a Brasil, y a veces hay los brasileños que no se incluyen y está bien.	

Tabla 7. Ejemplos de uso de adverbios en corpus HH y ELE

Los adverbios identificados permiten a los estudiantes manifestar su actitud sobre el contenido de su presentación (1, 2, 4, 5) y también mostrar su acuerdo (ejemplo 3). En los trabajos con artículos de investigación los autores se sirven de los adverbios, por ejemplo, para valorar de forma positiva o negativa aspectos de su investigación o de otros. Biber (2006) encuentra adverbios de actitud como *amazingly*, *importantly*, *surprisingly*. Por su parte, Azar y Hashim (2019) identifican adverbios como *only*, *necessarily*, *significantly*, *etc.*, con los que se enfatizan y evalúan otras publicaciones.

En español, Mur Dueñas (2010) encuentra que los adverbios no son nada frecuentes en su corpus y se limitan a las formas *solo*, *solamente* y *únicamente*. Por último, en el discurso político (Sukma 2017) con adverbios como *frankly*, lo que se busca sobre todo es

impactar a los votantes. Por lo tanto, nos encontramos con una función retórica diferente de la que caracteriza el corpus analizado de PAO.

7. CONCLUSIONES

A lo largo de este artículo se ha analizado el papel que desempeñan los marcadores de actitud en un corpus PAO producido por aprendices de español, hablantes de herencia y ELE, con el fin de contribuir a su caracterización.

Los resultados obtenidos nos permiten dar respuesta a las preguntas de investigación formuladas. Se identifica metadiscurso de actitud en los dos corpus, sin embargo, existe una diferencia significativa en su frecuencia, que resulta superior en el corpus de herencia. Esta diferencia puede deberse a una mayor identificación de los hablantes de herencia con los temas de las presentaciones del curso sobre la diáspora latinoamericana en Estados Unidos, puesto que conectan de manera más directa con su experiencia personal. El conocimiento experiencial de los hablantes de herencia sobre el contenido es posible que influya en cómo se manifiestan y, a su vez, se traduzca en un número mayor de marcadores de actitud. Por lo tanto, el dominio del tema da lugar a presentaciones más marcadas en cuanto a la actitud en el grupo de herencia al tratarse de contenido con el que se implican.

Los datos también revelan que ambos grupos presentan un patrón metadiscursivo similar en la distribución de las categorías léxico-gramaticales empleadas (adjetivos, sustantivos, verbos y adverbios) que coinciden, en mayor o menor medida, con investigaciones previas con artículos de investigación, discursos orales públicos y presentaciones orales.

En cuanto a las funciones retóricas identificadas, estas son compartidas por los dos grupos PAO. En términos generales, se puede afirmar que los estudiantes expresan sin ningún tipo de limitación su postura hacia el contenido de su presentación. Resulta interesante que estos resultados estén más próximos a los obtenidos en estudios con discursos públicos que a las conclusiones que se desprenden de investigaciones previas con PAO. Sin duda, las presentaciones analizadas se caracterizan por ser producciones personales y con identidad propia, al estar el material que se presenta sometido a la evaluación constante de los estudiantes.

Las limitaciones de esta investigación vienen motivadas por el tamaño reducido del corpus. Sin embargo, los resultados apuntan a que podría ser interesante estudiar con atención la carga emocional que expresan los marcadores de actitud en conexión con el contenido de las presentaciones. Esto quizás explicaría si la motivación o vinculación a un tema influye en el tipo de marcadores de actitud empleados.

Como conclusión, este estudio constituye una primera aproximación a la descripción del funcionamiento metadiscursivo de la actitud en las PAO y sería interesante que se ampliara en investigaciones futuras a través del análisis de las otras categorías que integran el metadiscurso sobre un corpus más extenso.

REFERENCIAS BIBLIOGRÁFICAS

Ainciburu, M.C., 2012. Recursos para estudiar el español académico a partir de corpus. *Revista Nebrija de Lingüística Aplicada a la Enseñanza de Lenguas*, (12), pp. 209-215.

- Ainciburu, M.C., 2012/2013. Corpus BANKIT. Presentación de Bancos y productos financieros. *Líneas de investigación y muestras de Comunicación Académica* 2. <http://www.comunicacionacademica.com>.
- Ainciburu, M.C., 2018. El estudio de los géneros académicos en la producción textual en ELE. *Journal of Spanish Language Teaching*, 5(2), pp.139-151.
- Ainciburu, M.C. and Villar, C.M., 2013. Coocurrencia de rasgos lingüísticos en la caracterización de la presentación académica oral EFE. In *Revista Nebrija de Lingüística Aplicada. Número especial-Actas de Congreso* (Vol. 13, pp. 30-44).
- Anthony, L., 2014. AntConc (Version 3.4.4) [Software]. Tokyo: Waseda University. <http://www.laurenceanthony.net>.
- Azar, A.S. and Hashim, A., 2019. The Impact of Attitude Markers on Enhancing Evaluation in the Review Article Genre. *GEMA Online® Journal of Language Studies*, 19(1).
- Beke, R., 2005. El metadiscurso interpersonal en artículos de investigación. *Revista signos*, 38(57), pp.7-18.
- Biber, D., 2006. Stance in spoken and written university registers. *Journal of English for Academic Purposes*, 5(2), pp.97-116.
- Dafouz-Milne, E., 2008. The pragmatic role of textual and interpersonal metadiscourse markers in the construction and attainment of persuasion: A cross-linguistic study of newspaper discourse. *Journal of pragmatics*, 40(1), pp.95-113.
- Gómez García, E., 2020. "Relevancia académica de la presentación oral en contextos multiculturales: análisis contrastivo de la interacción en el discurso oral formal de hablantes de herencia de español y estudiantes de ELE estadounidenses" en Licerias, J., Planelles, M. y Foucart, A. (eds.) *Perspectivas actuales en la enseñanza y el aprendizaje de lenguas en contextos multiculturales*. Cizur Menor: Civitas Aranzadi Thomson Reuters, pp.263-297.
- Hyland, K., 2005a. Metadiscourse. London. *Continuum*, 127.
- Hyland, K., 2005b. Stance and engagement: A model of interaction in academic discourse. *Discourse studies*, 7(2), pp.173-192.
- Hyland, K., 2009. Corpus informed discourse analysis: The case of academic engagement. *Academic writing: At the interface of corpus and discourse*, 110, p.128.
- Hyland, K. and Jiang, F., 2016. Change of attitude? A diachronic study of stance. *Written Communication*, 33(3), pp.251-274.
- Kahkesh, M. and Alipour, M., 2017. A comparative study of metadiscourse markers in English and Persian university lectures. *Research in Applied Linguistics*, 8(Proceedings of the Fourth International Conference on Language, Discourse and Pragmatics), pp.125-135.
- Lee, J.J. and Subtirelu, N.C., 2015. Metadiscourse in the classroom: A comparative analysis of EAP lessons and university lectures. *English for Specific Purposes*, 37, pp.52-62.
- MacWhinney, B., 2000. *The CHILDES Project: Tools for Analyzing Talk*. 3rd Edition. Mahwah, NJ: Lawrence Erlbaum Associates.
- Magnuczné Godó, Á., 2012. Are you with me? A metadiscursive analysis of interactive strategies in college students' course presentations. *International Journal of English Studies*, 12(1), pp. 55-78.
- Mameghani, A.A. and Ebrahimi, S.F., 2017. Realization of Attitude and Engagement Markers in Students' Presentations. *International Journal of Applied Linguistics and English Literature*, 6(2), pp.73-77.
- Martin, J., and White, P., 2005. *The Language of Evaluation: Appraisal in English*. London: Palgrave Macmillan.
- Mazidah, E.N., Ni'mah, F. and Ainiyah, M., 2019. Interactional Metadiscourse in Academic Presentation by Students of Fourth Semester of Universitas Qomaruddin. *Edulitics (Education, Literature, and Linguistics) Journal*, 4(2), pp.54-65.
- Mur Dueñas, P., 2010. Attitude markers in business management research articles: A cross-cultural corpus-driven approach. *International Journal of Applied Linguistics*, 20(1), pp.50-72.
- Sukma, B.P., 2017. Interpersonal Metadiscourse Markers as Persuasive Strategies in Barack Obama's 2012 Campaign Speeches. *Aksara*, 29(2), pp.283-292.

- Swales, J.M. and Burke, A., 2003. "It's really fascinating work": differences in evaluative adjectives across academic registers. In *Corpus Analysis* (pp. 1-18). Brill Rodopi.
- Tashi, T. and Suksawas, W., 2018. An analysis of interactional metadiscourse in public speaking: A case study in English speeches of the Prime Minister of Bhutan. *International Journal of Engineering and Technology (UAE)*, 7(4.38), pp.975-979.
- Vergara Padilla, M.Á., 2017. *La influencia de las tipologías textuales en la fluidez. Las presentaciones académicas orales de aprendientes estadounidenses de ele* (Tesis doctoral, Universidad Antonio de Nebrija).
- Vergara Padilla, M.Á. and Mavrou, I., 2019. Análisis macroestructural de presentaciones académicas orales en español L2. *Oralia: Análisis del discurso oral*, 22(1), pp.161-182.
- Villar, C.M., 2011. Las presentaciones académicas orales de los estudiantes alemanes de E/LE. *Revista Nebrija de Lingüística aplicada a la enseñanza de Lenguas*, (10), pp.130-172.
- Villar, C.M., 2011/2013. Corpus PRESAC. PRESentaciones Académicas de fin de curso. Líneas de investigación y muestras de Comunicación Académica 3. <http://www.comunicacionacademica.com/PAOs.html>
- Villar, C.M., 2014. *Las presentaciones académicas orales en E/LE de estudiantes alemanes*. Frankfurt: Peter Lang.

Análisis de Errores en el Corpus CAELE: estudio sobre la concordancia gramatical en el verbo "ser" en aprendientes francófonos y anglosajones

Error Analysis in the CAELE Corpus: study on the grammatical agreement in the verb "ser" in Francophone and Anglo-Saxon learners

Anita Ferreira

Universidad de Concepción
aferreir@udec.cl

Jessica Elejalde

Universidad de Concepción
jelejalde@udec.cl

Lorena Blanco

Universidad Católica de la Santísima Concepción
lblanco@ucsc.cl

RESUMEN

En este artículo se presenta un estudio sobre Análisis de Errores Asistido por el Computador. Para ello, se considera la identificación de los errores y la tendencia de uso erróneo en Español como Lengua Extranjera (ELE) en el nivel de competencia A2 en dos subcorpus del corpus CAELE, un subcorpus de 798 enunciados textuales de aprendientes anglosajones de ELE y otro subcorpus de 625 de aprendientes francófonos. Los objetivos generales son determinar las frecuencias de errores de concordancia gramatical en el verbo ser y comparar las semejanzas y diferencias observadas entre aprendientes francófonos y anglosajones. Los resultados muestran una mayor dificultad para los aprendientes anglosajones en la selección de las categorías gramaticales de género y número en los adjetivos, especialmente en la concordancia de los atributos. En cambio, los aprendientes del francés presentan una frecuencia de errores notoriamente menor en dichos aspectos gramaticales. Los errores podrían atribuirse a la transferencia directa de reglas gramaticales de la L1, incidiendo en la precisión gramatical de los enunciados con el verbo ser. Estos hallazgos contribuyen a dar mayor solidez a las tendencias de la literatura especializada.

Palabras clave: Análisis de Errores Asistido por el Computador, Corpus de Aprendientes, concordancia gramatical, precisión gramatical, Español como Lengua Extranjera.

ABSTRACT

This article presents a study on Computer Aided Error Analysis. For this, the identification of errors and the tendency of misuse in Spanish as a Foreign Language (ELE) at the A2 proficiency level is considered in two subcorpus of the CAELE corpus, a subcorpus of 798 textual sentences of English language learners of ELE and another subcorpus of 625 French-speaking learners. The general objectives are to determine the frequencies of grammatical concordance errors in the verb ser and to compare the similarities and differences observed between French-speaking and Anglo-Saxon learners. The results show greater difficulty for Anglo-Saxon learners in the selection of the grammatical categories of gender and number in the adjectives, especially in the concordance of the

attributes. In contrast, French learners show a noticeably lower frequency of errors in these grammatical aspects. The errors could be attributed to the direct transfer of grammar rules from L1, affecting the grammatical precision of the sentences with the verb to be. These findings contribute to giving more solidity to the trends in the specialized literature.

Keywords: Computer-Aided Error Analysis, Computer Learner Corpora, grammatical agreement, grammatical precision, Spanish as Foreign Language.

1. INTRODUCCIÓN

En la literatura especializada de adquisición de segundas lenguas (ASL), los estudios sobre la temática de los errores causados por transferencia lingüística negativa se han conducido a través de dos enfoques metodológicos, el análisis de errores asistido por el computador (del inglés *Computer Aided Error, CEA*) y el análisis contrastivo de la interlengua (ACI) (Gilquin, 2008; Granger, 1996). Granger (2009) sostiene que el ACI es una técnica heurística que permite descubrir rasgos de la lengua de los aprendientes y luego analizarlos desde una perspectiva de la L2. Los planteamientos actuales van más allá del error, señalando los aciertos, frecuencia de uso, usos infrautilizados, y la sobreutilización de formas de la L2 (Ishikawa, 2013; Sánchez, 2015).

En dicho contexto, el presente estudio se enmarca en las áreas de Corpus de Aprendientes en formato computacional (del inglés, *Computer Learner Corpora, CLC*). Los corpus computacionales fueron originados en la década de 1990, tras el interés por crear y generar recursos que fueran de utilidad para los aprendientes de lenguas. Como resultado, la consolidación de la Lingüística de Corpus y la evolución de la tecnología y de las metodologías de investigación permitieron el uso de dichos corpus como herramientas de investigación, posibilitando el desarrollo de nuevos tipos de estudios tales como los análisis frecuencias de coaparición o de patrones de uso lingüístico por parte de grupos de aprendientes (Granger, 2004, 2012, 2015; Hasselgard y Johansson, 2011; Sánchez, 2015).

Los corpus de aprendientes se construyen a partir de textos auténticos producidos por personas que están aprendiendo una determinada lengua con diferentes lenguas maternas (L1) y distintos grados de conocimiento de la lengua meta. Los textos se recolectan de acuerdo a criterios específicos cumpliendo con condiciones de tipología textual, ámbito geográfico, temporalidad, etc., que permiten que el corpus se constituya en una muestra representativa de una lengua particular o de un subconjunto de la misma (Bautista, 2019). Estos corpus permiten que el investigador detecte las fortalezas y debilidades del grupo lingüístico estudiado, por lo tanto, los resultados se consideran una fuente valiosa de datos lingüísticos (Buyse, 2011).

Las investigaciones basadas en corpus de aprendientes (Campillos, 2014; Granger, 2015 y Lozano, 2015) han contribuido en la descripción de la interlengua relacionada con los aspectos lingüísticos que se aprenden antes que otros, sobre cuáles son más difíciles de aprender y las diferencias con los hablantes nativos (HN), por medio de análisis contrastivos. Los corpus de aprendientes son especialmente útiles a la hora de investigar el proceso de aprendizaje y constituyen una fuente empírica de datos para líneas de

investigación como el CEA (del inglés *Computer aided error analysis*) y los estudios en análisis de interlengua (Cruz Piñol, 2012). Al mismo tiempo, estas descripciones mejoradas de la lengua del aprendiente pueden ser utilizadas en la enseñanza de las L2 (Granger, 2012; Pastor Cesteros, 2004).

Para encauzar la investigación en este ámbito interdisciplinario, hemos estado implementando un corpus de aprendientes de Español como Lengua Extranjera, el cual hemos denominado, CAELE. Este corpus se constituye en una colección abierta que año tras año considera la recolección y el procesamiento de nuevos textos. El conjunto de textos recolectados ha sido anotado y procesado a través del software *Sketch Engine* (Kilgarriff et al., 2004; 2014), configurándose en una colección de 1217 textos producidos en un ambiente presencial en un aula virtual. Los aprendientes, que escribieron dichos textos, provienen de diferentes países, tales como, Francia, Bélgica, Alemania, Sudáfrica, Estados Unidos, Inglaterra, Brasil, Portugal, entre otros. Los estudiantes pertenecen a diferentes áreas de estudio: ingeniería, matemáticas e industria, administración y astrofísica, áreas más relacionadas con las ciencias exactas y el uso del lenguaje matemático. También provienen de las áreas de las humanidades y ciencias sociales.

El CAELE permite observar distintas problemáticas lingüísticas y realizar análisis léxico-gramatical, que puede incluir listas de frecuencias, colocaciones, expresiones multipalabras, palabras clave, n-gramas, entre otros. La problemática que se aborda en el presente artículo se relaciona con las dificultades que se observan en las estructuras nominales y adjetivas con el verbo *ser* en aprendientes de ELE de nivel de competencia A2 con L1 francés e inglés. En general, los estudios relacionados con el verbo *ser* abordan el tema sobre el carácter intercambiable de los verbos copulativos *ser* y *estar* (Baralo, 1998; Guijarro-Fuentes, 2008). La distinción entre el uso del verbo *ser* y *estar* es uno de los temas gramaticales de mayor dificultad para los aprendientes de ELE, no solo para aprendientes francófonos o anglosajones, sino también para estudiantes de muchas otras lenguas. Tanto en el francés como en el inglés sólo existe un verbo *être* y *to be*, respectivamente, para lo que en español tenemos el verbo *ser* y el verbo *estar*.

Otra problemática interesante de explorar y describir es la relativa a la concordancia gramatical en las estructuras de sujeto y predicativas con el verbo *ser*. En efecto, una de las áreas que presenta bastante dificultad para los aprendientes de ELE anglosajones lo constituyen los errores de concordancia (Fleissig, 2009; Méndez y Rojas, 2013; Naranjo, 2009; Paquet, 2017; Prieto, 2011). Como bien lo señala Pastor Cesteros (2001) se debería analizar los problemas de transferencia en la concordancia gramatical, especialmente en los niveles de competencia principiante, para evitar la fosilización de dichos errores. La L1 y el nivel de competencia en la L2 se consideran factores influyentes en la transferencia. En consecuencia, determinar en qué medida la transferencia y el nivel de competencia influyen en los errores de concordancia gramatical es todavía un tema no resuelto que requiere de mayor investigación (Paquet, 2017).

En el presente artículo se muestran los resultados de un estudio descriptivo de carácter mixto, cuantitativo y cualitativo para delimitar los errores más frecuentes cometidos por aprendientes de ELE francófonos y anglosajones de nivel A2 en las estructuras de sujeto y atributivas con el verbo *ser*, así como las principales semejanzas y diferencias observadas entre ambos grupos. El artículo se organiza en las siguientes secciones: En la sección 2, nos referimos a los fundamentos teóricos en materia de la concordancia gramatical y los estudios sobre aprendientes de ELE. En la sección 3, abordamos el estudio sobre un análisis de errores de concordancia gramatical. En la sección

4, presentamos los resultados cuantitativos y cualitativos. Finalmente, presentamos algunas conclusiones sobre los avances y logros obtenidos en esta investigación.

2. FUNDAMENTOS TEÓRICOS

2.1. La concordancia gramatical en español

El Diccionario de la Real Academia Española (2014: 531) define la concordancia gramatical como la "congruencia formal que se establece entre las informaciones flexivas de dos o más palabras relacionadas sintácticamente". A su vez, en el Diccionario Panhispánico de Dudas (2005: 158), se explica la concordancia como "la coincidencia obligada de determinados accidentes gramaticales (género, número y persona) entre distintos elementos variables de la oración". En dicho diccionario se distinguen dos tipos de concordancias:

a) nominal, en la cual el sustantivo establece concordancia de género y número con sus determinantes, tales como artículos y adjetivos. Por ejemplo, "Enriqueta es malintencionada". En esta oración el número y el género del sujeto se corresponden con el atributo nominal.

b) verbal, establecida entre el sujeto y el verbo de una oración. Por ejemplo, "Los alumnos están muy bien". La frase nominal desempeña función sintáctica de sujeto y concuerda en número plural con el verbo.

La problemática de la concordancia se rige por algunas reglas generales. Entre ellas:

1) La coordinación de dos o más sustantivos o pronombres en singular, siempre que cada uno de ellos se refiera a un ente distinto, forma un grupo que concuerda en plural con el adjetivo o el pronombre, o con el verbo del que son sujeto. Por ejemplo, "Los estudios y los viajes son compatibles".

2. La coordinación de dos o más sustantivos o pronombres de diferente género gramatical forma un grupo que concuerda en masculino con el adjetivo o con el pronombre: "En la universidad, las flores y los murales eran coloridos".

La gramática española reconoce el valor de dos propiedades de esta naturaleza que están implicadas en la concordancia. Estas son género y número.

2.1.1. Género gramatical en español

Según el Manual de la Nueva Gramática de la Lengua Española (2010: 23), el género es "una propiedad gramatical de los sustantivos y de algunos nombres que incide en la concordancia con los determinantes, los cuantificadores y los adjetivos o los participios". Montes se refiere a este fenómeno como:

El género es un sistema de clasificación de los nombres en grupos o clases que pueden incluir o no la clase 'sexo'. Además, agrega que la concordancia es la que hace posible dicha clasificación. Cuando se hace alusión al género, el nombre es el que lleva la información gramatical, pero cuando actúa dentro de una expresión debe concordar con las otras categorías gramaticales tales como el adjetivo o el pronombre que lo acompañan (Montes, 1997: 144-145).

2.1.2. Número gramatical en español

El Diccionario de la Real Academia (2014: 1444) define en una de sus acepciones el concepto de número como la "categoría gramatical que en el nombre y el pronombre

expresa la referencia a una o a varias entidades, y que en otras clases de palabras se manifiesta a través de la concordancia". En el Manual de la Nueva Gramática de la Lengua Española este concepto significa:

una propiedad gramatical característica de los sustantivos, los nombres, los adjetivos, los determinantes (en el sentido amplio, que abarca los cuantificadores) y los verbos. Se presenta en dos formas: singular (árbol, quien, grande, este, come) y plural (árboles, quienes, grandes, estos, comen) (RAE, 2010: 35).

Sobre la base de las conceptualizaciones expuestas, se puede señalar que ambas propiedades gramaticales tienen un sistema de valoración dual, ya que distinguen entre el singular y el plural para el caso de número, y entre el masculino y el femenino en el caso del género. Ambas categorías se manifiestan por medio de desinencias o sufijos flexivos que se agregan a la raíz de los sustantivos, adjetivos, entre otros (por ejemplo, o/a-o/e-o/a en chico/chica, cliente/clienta, león/leona para el género).

2.2. Estudios sobre la concordancia gramatical en aprendientes de Español como lengua Extranjera

En los estudios sobre la concordancia basados en el análisis contrastivo en la adquisición de la sintaxis del español pueden destacarse las investigaciones de la evolución de la interlengua en hablantes de lengua nativa inglesa (Liceras, 1983), estudios sobre el análisis de errores en los tiempos de pasado (Fernández, 1997); estudios de la interlengua y análisis de errores de hablantes de L1 alemana (Rakasaeder y Schmidhofer, 2014; Vázquez, 1991), serbocroata (Santos Gargallo, 1993) o de diversas L1 (Fernández, 1997). Asimismo, existen estudios de la interlengua del español en aspectos morfológicos como la formación de palabras (Baralo, 1996) o en otros como los clíticos, las oraciones relativas especificativas o el parámetro del sujeto nulo (Liceras, 1996); y, finalmente, trabajos de lingüística contrastiva del español con lenguas como el portugués, el inglés o el chino (Penadés, 1999). Aun cuando, han sido pocos los estudios que se focalizan en el fenómeno de la concordancia, no obstante, dichos estudios contribuyen en los análisis de la conformación de la interlengua y los procesos que subyacen a este proceso.

En el estudio de Vázquez (1991), se explica que en la producción de la concordancia en el sintagma nominal, el aprendiente activa diversas operaciones mentales. Estas "operaciones cognitivas, por un lado, a partir del conocimiento de las reglas morfológicas de formación de género y número, y operaciones psicológicas, por otro, en la medida en que se presupone un análisis semántico del sustantivo, especialmente en el caso de los colectivos" (Pastor Cesteros, 2001: 10). En dicha explicación Vázquez (1991) considera que este tipo de operaciones ocurren dada la experiencia y el conocimiento que tiene el aprendiente de su L1, lo que podría resultar en apariciones de interferencias cuando no haya equivalencias entre ambas lenguas (Pastor Cesteros, 2001). En este contexto, Vázquez (1991) señala que en la mayoría de los casos los errores de concordancia se pueden explicar en términos de uso de estrategias de simplificación por parte del aprendiente y en su investigación, se refiere a la interferencia de la L1, además de otros factores mencionados.

Por su parte, Fernández (1997) analiza la interlengua de aprendientes de diversas L1, profundiza diversos errores gramaticales cometidos, señalando los problemas de concordancia con mayor frecuencia. El estudio se realizó sobre la base de cuatro grupos de

L1 alemana, japonesa, árabe y francesa, de distintos niveles de adquisición, a continuación se referencia lo indicado por Fernández (1997): 1) frecuencia de errores de concordancia cuando el sustantivo y el elemento que concuerda con este no se sitúan contiguamente, 2) el cambio de género del elemento que no concuerda es más el masculino por el femenino que viceversa, 3) respecto de la concordancia en número, hay preferencia por la forma del singular frente a la de plural; la falta de concordancia se produce mayormente en la categoría de adjetivo; y un factor que la favorece es precisamente la no distinción entre adverbio y adjetivo. Según Fernández (1997), el aumento de errores, en un determinado momento, no es signo de regresión sino de progreso, pues ello implica en el aprendiente un proceso más creativo, de ensayar y practicar atreviéndose a arriesgar más en la nueva lengua.

Con otro enfoque, Pastor Cesteros (2001) lleva a cabo un análisis de un corpus multilingüe y de niveles iniciales, intermedios y avanzados compuesto por pruebas de español como LE. Los resultados en relación con la concordancia gramatical establecen cuatro tipos de errores que explican el avance del nivel de competencia y la conformación de la interlengua. Estos corresponden a: 1) la concordancia entre determinante y sustantivo, 2) sujeto y atributo, 3) adjetivo y sustantivo. Estos hallazgos determinan cómo en el nivel inicial se cometen errores prácticamente en todos los ítems, mientras que en los siguientes niveles va aumentando el dominio de este aspecto, hasta el punto de que en el nivel superior el error aparece sólo en un tercio de los ítems y en un porcentaje mucho menor. En lo referido a la concordancia entre sustantivo y adjetivo (sean estos elementos de un único SN o sujeto y atributo, respectivamente), se especifica que a partir del nivel avanzado desaparecen gradualmente los problemas con adjetivos terminados en -e (y de un modo definitivo ya en el nivel superior). Se concluye que en la concordancia sustantivo/adjetivo (tanto en el seno del SN, como en la relación sujeto/atributo), surgen errores en el nivel inicial, los cuales permanecen en el nivel intermedio en los casos de adjetivos acabados en -e, cuando hay alternancia en la adscripción del determinante y cuando hay inversión de las normas generales respecto al género y se encuentra ante sustantivos masculinos acabados en -a que se hacen, por tanto, concordar en femenino. Este tipo de problemas disminuye considerablemente en el nivel avanzado y prácticamente es inexistente en el superior.

Por su parte Aránzazu (2004) especifica la importancia de estudiar la concordancia de género de los sustantivos inanimados de género no semántico, como *problema*, *mano*, etc., con sus modificadores en aprendientes de ELE anglosajones. La autora señala que la adquisición de la concordancia de los sustantivos problemáticos resulta complicada para cualquier alumno de ELE, cuya L1 carezca de esta categoría morfosintáctica. Sobre la base de dichos resultados Aránzazu (2004) propone la enseñanza de estos aspectos gramaticales centrada en el procesamiento del input y su realce para facilitar el aprendizaje en el ELE.

En estudios más recientes en otras lenguas, como en el caso del análisis de la concordancia en la expresión escrita de aprendientes serbios de ELE (Rubinjoni, 2014), se señala que las causas de los errores podrían residir en la transferencia del comportamiento gramatical de las unidades de una lengua a las de la otra, lo que supone considerar el conocimiento de la lengua y su influencia en la conformación de oraciones y discurso con sentido.

En las investigaciones de Elejalde y Ferreira (2016) y Ferreira y Elejalde (2020b) referidos al análisis de frecuencia de errores de transferencia de la L1 al ELE de aprendientes anglófonos y germanoparlantes las autoras concluyen que la tendencia de la falsa selección de la concordancia gramatical en el atributo de oraciones copulativas y la elección del género

gramatical se corrobora con los hallazgos en el estudio de Pastor Cesteros (2001) y Rakasaeder y Schmidhofer (2014). No obstante, los errores del género se observaron con mayor frecuencia en el alemán, dadas las características del género morfológico en dicha lengua. En esta última se transfieren los rasgos de género propios del alemán, induciendo al error de falsa selección de género gramatical y morfológico tanto en el sintagma nominal como verbal.

Finalmente, en el estudio de Agudelo (2018) enfocado en identificar y describir los errores y causas del verbo *ser* y *estar* en estudiantes francófonos de ELE, se parte de la base de que el uso entre estos dos verbos del español es una dificultad para los estudiantes que poseen un sistema verbal de atributo diferente. En sus hallazgos la investigadora señala con respecto al verbo *ser* en el nivel A2 que, la mayoría de los errores cometidos por este grupo de aprendientes está presente en las estructuras conformadas por verbo + atributo adjetivo (que solo van con *ser*). Afirma también que en este nivel de competencia del español la concordancia en la conjugación constituye una dificultad lingüística y no la selección entre el verbo *ser* y *estar*. En relación con las causas, se atribuye la confusión morfológica, definida como los errores de conjugación en los cuales no hay concordancia entre la forma y el sujeto. Por ejemplo, "*¿Dónde está ellas?*"

3. EL ESTUDIO

El diseño del estudio que se presenta a continuación es de carácter descriptivo con un enfoque de análisis de datos mixto: cuantitativo y cualitativo. En cuanto a los datos de corte cuantitativo, permitirán identificar las frecuencias y sistematicidad de los errores, así como también determinar las tendencias de uso de acuerdo con las funciones gramaticales o sintácticas que correspondan. Con respecto a los datos cualitativos, estos proporcionarán las explicaciones sobre el posible origen de los errores según los análisis de la influencia directa de la L1. Se busca avanzar en responder la pregunta de investigación sobre ¿Cuáles son las principales dificultades que manifiestan los alumnos universitarios francófonos y anglosajones de ELE en tareas de escritura académica?

El enfoque metodológico de la investigación se enmarca en los procedimientos de la lingüística de corpus, el análisis de errores y el análisis contrastivo de interlengua. Los procedimientos se basan en los lineamientos propuestos en el Proyecto Fondecyt No.1180974. A continuación, presentamos los objetivos generales y específicos del estudio:

Objetivos generales

1. Determinar las frecuencias de errores de concordancia gramatical de género y número en estructuras oracionales con el verbo *ser* a nivel de sujeto, sujeto verbo y atributos nominales y adjetivales, producidas por aprendientes francófonos y anglosajones de nivel A2 de ELE.
2. Comparar las principales diferencias y semejanzas observadas en los errores de aprendientes francófonos y anglosajones en los contextos de estudio con el verbo *ser*.

Objetivos específicos

- 1.1 Determinar y explicar las mayores frecuencias de errores en la concordancia gramatical en estructuras oracionales con el verbo *ser* en aprendientes francófonos.
- 1.2 Determinar y explicar las mayores frecuencias de errores en la concordancia gramatical en estructuras oracionales con el verbo *ser* en aprendientes anglosajones.

3.1. El Corpus de Aprendientes de Español como Lengua Extranjera, CAELE

El CAELE es una colección de 1217 textos producidos en un ambiente presencial a través de un aula virtual, procesados a través del software *Sketch Engine*. Estos textos han sido recolectados durante los periodos de clases entre los años 2014 y 2019 y se distribuyen en los niveles de competencia del ELE, A2 y B1. CAELE está conformado por 265.362 palabras, por 296.136 tokens y por 14.902 oraciones. La recolección de los textos se realizó durante los cursos de ELE ofrecidos semestralmente cada año en una universidad chilena. La extensión es de 150 a 200 palabras (A2) y de 250 a 300 (B1). Las temáticas de escritura se determinaron mediante un análisis de necesidades y consideraron temas del ámbito de español para fines académicos científicos y humanísticos: literatura, ciencia y tecnología, cultura latinoamericana, etc. Las secuencias textuales recolectadas en los textos responden a la estructura narrativa, descriptiva y argumentativa (Ferreira y Elejalde, 2019). Para ratificar los niveles de competencia que los aprendientes declaran en una encuesta de registro, se aplicó la prueba de multinivel de competencia (Ferreira, 2016). Los aprendientes que escribieron los textos son 210 sujetos y provienen de los programas de intercambio de una universidad chilena en el marco de la movilidad estudiantil.

Para este estudio se crearon 2 subcorpus de enunciados textuales a partir del CAELE, relacionados con el objeto de estudio de esta investigación. Un subcorpus que está constituido por 625 enunciados textuales con el verbo *ser* producidos por aprendientes francófonos y otro subcorpus está constituido por 798 enunciados textuales producidos por aprendientes anglosajones.

A continuación se explican los procedimientos de extracción de los subcorpus a partir del corpus CAELE y los procesos de anotación a partir de la taxonomía de errores y procesamiento de los datos de dichos corpus.

3.2. Implementación de los subcorpus

El corpus CAELE fue implementado en el software *Sketch Engine*. Para extraer los 2 subcorpus francófono y anglosajón se ingresó la búsqueda de concordancias con el verbo *ser* en la herramienta *Word Sketch*. En seguida, se seleccionaron los atributos más representativos (nominales y adjetivos) para finalmente filtrar los resultados por la L1 de los aprendientes (francés e inglés) y nivel de competencia A2. En la Figura 1 se muestra parte del resultado de la búsqueda en la interfaz concordancia, correspondiente al atributo nominal. Se ilustra la interfaz con la que funciona el análisis del software *Sketch Engine*. Los enunciados pueden analizarse de acuerdo con su concordancia, los tipos de palabras, el contexto de la oración y el número de frecuencias o apariciones en las consultas por cada caso.

CONCORDANCIA CAELE

cql ser (objects of "ser") 349 > filtro A2,INGLES 349 (1178,51 por millón) X

Contexto izquierdo KWIC Contexto derecho

1	<input type="checkbox"/>	S100A2PRU112019...	aire libre que estoy pasando en Chile . </s> <s>	Es el invierno en los Estados Unidos , y en mi esta
2	<input type="checkbox"/>	S100A2T112019.t...	comida . </s><s> Una persona que yo admiro mucho	es mi madre . </s><s> Mi madre trabaja en el Depa
3	<input type="checkbox"/>	S100A2T112019.t...	madre trabaja en el Departamento de Justicia . </s> <s>	Es un trabajo muy complejo y estresante . </s><s> Aunque el
4	<input type="checkbox"/>	S100A2T112019.t...	os demás antes de piensa en ella misma . </s><s> Ha	sido un buen ejemplo para mí en toda mi vida .
5	<input type="checkbox"/>	S100A2T112019.t...	vida . </s><s> Ella me ha enseñado que puedo	ser cualquier persona que quiero ser si sigo mis sueñ

Figura 1. Concordancias con el verbo *ser* en *Sketch Engine*

Por otra parte, en el Gráfico 1 se muestra la conformación de los subcorpus del CAELE para el análisis de los errores de concordancia gramatical. Cada subcorpus de enunciados se subdivide en dos tipos de atributo, el nominal y el adjetivo. Respecto de las oraciones con atributo nominal el total corresponde a 663 frecuencias, de las cuales 349 corresponden a la producción de hablantes nativos de inglés y 314 de hablantes nativos de francés. En relación con las oraciones con atributo adjetivo se refleja un total de 760 frecuencias distribuidas en 449 frecuencias de hablantes nativos de inglés y 311 de francés. En total se consideraron para este análisis 1423 oraciones con concordancia gramatical de atributo con el verbo copulativo *ser*.

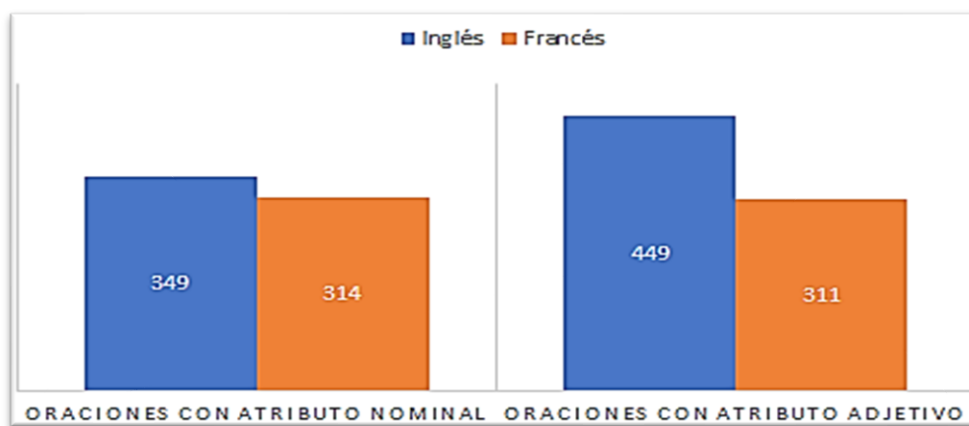


Gráfico 1. Distribución por lengua materna

En el Gráfico 1 se puede observar la distribución de las oraciones con atributo nominal y con atributo adjetivo en las lenguas en estudio. Se observa que en inglés hay mayor uso de oraciones con atributo adjetivo (449 apariciones) respecto de las oraciones con atributo nominal con 349 usos, mientras que en francés se usa equilibradamente ambos tipos de oraciones con un total de 314 de tipo nominal y 311 de tipo atributo adjetivo.

3.3. La anotación

Respecto del análisis y clasificación de los errores se consideró la taxonomía propuesta para el etiquetado de errores del criterio etiológico (Ferreira y Elejalde, 2020a). Para la identificación de los errores de concordancia gramatical se incluyó 9 tipos de etiquetas que se especifican en la Tabla 1. Las etiquetas se forman por dos bases principales, la primera (FS) nomenclatura que responde a *falsa selección*, en términos de la modificación de la lengua objeto de estudio y la segunda (CG) nomenclatura que significa *concordancia gramatical* para referirse a la temática central de este estudio. Las otras nomenclaturas dependen del tipo de concordancia en estudio. Las etiquetas corresponden a: 1) FS-CG-GEN, falsa selección de la concordancia gramatical de género en el sujeto, 2) FS-CG-NUM, falsa selección de la concordancia gramatical de número, 3) FS-CG-SUJ-VSER, falsa selección de la concordancia gramatical entre el sujeto y el verbo ser, 4) FS-CG-ATRIB-GEN-FA falsa selección de la concordancia gramatical del atributo en el género frase adjetival, 5) FS-CG-ATRIB-NUM-FA falsa selección de la concordancia gramatical del atributo en el número frase adjetival, 6) FS-CG-ATRIB-GEN-FN falsa selección de la concordancia gramatical del atributo en el género frase nominal, 7) FS-CG-ATRIB-NUM-FN falsa selección de la concordancia gramatical del atributo en el número frase nominal, 8) FS-CG-GEN falsa selección de la concordancia gramatical en el género en el atributo interno, 9) FS-CG-NUM falsa selección de la concordancia gramatical en el número en el atributo interno.

Tipos CG	Etiqueta	Descripción
1) Sujeto	FS-CG-GEN	Errores de concordancia gramatical de género y de número en la frase nominal que funciona como sujeto de la oración con verbo <i>ser</i>
	FS-CG-NUM	
2) Sujeto-predicado	FS-CG-SUJ-VSER	Error gramatical de concordancia gramatical de número entre el sujeto y el verbo <i>ser</i>
3) Atributo CG	FS-CG-ATRIB-GEN-FA	Errores en el atributo frase nominal o en el atributo frase adjetivo que generan falta de concordancia gramatical con el verbo <i>ser</i> o con el sujeto del enunciado.
	FS-CG-ATRIB-NUM-FA	
	FS-CG-ATRIB-GEN-FN	
	FS-CG-ATRIB-NUM-FN	
4) Atributo interno	FS-CG-GEN	Errores de concordancia gramatical de género y número a nivel interno en el atributo ya sea frase adjetival o nominal.
	FS-CG-NUM	

Tabla 1. Especificación de la taxonomía y las etiquetas para identificación y clasificación de errores (Ferreira y Elejalde, 2020b)

Como se observa en la Tabla 1, de acuerdo con la problemática de este estudio se delimitaron cuatro tipos de concordancia gramatical para el análisis, las que corresponden a los siguientes ejemplos:

(1) Tipo 1. La concordancia entre los elementos del sintagma nominal en función de sujeto de la oración copulativa. Un ejemplo de este caso corresponde a: "*Hay muchas tipos de música que existe en Latino América, algunos ejemplos **son** la salsa, la bachata y la cumbia*" (S43A2T4I2016)¹ y "*Me gusta la comida del país. Los paisaje **son** muy captivo y magnifico*" (S66A2PRUMF2018).

En los enunciados puede observarse la falta de concordancia en el sujeto, tanto a nivel de género "*algunos ejemplos*" como en el número "*los paisaje*".

(2) Tipo 2. La concordancia gramatical de número entre el sujeto y el verbo *ser*. Los ejemplos para este caso son: "*En el centro hay muchas departamentos de gobierno y muchas edificios que es importante de historia de Chile*" (S51A2POSDIFI2016).

En el enunciado se muestra la falta de concordancia gramatical entre el sujeto pronombre relativo (que) el cual se refiere a "muchas edificios" y la forma verbal del verbo *ser* "es".

(3) Tipo 3. La concordancia gramatical de género y número del atributo respecto del sujeto. Los ejemplos son: "*En mi país la música folclórica **es** sencillo y en general hay solo un cantante*" (S11A2POSINMI2014) y "*Todas las personas **son** diferente por eso hay personas a que gustan la música folclórica*" (S19A2PREF2015).

En los enunciados se ilustra la falta de concordancia gramatical entre el sujeto y el atributo del verbo *ser*. Es decir, hay una falsa selección del género entre "la música" y "sencillo". Además, se observa falta de concordancia de número entre "las personas" y "diferente".

(4) Tipo 4. La concordancia gramatical del atributo considerando el sintagma nominal o adjetival respecto de su función como atributo. Por ejemplo: "*Las ferias con soporte local **son** buenas lugares para la gente que quieren vivir con las ventas de artesanía*" (S11A2T1I2014)

En este último ejemplo, se muestra la falta de concordancia gramatical de género entre el sujeto y el atributo de adjetivo en "las ferias" y "buenas lugares".

Con la delimitación de estos cuatro tipos de concordancia gramatical se busca determinar no sólo la concordancia relacionada con el sujeto y el atributo, sino también las concordancias internas a nivel del sujeto y sujeto verbo *ser*. De esta forma, se cautela la revisión de la concordancia que podría afectar el sentido parcial o completo de la oración.

3.4. Procesamiento de los datos del subcorpus CAELE con el verbo *ser*

Para realizar el análisis de errores, como ya se señaló, se extrajo a través del software *Sketch Engine* un subcorpus de concordancias textuales conformado por 780 concordancias de aprendientes anglófonos y 625 concordancias de francófonos. Una vez obtenido los

subcorpus, se procedió a la identificación y clasificación de los errores. Para ello, se llevó a cabo de forma manual un registro de los errores que se encontraban y las frecuencias. Este registro se alojó en Excel para el posterior tratamiento de los datos.

Con los registros de las frecuencias en el Excel se calculó el total de errores y las frecuencias relativas por cada caso; al mismo tiempo, se determinaron los errores más frecuentes y cuáles de estos tenían mayor incidencia en la conformación de las oraciones copulativas del verbo *ser*. Respecto a la distribución de los enunciados textuales para el análisis, se consideró la variable *investigador*, por lo que estos fueron distribuidos y asignados de manera aleatoria para ser analizados por el equipo de investigación.

4. RESULTADOS Y DISCUSIÓN

De acuerdo con nuestro Objetivo General 1 se determinó los errores más frecuentes para establecer las mayores dificultades en el uso de la concordancia gramatical observadas en los enunciados de los aprendientes. Para ello, se realizó el conteo de los errores identificados en los cuatro tipos especificados anteriormente. Para responder a los objetivos específicos 1 y 2 se distribuyeron los resultados generales acorde con la variable L1 francés e inglés.

4.1. Errores más frecuentes en la concordancia gramatical con el verbo *ser*

El total de errores identificados fue de 282 errores en ambas lenguas en estudio, francófonos y anglosajones. En la Tabla 2 se presenta el total de los errores en los cuatro tipos de concordancias analizadas. En el Tipo 1, sujeto, se encontraron 55 errores en la concordancia gramatical de género y número en el sintagma nominal del sujeto. Seguido de este se observa el Tipo 2 (sujeto-predicado) con un total de 35 errores de concordancia entre el sujeto y el verbo *ser*. En el Tipo 3, se hallaron 136 errores en la concordancia gramatical entre el atributo y el sujeto de la oración copulativa. Por último, en el Tipo 4 hay un total de 55 errores en la concordancia gramatical de género y número del atributo a nivel del sintagma nominal o adjetival. De esta manera, puede observarse con precisión qué ocurre tanto a nivel del sintagma como la estructura completa de la oración copulativa.

Por otra parte, el análisis de estas frecuencias arroja como error más frecuente, con 80 errores, la falsa selección de concordancia gramatical de número en el atributo en la frase adjetival (*FS-CG-TRIB-NUM-FA*), seguido de la falsa selección de concordancia gramatical de género en el atributo interno (*FS-CG-GEN*) con 48 errores y la falsa selección de concordancia gramatical de género presente en el sujeto de la oración (*FS-CG-GEN*) con 42 errores. Estas frecuencias reflejan cuáles son las problemáticas más relevantes en el CAELE en cuanto a la concordancia gramatical en oraciones copulativas con el verbo *ser*.

No. Tipo	Tipos CG	Etiqueta	Total	Fr. %
1	Sujeto	FS-CG-GEN	42	15%
		FS-CG-NUM	13	5%
2	Suj-pred	FS-CG-SUJ-VSER	35	12%
3	Atributo	FS-CG-TRIB-GEN-FA	28	10%

	Concordancia Gramatical en FAdj y FN	FS-CG-TRIB-GEN-FN	12	4%
		FS-CG-TRIB-GEN-FA	3	3%
		FS-CG-TRIB-GEN-FN	3	3%
		FS-CG-TRIB-GEN-FN	2	2%
4	Atributo interno	FS-CG-GEN	17	18%
		FS-CG-NUM	3	3%
Total de todos los tipos			282	100%

Tabla 2. Total de frecuencias de errores según los 4 tipos de concordancias gramaticales en estudio

4.2. Frecuencia de errores de concordancia gramatical en aprendientes francófonos

De acuerdo con el Objetivo Específico 1, en esta subsección se presentan los errores más frecuentes de los aprendientes de L1 francés y su distribución según los tipos de concordancia en estudio (ver Tabla 3).

El error más frecuente corresponde a la falsa selección en la concordancia gramatical de número en la frase adjetival con un total de 30 apariciones (32%) en el Tipo 3 (atributo CG). Seguido se encuentran dos errores con una frecuencia menor, la falsa selección de la concordancia gramatical entre el sujeto y el verbo copulativo con un total de 18 errores (19%) en el Tipo 2 y la falsa concordancia en el género dentro del sujeto con 12 errores (13%) en el Tipo 1.

De acuerdo con lo señalado por Doquin de Saint Preux y Sáez (2014) la concordancia de la categoría gramatical de número es uno de los errores más frecuentes cometidos por aprendientes francófonos de nivel de competencia avanzada. En esta categoría gramatical los errores se podrían considerar errores de transferencia dado que los francófonos comenten de manera recurrente dicho tipo de error en su propio idioma. Según Doquin de Saint Preux y Sáez (2014) la concordancia de número, en francés, presenta el mismo mecanismo que en español, sin embargo, al no pronunciarse la "s" final del sustantivo tienden a omitirla, también de manera equivocada, en la lengua escrita.

No.Tipo	Tipos CG	Etiqueta	L1 Francés	Fr. %
1	Sujeto	FS-CG-GEN	12	13%
		FS-CG-NUM	6	6%
2	Suj-pred	FS-CG-SUJ-VSER	18	19%
3	Atributo CG	FS-CG-TRIB-GEN-FA	3	3%
		FS-CG-TRIB-GEN-FA	30	32%
		FS-CG-TRIB-GEN-FN	3	3%
		FS-CG-TRIB-GEN-FN	2	2%
4	Atributo interno	FS-CG-GEN	17	18%
		FS-CG-NUM	3	3%
Total de errores			94	100%

Tabla 3. Frecuencia de errores en la concordancia gramatical de oraciones copulativas con el verbo *ser* en aprendientes francófonos

Otro error frecuente es la falsa selección del género gramatical en la composición interna del atributo de la frase nominal con un total de 17 errores (18%) en el Tipo 4. Este tipo de errores en la categoría de género gramatical se pueden explicar por la transferencia desde la L1, ya que los aprendientes le adjudican al sustantivo el género que tiene dicha palabra en francés (Doquin de Saint Preux y Sáez, 2014).

A continuación, se presentan ejemplos y la explicación de los casos más frecuentes en los aprendientes francófonos:

- 1) Ejemplo Tipo 1, FS-CG-GEN: "El segundo lugar que tienes absolutamente que ver es la ciudad de Valparaiso. *Este ciudad es* el puerto el más importante del Chile y se puede encontrar muy grande barcos" (S52A2PRUMF2017).

En este ejemplo no hay concordancia de género gramatical a nivel de la frase nominal del sujeto, probablemente debido al desconocimiento del uso del adjetivo determinante "este".

- 2) Ejemplo Tipo 2, FS-CG-SUJ-VSER: "*Mi rutina en Chile son* muy diferentes de mi vida" (S90A2PRU2F2019).

En este ejemplo hay una falsa selección del número gramatical del verbo, el cual no concuerda con el sujeto que está en singular. Según afirma Agudelo (2018) estos errores entre la concordancia del verbo se podrían explicar en términos de confusión morfológica donde el aprendiente no identifica claramente la categoría de número en el verbo y su concordancia con el atributo.

- 3) Ejemplo Tipo 3, FS-CG-TRIB-NUM-FA "*Los autos electricos o que funcionan con agua son* muy caro" (S58A2PRU2F2017).

En el ejemplo 3, se observa la falsa selección de la concordancia gramatical del número en la frase adjetiva del atributo. Dicho atributo no concuerda con el sustantivo "*los autos*". Este error podría deberse a la distancia entre la frase adjetiva "muy caro" y la frase nominal referenciada por el pronombre relativo "que". El aprendiente se confunde por la subordinación. Estas confusiones igualmente ocurren en el español cuando hay una distancia mayor entre el verbo *ser* y los predicativos. También se observa este fenómeno en oraciones con pronombres relativos y referenciales en su función anafórica.

- 4) Ejemplo Tipo 4, FS-CG-GEN "*La pintura describe el mesclar de los racas en sur america . es* un pintura de acrylico" (S19A2T3F2015).

Finalmente, en el ejemplo 4 hay una falsa selección del género en la función interna del atributo "un pintura de acrylico". Este error puede deberse al desconocimiento del género gramatical del sustantivo induciendo al error. El desconocimiento del género de los sustantivos representa uno de los aspectos más problemáticos en el aprendizaje del español como LE (Pastor Cesteros, 2001).

4.3. Frecuencia de errores de concordancia gramatical en aprendientes anglosajones

De acuerdo con el Objetivo Específico 2, en esta subsección se presentan los errores más frecuentes de los aprendientes de L1 inglés y su distribución acorde con los tipos de concordancia en estudio (ver Tabla 4). Los errores más frecuentes identificados de mayor a menor en la Tabla 5 se distribuyen de la siguiente forma: (1) del Tipo 3, falsa selección de concordancia gramatical de número en la frase adjetiva del atributo con un total de 50 errores (27% del total de los errores); (2) del Tipo 4, la falsa selección de género en la conformación interna del atributo con un total de 31 errores (16%); (3) del Tipo 1, la falsa selección del género en la concordancia interna del sujeto con un total de 30 errores (16%); (4) del Tipo 3, la falsa selección de concordancia gramatical de género con un total de 25 errores (13%) en la frase adjetiva del atributo, y (5) del Tipo 3, la falsa selección de la concordancia gramatical de número en la frase nominal del atributo de la oración copulativa con un total de 15 errores (8%).

No. Tipo	Tipos CG	Etiqueta	L1 Inglés	Fr. %
1	1.Sujeto	FS-CG-GEN	30	16%
		FS-CG-NUM	7	4%
2	2.Suj-pred	FS-CG-SUJ-VSER	17	9%
3	3.Atributo CG	FS-CG-ATRIB-GEN-FA	25	13%
		FS-CG-ATRIB-NUM-FA	50	27%
		FS-CG-ATRIB-GEN-FN	9	5%
		FS-CG-ATRIB-NUM-FN	15	8%
4	4.Atributo interno	FS-CG-GEN	31	16%
		FS-CG-NUM	4	2%
Total de errores			188	100%

Tabla 4. Frecuencia de errores en la concordancia gramatical de oraciones copulativas con el verbo *ser*, aprendientes anglosajones

A continuación, se presentan algunos ejemplos con su respectiva explicación:

- 1) Ejemplo del Tipo 1, FS-CG-GEN: "*La ambiente es bueno y personas necesita reservado una mesa*" (S51A2T1I2016).

Se observa una falsa selección del género a nivel del sujeto, "la ambiente". Este tipo de error se puede explicar también (al igual que en el caso de los aprendientes francófonos) por un desconocimiento del género del sustantivo. Errores muy frecuentes según la literatura especializada.

- 2) Ejemplo Tipo 3, FS-CG-ATRIB-NUM-FA: "*Los otros estudiantes de intercambio son muy amable y nosotros trabajamos junto para hacer muchas cosas*" (S76A2PRUMI2018).

En este ejemplo se evidencia la falta de concordancia en el número gramatical entre el sujeto y el atributo, "*los otros estudiantes son muy amable*". En inglés los adjetivos no tienen rasgos morfológicos de género y número por lo que se podría interpretar que el

aprendiente, en este caso, transfiere esta regla gramatical al español induciendo al error en la concordancia.

- 3) Ejemplo Tipo 3, FS-CG-ATRIB-GEN-FA: "El Gran Cañon **es** muy bonita y gente de todos lugares viene para verlo" (S40A2PREI2016)

En este ejemplo hay una falsa selección del género gramatical en el atributo induciendo al error en la concordancia con el sujeto, "El Gran Cañón es muy bonita". Esto podría deberse al desconocimiento del género del sustantivo o bien a las estrategias intralingüísticas que operan en la interlengua del aprendiente. Es decir, muchas de las producciones de los aprendientes de lenguas se enfocan en la reducción o invención, la generalización, en la evasión o en la superproducción. En este caso, podría explicarse en términos de reducción de la morfología del adjetivo, induciendo al error en la falta de concordancia gramatical.

- 4) Ejemplo Tipo 3, FS-CG-ATRIB-GEN-FN: "*La artesanía es una producto que es*" (S25A2T1I2015)

Finalmente, en el ejemplo 4 se presenta un error en la concordancia de género gramatical del atributo. Al parecer, el sujeto intenta concordar el atributo nominal "una producto" con el sujeto "la artesanía". Este tipo de error podría explicarse en términos de la estrategia interlingüística de generalización de las reglas que resultan de fácil comprensión para el aprendiente.

En síntesis: a diferencia de los aprendientes francófonos, los anglosajones tienden a transferir las estructuras lingüísticas de su L1 dada la distancia entre ambas lenguas. La explicación en los errores de los francófonos podría estar relacionada con la transferencia del sistema lingüístico que se asemeja o se acerca a la conformación del español. Estas dos fuentes del error muestran la interferencia del conocimiento del sistema lingüístico de la L1 como estrategia de aprendizaje para sortear con éxito los vacíos de aprendizaje o conocimiento lingüístico a nivel del género gramatical de los sustantivos. Los resultados aquí encontrados guardan relación con estudios de transferencia, reflejándose en su mayoría tendencias similares en el uso incorrecto de la concordancia gramatical (Agudelo, 2018; Doquin de Saint Preux y Sáez, 2014; Elejalde y Ferreira, 2016; Ferreira y Elejalde, 2017; Pasquet, 2017; Pastor Cesteros, 2001).

4.3. Comparación de errores de concordancia entre aprendientes francófonos y anglosajones

En cuanto al Objetivo General 2, esto es, "comparar las principales diferencias y semejanzas observadas en los errores de aprendientes francófonos y anglosajones en los contextos de estudio con el verbo SER", a continuación se presenta los resultados de las frecuencias por cada L1, considerando la relación entre la frecuencia relativa (respecto del número de errores por lengua) y la frecuencia aumentada (referida a la proporción entre los errores de todo el corpus y los errores encontrados en cada lengua). Para ello, también se ha intentado identificar las posibles fuentes que inducen al error respecto de la conformación del sistema lingüístico tanto del inglés como del francés. El vasto número de adjetivos que existen en español, el aspecto semántico de los mismos, la pragmática y la

variedad geográfica son factores que hacen complejo el campo atributivo y contribuyen a la dificultad en el uso del verbo *ser* (Agudelo, 2018).

Tipos CG	Etiqueta	L1 Francés	Fr. %	Fr. Aum	L1 Inglés	Fr. %	Fr. Aum
1.Sujeto	FS-CG-GEN	12	13%	4%	30	16%	11%
	FS-CG-NUM	6	6%	2%	7	4%	2%
2.Suj-pred	FS-CG-SUJ-VSER	18	19%	6%	17	9%	6%
3.Atributo CG	FS-CG-ATRIB-GEN-FA	3	3%	1%	25	13%	9%
	FS-CG-ATRIB-NUM-FA	30	32%	11%	50	27%	18%
	FS-CG-ATRIB-GEN-FN	3	3%	1%	9	5%	3%
	FS-CG-ATRIB-NUM-FN	2	2%	1%	15	8%	5%
4.Atributo interno	FS-CG-GEN	17	18%	6%	31	16%	11%
	FS-CG-NUM	3	3%	1%	4	2%	1%
Total de errores		94	100%	33%	188	100%	67%

Tabla 5. Frecuencia de errores por L1 en estudio (Francés e inglés)

En la Tabla 5 se puede ver que en los aprendientes de L1 francés hubo un total de 94 errores que corresponde al 33% en relación con el 100% del total de los errores. Por el contrario, la producción de errores de los aprendientes de L1 inglés arrojó un total de 188 errores lo que representa el 67% en relación con el 100% del total de los errores. Este primer análisis refleja que hay una notoria diferencia en cuanto a los porcentajes de errores, evidenciando que se presentan mayores problemáticas de concordancia en la producción de aprendientes anglosajones.

En cuanto al porcentaje de errores en el francés, puede inferirse que las similitudes entre esta lengua y el español favorecen al aprendizaje de este último.

Además, puede observarse que en cada lengua los errores más frecuentes corresponden a:

- 1) En L1 inglés se identifican mayores frecuencias en 4 tipos de errores: (1) errores de concordancia de falsa selección en el atributo en el número en la frase adjetiva con 50/188 errores (Tipo 1). (2) Errores de falsa selección del género en la frase interna del atributo con 31/188 errores (Tipo 4). (3) Errores de concordancia de género gramatical a nivel interno de la FN del sujeto con 30/188 errores (Tipo 1) y por último (4) Errores en la concordancia de género gramatical entre sujeto y el atributo con 25/188 errores (Tipo 3).
- 2) En la L1 francés, puede constatarse tres errores más frecuentes: (1) la falsa selección de número en la concordancia del atributo y el Sujeto con 30/94 errores (Tipo 3). (2) La falsa selección de número del sujeto y predicado del verbo *ser* con un total de 18/94 errores (Tipo 2). Y por último, (3) el error referido a la concordancia de género gramatical a nivel interno del atributo con 17/94 errores (Tipo 4).

Con estos resultados obtenidos, se puede inferir que los errores más frecuentes corresponden al desconocimiento del género gramatical del sustantivo en español, lo que induce al error de concordancia. Asimismo, podría explicarse que en el inglés los errores son más frecuentes en cuanto al género dado que esta lengua no hace una distinción entre el género y el número de los sustantivos, lo que supondría una dificultad mayor para los aprendientes anglosajones (Elejalde y Ferreira, 2016). A diferencia en el francés, algunos de los errores de género y número se reflejan porque muchos de estos no coinciden con el género ni el número en francés y en español, lo que podría determinarse como una interferencia directa.

Los hallazgos aquí encontrados pueden ser corroborados con otros estudios respecto de la complejidad que tienen los rasgos morfológicos y sintácticos del género y número en el español y la distancia de esta lengua con lenguas que no pertenecen al grupo romance (Agudelo, 2018; Elejalde y Ferreira, 2016; Paquet, 2017). Por su parte, Bond et al. (2011) examinaron la influencia de la L1 a partir de un grupo de estudiantes de L1 inglés de nivel competencia inicial y delimitaron que tres tipos de concordancias: de número sujeto verbo, de número sustantivo adjetivo y de género sustantivo adjetivo. Los resultados mostraron mayor sensibilidad de los aprendientes para percibir errores de concordancia de número que de género.

Para una mejor visualización de los resultados logrados con respecto al Objetivo General 2 y poder observar las principales diferencias y semejanzas en los errores en ELE según las L1 en estudio se presentan los Gráficos 1 y 2, que reflejan las tendencias de uso erróneo en ambos grupos de aprendientes.

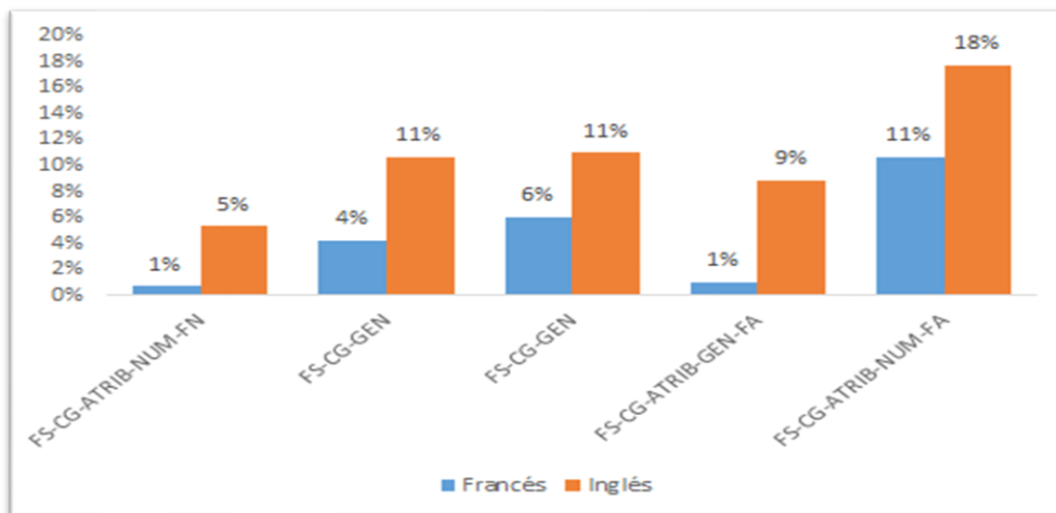


Gráfico 3. Diferencias en las frecuencias de errores entre los francófonos y anglosajones

En el Gráfico 3, se constata las principales diferencias entre los 5 errores más frecuentes del análisis de la concordancia gramatical en los enunciados textuales con el verbo *ser*. Según la distribución de mayor a menor de los porcentajes de errores se visualiza (1) la falsa selección del número gramatical en la concordancia de atributo en la frase adjetiva con un 18% en L1 inglés y 11% en L1 francés, reflejando 7 puntos de diferencia. (2) El error de falsa selección de la concordancia de género gramatical en la frase adjetiva

con un 1% de L1 francés y un 9% de la L1 inglés, con una diferencia de 8 puntos. (3) El error de falsa selección de concordancia gramatical de género a nivel interno en el atributo con un 6% en L1 francés y un 11% en L1 inglés con una diferencia de 5 puntos. (4) La falsa selección del género en la concordancia gramatical a nivel interno del sujeto con un 4% en L1 francés y con un 11% en L1 inglés, marcando una diferencia de 7 puntos. Finalmente, (5) la falsa selección de la concordancia gramatical de número en el atributo de la frase nominal con un 1% en el francés y un 5% en el inglés (4 puntos de diferencia).

Como explicación de estas diferencias, se puede sostener que para los aprendientes anglosajones hay mayores dificultades en la selección del género gramatical y de número en los adjetivos, especialmente en la concordancia de los atributos, dado que, en inglés, como ya se ha dicho, la estructura de los sintagmas nominales no concuerda ni en género ni en número con el sustantivo. Es decir, que estos errores podrían atribuirse a la transferencia directa de estas reglas gramaticales, afectando la precisión gramatical de la oración copulativa del verbo *ser*. Respecto del francés, no se observa que estas problemáticas sean muy frecuentes en las oraciones, dado que esta lengua tiene rasgos morfológicos de género y número que coinciden con las reglas del español (Elejalde y Ferreira, 2016; Agudelo, 2018). No obstante, existe otra posible explicación en la cual se le atribuye la distancia lineal o sintáctica que expone que, a mayor distancia entre los elementos de una oración, la concordancia podría presentar dificultades (Paquet, 2017). Si se remite a los ejemplos anteriores, muchas de las oraciones tienen varios modificadores del sujeto que podrían afectar la concordancia tanto a nivel de sintagma nominal o adjetival como el sujeto y el atributo.

En cuanto a las similitudes en los resultados encontrados entre las frecuencias de errores en ELE, el Gráfico 4 muestra las principales estructuras que fueron menos problemáticas para ambos grupos de aprendientes. Esto es: 1) la falsa selección del número gramatical en la conformación interna del atributo con un porcentaje de un 1%. 2) La falsa selección de la concordancia de número gramatical en el sujeto con un porcentaje de 2% en ambas L1. 3) La falsa selección de la concordancia de género gramatical en la frase nominal del atributo con un 3% en L1 inglés y un 1% en L1 francés y, finalmente, 4) la falsa selección de la concordancia gramatical entre el sujeto y el verbo *ser* que en ambas L1 corresponde al 6%.

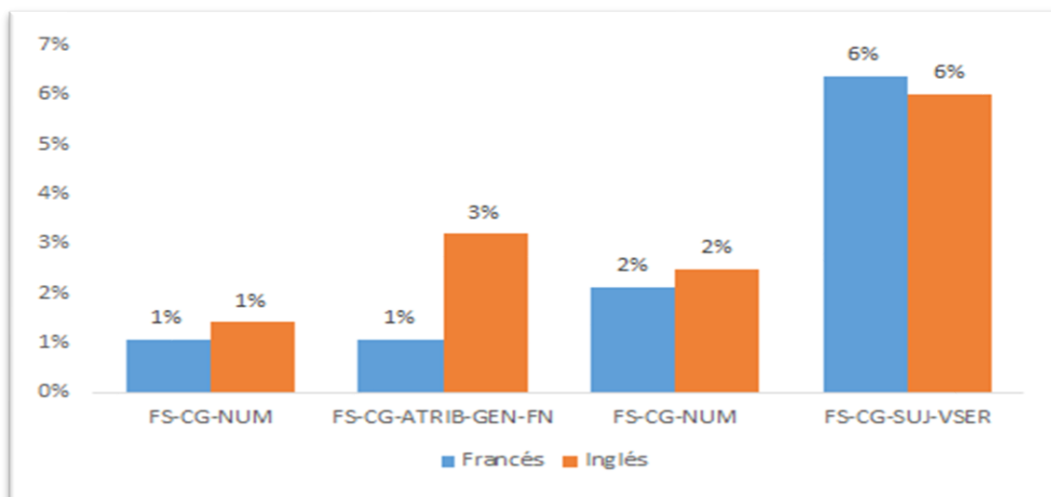


Gráfico 4. Similitudes en las frecuencias de errores entre los francófonos y anglosajones

Esta distribución de los errores con menor frecuencia revela una menor dificultad para este tipo de concordancias y un uso con mayor precisión. Estos hallazgos suponen un avance importante, dado que permiten identificar cuáles son los errores en los que debería centrarse el análisis respecto de la conformación de la interlengua. Es decir, la concordancia gramatical de género y número en el atributo prueban ser una dificultad mayor dadas las características morfológicas de los sustantivos y adjetivos en español, así también la distancia lineal entre los elementos entre los cuales debe darse la concordancia gramatical.

5. CONCLUSIONES

En este artículo, hemos presentado un estudio descriptivo con un enfoque de análisis de datos mixto con el propósito de explorar en dos subcorpus del corpus de aprendientes de Español como Lengua Extranjera, CAELE, las dificultades que se observan en aprendientes anglosajones y francófonos de nivel A2 con respecto al uso del verbo *ser* en las estructuras de atributos nominales en un contexto de tareas de escritura académica.

El CAELE se constituye en un corpus robusto de 1217 textos que permite examinar distintas problemáticas lingüísticas y realizar análisis de corpus a nivel léxico-gramatical. Este importante recurso lingüístico y su metodología de trabajo facilita realizar análisis de muestras textuales auténticas y representativas de aprendientes de ELE. La investigación realizada a partir del CAELE aporta información relevante sobre los procesos que sigue el aprendiente hasta alcanzar un determinado dominio en las estructuras de concordancia gramatical. Como se ha señalado, los estudios de corpus de aprendientes en formato computacional han contribuido con información relevante sobre las fases por las que pasa el aprendiente hasta la completa adquisición y las características de su interlengua en función del nivel de dominio en el que se encuentre. CAELE se constituye en un inventario abierto que año tras año considera la recolección y el procesamiento de nuevos textos.

Los resultados de la investigación han arrojado que para los aprendientes anglosajones hay mayores dificultades en la selección de las categorías de género y número en los adjetivos, especialmente en la concordancia de los atributos. En cambio, para los aprendientes del francés, si bien presentan un número de errores en dichos aspectos gramaticales, su frecuencia es notoriamente menor.

La interpretación de las posibles explicaciones de por qué los aprendientes de ambas L1 cometieron los errores está supeditada tanto al sistema lingüístico de su L1 como a la conformación y evolución de la interlengua en ELE. Para el caso del francés, puede concluirse que las estructuras erróneas podrían atribuirse a tres estrategias: (1) la transferencia de la L1 respecto de la semejanza en ambas lenguas dado su origen romance, (2) a la confusión que podría generar las reglas entre el español y el francés similares, como en el caso del género morfológico de los sustantivos y su correlación con los adjetivos en español. En este último se infiere que, a pesar de las similitudes entre el género del español y el francés, muchos de estos son diferentes en cada lengua, y (3) la posibilidad de la distancia entre los elementos que concuerdan tanto en el sintagma nominal como en el atributo, es decir, "los elementos de concordancia pueden situarse muy cercanos entre sí, como un sustantivo y su determinante definido o indefinido, o muy lejanos, tal como el sustantivo y su adjetivo" (Paquet, 2017: 60). Este distanciamiento podría afectar en la concordancia, considerando que muchas de las oraciones se apoyan en elementos modificadores de los sustantivos provocando un distanciamiento que conduciría al error.

Para el caso del inglés, los errores podrían atribuirse a tres estrategias: (1) la transferencia directa negativa dada la distancia entre las lenguas (ausencia de género y número gramatical de los adjetivos y sustantivos en casos específicos), (2) la sobregeneralización de reglas de fácil acceso para el aprendiente y (3) la reducción de estructuras complejas a simples como en el caso de la concordancia entre el sujeto y el verbo copulativo *ser*. En estas tres estrategias, la que mayor incidencia tiene es la transferencia negativa, dado que entre el inglés y el español muchas reglas reflejan la ausencia de rasgos morfológicos y gramaticales, la ausencia en el uso de elementos que faciliten la concordancia gramatical (como artículos o modificadores del sustantivo, donde el género garantiza la cohesión sintáctica en la oración (Paquet, 2017)) y las diferencias entre la organización interna de los sintagmas nominales y verbales. En cuanto a la sobregeneralización, es común observar que los aprendientes de inglés suelen generalizar reglas del género que resultan simples a la vista, dadas las terminaciones en la vocal "o" y la vocal "a" evocando la primera al género masculino y la segunda al femenino. Sin embargo, en las palabras con género neutro, se observa una tendencia a masculinizar las palabras terminadas en "e". Respecto de estudios que corroboran la dificultad que presenta aprender una LE a partir de una L1 sin género (Keating, 2009), puede adquirirse la asociación de elementos y la concordancia gramatical, sin embargo, no será procesada de la misma forma que un hablante nativo. También afirman que la influencia de la distancia entre los elementos podría afectar a la comprensión y sensibilidad al momento de aprender las reglas de la concordancia y, por ende, sería un proceso más lento y tardío que la adquisición de otras reglas gramaticales (Keating, 2009; Paquet, 2017).

Por lo tanto, la consideración de estas diferencias o similitudes en ambas lenguas es preponderante al momento de diseñar el proceso de aprendizaje en estas estructuras del verbo copulativo *ser*. Esto debido a que, en consonancia con los errores identificados, se pueda considerar la distancia y similitudes en ambas lenguas. Asimismo, puede arrojar luces sobre cómo la organización de las estructuras de las oraciones copulativas que son base para la precisión gramatical (concordancia entre el sujeto y el verbo, concordancia de género y número) frente al aprendizaje de la concordancia gramatical, es decir, cuáles de estos podrían priorizarse para la enseñanza y así cautelar un adecuado desarrollo de las estructuras más complejas del español respecto de las lenguas en estudio para esta investigación.

Finalmente, el proceso de enseñanza-aprendizaje en torno a la concordancia gramatical presente en las oraciones con el verbo *ser* es una temática morfosintáctica muy relevante en ELE y genera un impacto en la producción de las tareas de escritura académica por parte de los aprendientes francófonos y anglosajones. En el contexto del aula, la concordancia debe articularse con actividades que incorporen las cuatro habilidades comunicativas y ejemplos auténticos que reflejen la lengua en uso. Dichos ejemplos permiten al docente proporcionar explicación gramatical y a los aprendientes de ELE de competencia lingüística de nivel A2 visualizar, reflexionar y corregir los usos incorrectos.

AFILIACIÓN DEL PROYECTO

Este artículo se ha desarrollado en el contexto del Proyecto Fondecyt 1180974 "Diseño e implementación de un corpus escrito de aprendientes de Español como Lengua Extranjera (ELE) para el análisis de la interlengua", 2018-2021.

NOTAS

1 En el corpus CAELE cada texto está identificado con una etiqueta del tipo S43A2T4I2016, en la cual S43 corresponde al número del sujeto de la muestra, A2 al nivel de competencia; T4 al número del texto escrito por el sujeto; I corresponde a la L1 inglés y finalmente 2016 al año en que se recolectó el texto.

REFERENCIAS BIBLIOGRÁFICAS

- Agudelo, S. (2018). La interlengua de los estudiantes francófonos de ELE: el caso de ser y estar. *Forma y función*, 31(2), pp. 93-128.
- Aránzazu, M. (2004). La atención a la forma aplicada a la adquisición de la concordancia del género de los sustantivos "problemáticos" en español. *Interlingüística*, (15), pp. 1217-1224.
- Baralo, M. (1996). Adquisición y/o aprendizaje del español/LE. En: Tendencias actuales en la enseñanza del español como lengua extranjera II. Actas del VI Congreso Internacional de ASELE. [en línea] León: Universidad de León, págs. 63-67. Disponible en: https://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/06/06_0062.pdf [Consultado el 30 de julio de 2020].
- Baralo, M. (1998). Ser y estar en los procesos de adquisición de lengua materna (lm) y de lengua extranjera (le). En: Español como Lengua Extranjera: Enfoque Comunicativo y Gramática. Actas del IX Congreso Internacional de ASELE. [en línea] Santiago de Compostela: Universidade de Santiago de Compostela, págs. 291-300. Disponible en: https://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/09/09_0294.pdf [Consultado el 10 de agosto de 2020].
- Bautista, M. (2019). *Terminología y ontologías: un estudio alemán-inglés-español basado en corpus*. Granada: Editorial Comares.
- Bond, K, Gabriele A, Fiorentino, R. y Alemán Bañón, J. (2011). Individual differences and the role of the L1 in L2 processing: An ERP investigation. In: Herschensohn J and Tanner D (eds) Proceedings of the 11th Generative Approaches to Second Language Acquisition. Somerville, MA: Cascadilla Proceedings Project, 17-29.
- Buyse, K. (2011). ¿Qué corpus utilizar para qué fines en la clase de ELE? En: Del texto a la lengua: la aplicación de los textos a la enseñanza-aprendizaje del español L2-LE. Ponencia llevada a cabo en el XXI Congreso Internacional de ASELE. [en línea] Salamanca: págs. 277-288. Disponible en: <https://dialnet.unirioja.es/servlet/articulo?codigo=5419362> [Consultado el 27 de agosto de 2020].
- Campillos Llanos, L. (2014). Errores léxicos en el español oral no nativo: análisis de la interlengua basado en corpus. *Revista ELUA*, Alicante, v. 28, pp.85-124.
- Cruz Piñol, M. (2012). *Lingüística de corpus y enseñanza del español como 2/L*. Madrid: Arco Libros.
- Doquin de Saint Preux, A. y Sáez Garcerán, P. (2014). Errores gramaticales persistentes en la interlengua de estudiantes francófonos de ELE de nivel superior. *MARCO ELE Revista de Didáctica*, 19, 1, pp.-19.
- Elejalde, J. y Ferreira, A. (2016). Errores de transferencia en comunidades de aprendizaje en línea por aprendientes de español como lengua extranjera (ele). *Trabalhos em Linguística Aplicada*, 55(3), pp. 619-650.
- Fernández, S. (1997). *Interlengua y análisis de errores*. Madrid: Edelsa.
- Ferreira, A. (2016). *Prueba de multinivel con fines específicos académicos*. Universidad de Concepción, Chile.

- Ferreira, A. y Elejalde, J. (2017). Análisis de errores recurrentes en el Corpus de Aprendices de Español como Lengua Extranjera, CAELE. *Revista Brasileira de Lingüística Aplicada*, 17(3), pp. 509-538. <https://doi.org/10.1590/1984-6398201710927>
- Ferreira, A., y Elejalde, J. (2019). Hacia un perfil lingüístico-comunicativo del estudiante de Español como Lengua Extranjera para fines Académicos. *Revista Nebrija De Lingüística Aplicada a La Enseñanza De Lenguas*, 13(27), 145-165. <https://doi.org/10.26378/rnlael1327326>
- Ferreira, A., y Elejalde, J. (2020). Propuesta de una taxonomía etiológica para etiquetar errores de interlengua en el contexto de un corpus escrito de aprendientes de ele. *Forma y Función*, 33(1), 115-146 <https://revistas.unal.edu.co/index.php/formayfuncion/article/view/84182>
- Ferreira, A. y Elejalde, J. (2020). Análisis de Errores en ELE por Transferencia Negativa del inglés como L1 en clases presenciales versus clases no presenciales en comunidades virtuales. *Revista Brasileira de Linguística Aplicada*, 20(3), pp. 591-626. <https://doi.org/10.1590/1984-6398201710927>
- Ferreira, A. (2018-2021). FONDECYT- CONICYT N° 1180974. "Diseño e implementación de un corpus escrito de aprendientes de ELE en formato computacional para el análisis de la interlengua". Proyecto de Investigación. Conicyt, Chile.
- Fleissig, D. (2009). *El análisis de los errores de la concordancia de los estudiantes universitarios anglohablantes*. Tesis de Licenciatura. Universidad de Michigan.
- Gilquin, G. (2008). Combining contrastive and interlanguage analysis to apprehend transfer: detection, explanation, evaluation. In: G. Gilquin, S. Papp y M^a. Díez-Bedmar, eds., *Linking up contrastive and learner corpus research*, 1^a ed. Amsterdam/New York, NY: Rodopi. Language and Computers - Studies in Practical Linguistics, págs. 3-33.
- Granger, S. (1996). From CA to CIA and back. An integrated approach to computerized bilingual and learner corpora. In: K. Aijmer, B. Altenberg, y M. Johansson, eds., *Languages in Contrast. Papers from a Symposium on Text-based Cross-linguistic Studies*, 1^a ed. Cambridge: University Press, págs. 37-51
- Granger, S. (2002). A bird's-eye view of learner corpus research. In: S. Granger, J. Hungand y S. Petch-Tyson, eds., *Computer learner corpora, second language acquisition and foreign language teaching*, 1^a ed, Philadelphia: EE.UU: John Benjamins, págs. 3-33.
- Granger, S., Hung, J. y Petch-Tyson, S. (2002). *Computer learner corpora, second language acquisition and foreign language teaching*. Amsterdam: John Benjamins.
- Granger, S. (2003). Error-tagged learner corpora and CALL: A promising synergy. *CALICO Journal*, 20(3), pp. 465-480.
- Granger, S. (2004). Computer learner corpus research: Current status and future prospects. *Language and Computers*, 52(1), pp. 123-145.
- Granger, S. (2009). The contribution of learner corpora to second language acquisition and foreign language teaching. In: Aijmer, K., eds., *Corpora and language teaching*, 1^a ed. Amsterdam: John Benjamins, págs. 13-32.
- Granger, S. (2012). How to use foreign and second language learner corpora. In: A. Mackey y S. Gass. eds., *Research methods in Second Language Acquisition: A practical guide*, 1^a ed. Malden: Blackwell Publishing, págs. 7-29.
- Granger, S. (2015). Contrastive interlanguage analysis: A reappraisal. *International Journal of Learner Corpus Research*, 1(1), pp. 7-24.
- Guijarro-Fuentes, P. (2008). ¿Ser o estar? Cuestiones sin resolver para el español como lengua extranjera. *Revista Nebrija de Lingüística Aplicada a la Enseñanza de Lenguas*, 2(3), pp. 44-55.
- Hasselgard, H. y Johansson, S. (2011). Learner corpora and contrastive interlanguage analysis. In: F. Meunier, S. De Cock, G. Gilquin y M. Paquot, eds., *A taste for corpora: In honour to Sylviane Granger*, 1^a ed. Amsterdam/Philadelphia: John Benjamins Publishing Company, págs. 33-62.

- Ishikawa, K. (2013). The ICNALE and sophisticated contrastive interlanguage analysis of Asian learners of English. In: S. Ishikawa, Ed., *Learner corpus studies in Asia and the world*, 1ª ed. Kobe: Kobe University, págs. 91-118.
- Keating, G.D. (2009). Sensitive to violations of gender agreement in native and no native Spanish: An eye-movement investigation. *Language Learning*, 53(3), pp. 503-535.
- Kilgarriff, D., Baisa, V., Bušta, J., Jakubíček, M., Kovář, V., Michelfeit, J., Rychlý, P. y Suchomel, V. (2014). *The Sketch Engine: Ten years on. Lexicography*, 1: 7-36.
- Kilgarriff, D., Rychlý, P., Smrž, P. y Tugwell, D. (2004). *The Sketch Engine. Information Technology*. 105, 116.
- Liceras, J. (1983). *Contrastive analysis and the acquisition of Spanish syntax by English speakers*. Tesis Doctoral. Universidad de Toronto.
- Lozano, C. (2015). Learner corpora as a research tool for the investigation of lexical competence in L2 Spanish. *Journal of Spanish Language Teaching*, 2(2), pp. 180-193.
- Méndez de Sifontes, I. y Rojas, I. (2013). An error analysis of persistent grammatical errors carried over from beginner to intermediate Spanish. *Revista Electrónica de Lingüística Aplicada*, 12, pp. 63-85.
- Montes, J. (1997). Notas sobre el género gramatical en Bello y en la actualidad. *Thesaurus. Boletín del instituto Caro y Cuervo*, 52(1-3), pp.144-155.
- Naranjo, J. (2009). *Análisis de errores en la interlengua de estudiantes de español como lengua extranjera. Un estudio sobre la correlación de los errores con la comunicabilidad en la producción escrita de estudiantes de habla inglesa*. Tesis de magister. Universidad de Bergen.
- Paquet, P. (2017). *Influencia de las propiedades de la L1 y del nivel de competencia en la adquisición de la concordancia de género: una comparación entre estudiantes francófonos y anglófonos de español L2*. Tesis de Doctorado. Universidad de Alicante.
- Pastor Cesteros, S. (2001). La concordancia en la interlengua de los aprendices de español como lengua extranjera. *ELUA. Estudios de Lingüística, Anexo 1*, pp. 95-125.
- Pastor Cesteros, S. (2004). *Aprendizaje de segundas lenguas: Lingüística aplicada a la enseñanza de idiomas*. Alicante: Editorial Universidad de Alicante.
- Penadés, I. (1999). *Lingüística contrastiva y análisis de errores*. Madrid: Edinumen.
- Prieto, C. (2011). *Análisis de errores de estudiantes anglófonos*. Memoria de investigación. Universidad de Alcalá.
- Rakaseder, J. y Schmidhofer, A. (2014). Errores de producción de textos por parte de germano parlantes en las primeras etapas de aprendizaje de ele. *MARCO ELE Revista de Didáctica*, 19, pp. 1-17.
- Real Academia Española. (2005). *Diccionario Panhispánico de dudas*, 1ª ed. Madrid: Taurus.
- Real Academia Española. (2010). *Manual de la Nueva Gramática de la Lengua Española*. Buenos Aires: Espasa.
- Real Academia Española. (2014). *Diccionario de la lengua española*, 23ª ed. Buenos Aires: Espasa.
- Rubinjon, L. (2014). Análisis de los errores de concordancia en la expresión escrita de los aprendientes serbios de ELE. *Linred: Lingüística en la Red*, 11, pp. 1-34.
- Sánchez, A. (2015). La Investigación de Corpus de Aprendientes y el desarrollo de los estudios de la interlengua del español. *Language design: Journal of theoretical and experimental linguistics*, 17, pp. 57-83.
- Santos Gargallo, I. (1993). *Análisis contrastivo, análisis de errores e interlengua en el marco de la lingüística contrastiva*. Madrid: Síntesis.
- Vázquez, G. (1991). *Análisis de errores y aprendizaje de español/lengua extranjera*. Frankfurt am Main: Peter Lang.

Dinámica del error de concordancias plurales en un corpus de cuatro aprendientes italianos de español L2

The dynamics of agreement errors in a corpus of four Italian learners of Spanish L2

Pablo Ezequiel Marafioti

Universidad Nacional de Córdoba (Argentina)

kenavo38@yahoo.com

RESUMEN

Se propone analizar la evolución de la producción de concordancia plural en cuatro aprendientes italianos de español L2, en un estudio de caso longitudinal, mediante el marco teórico de los sistemas dinámicos. Se crearon dos series temporales para cada alumno, tipificando la producción según: no error, error de género, de no inserción de -e- epentética ("-es"), de plural y mixtos. Luego dichas series se abordaron mediante técnicas no lineales cuantitativas que involucraban entropías y análisis de recurrencias. El objetivo consistió en encontrar periodos de transición en la dinámica del error e identificar sesiones donde se promoviera el aprendizaje.

Palabras clave: lengua segunda, sistemas complejos, concordancia, número, plural

ABSTRACT

This article analyzes the evolution in the production of plural agreement in four Italian learners of Spanish L2, a longitudinal case study, using the dynamical systems approach. Two time series for each student were created, tagging production of agreement with: no error, gender error, error due to no insertion of epenthetic -e- ("-es"), plural error and mixed error. Then these series were treated with quantitative nonlinear techniques involving entropies and recurrence analysis. The main goal consisted in finding transitions periods in error dynamics and identifying sessions in which learning was promoted.

Keywords: second language, complex systems, agreement, number, plural

1. GÉNERO, NÚMERO Y CONCORDANCIA EN ESPAÑOL

La categoría gramatical del GÉNERO asigna sustantivos a dos (o más) clases y es una propiedad léxica inherente al nombre. Dicha asignación puede basarse en criterios semánticos, morfológicos o fonológicos (Corbett, 1991). En español algunos sustantivos animados poseen su género basado en alguna característica de su referente, como por ejemplo el sexo biológico (criterio semántico): (a) *el hombre / la mujer*; (b) *el gat-o / la gat-a*. En (a) la asignación a las clases "femenino" y "masculino" se expresa mediante bases léxicas diferentes; en cambio, en (b) la misma base léxica toma diferentes morfemas para expresarla; es decir que el criterio resulta también morfológico. En lo concerniente al criterio fonológico, a modo de ejemplo, las formas en *-i* y *-u* tienden a ser masculinas (*el espíritu, el taxi, pero: la tribu*). Además, los segmentos *-ma* y *-ta* también determinan la pertenencia al masculino (*el problema, el planeta*). Formalmente, los nombres terminados en *-a* son en general femeninos (96.3 %) y los terminados en *-o*, masculinos (99.8 %) [Alarcón, 2011]. Sin embargo, existen formas menos prototípicas que asignan ambas clases; aquellas en *-e*: *el puente, la suerte*, y en consonante: *el camión, la canción*; así como excepciones infrecuentes del tipo: *la mano, la foto, el día, la radio*. Incluso hay casos donde el cambio de clase implica un cambio de significado: *el libro / la libra; el capital / la capital*. Si bien los nombres en español solo se asignan a una de las clases de género, casi todos ellos pueden exhibir la categoría gramatical de NÚMERO en sus rasgos 'singular' o 'plural'. Formalmente, el plural de los nombres acabados en vocal no acentuada se forma añadiendo *-s* (*libros, barcos*); mientras que agregan *-es* aquellos acabados en consonante (*relojes, paredes*), a través de inserción fonológica de una *-e-* epentética. Los que finalizan en *-e* acentuada solo admiten plural en *-s* (*bebés, cafés*), mientras que las demás vocales acentuadas tienden a presentar dobles (*esquí / esquís; tabú / tabúes*). Por otro lado, los nombres terminados en vocal inacentuada seguida de */-s/* no varían (*lunes, tórax, crisis, hipótesis*). En general en los sustantivos concretos y contables, la oposición singular / plural indica cuantificación ('uno' / 'más de uno') de un *mismo* referente. En cambio, en los nombres de sustancia, abstractos y colectivos, su pluralización apunta en su mayoría a un referente distinto y/o conlleva un significado diferente: (a) contable / incontable: *algodón / algodones*; (b) abstracto / persona u evento: *autoridad / autoridades, locura / locuras*; (c) plural estilístico: *gente / gentes*. Asimismo, los plurales que refieren a objetos complejos no son significativos (*tijera / tijeras*); mientras que en otros casos cambia completamente el referente al pluralizarse (*esposa / esposas*). Por último, existen algunas formas sin plural (*fe, salud*). A diferencia del género (inherente al nombre), el número es transcategorial: lo pueden manifestar sustantivos, verbos, adjetivos, artículos y pronombres. En general, adjetivos, artículos y pronombres forman el plural siguiendo las reglas morfofonológicas arriba descritas; a excepción de las formas invariables en los relativos: *lo que, lo cual, cuanto*; los interrogativos: *qué, cuánto*; y los indefinidos: *nada, ninguno/a, cada, más, menos, demás*. Todos los demostrativos, posesivos y artículos cuentan con formas de singular y plural, dejando de lado a las formas neutras *esto, eso, aquello, lo* (Ambadiang, 1999).

La concordancia se define como una relación entre rasgos sub-léxicos (pares 'valor: atributo') de los ítems léxicos (O' Grady, 2005). En español dichos rasgos son 'persona', 'número' y 'género', junto a sus valores. Corbett (2006) denomina *controlador* al ítem léxico que determina la concordancia y *objetivo* al elemento cuya forma es determinada por aquel. Se denomina *dominio* al entorno sintáctico en el cual ocurre la concordancia. La

concordancia se establece por covarianza sistemática de rasgos. En el presente trabajo el controlador será nominal, presente y expresará sus rasgos abiertamente. Por otra parte, los objetivos consistirán en artículos (definidos e indefinidos), adjetivos y pronombres (demostrativos, posesivos, indefinidos). Usarán morfemas ligados (flexión) para expresar concordancia; con marcado obligatorio; usando morfología regular, productiva y con diferentes grados de aliteración. Concuerdan con un solo controlador obligatoriamente. Los dominios relevantes serán el sintagma nominal; el sintagma verbal (predicativo) y la oración subordinada. La concordancia se considerará asimétrica (el género y número de los objetivos dependen del controlador nominal). En lo que atañe a la concordancia a larga distancia, se considerarán como tales aquellas del adjetivo (en función predicativa) con el controlador con verbos copulativos (*ser, estar, parecer, quedar, etc.*), también dentro de una oración subordinada relativa. El caso de elipsis en posición predicativa no será considerado como a larga distancia sino como local [*Los libros interesantes son los [libros] rojos*]. Tampoco se considerará la concordancia en la voz pasiva con *ser* pero sí con *estar* [*Las mesas están servidas*].

2. SISTEMAS DINÁMICOS EN L2

Un *sistema dinámico* es aquel que cambia en el tiempo. El sistema posee un conjunto de componentes que interactúan de una forma determinada para generar un *estado* particular (Larsen Freeman y Cameron, 2008). Un *estado* se define como un patrón de actividad determinado del sistema en un punto determinado del tiempo. Los componentes pueden ser entidades, procesos e incluso otros (sub-)sistemas. Si un sistema se cataloga como *simple*, dichos componentes se hallan conectados de una manera que resulta en un comportamiento predecible y sin cambios abruptos. Ello quiere decir que una vez especificadas las reglas que gobiernan el sistema, es posible conocer sus estados futuros. Por otro lado, un sistema que es *complejo* conlleva una serie de características que lo diferencian de sus contrapartes simples. A saber:

- (1) Heterogeneidad. Los componentes del sistema pertenecen a tipos diferentes y podrían ser incluso otros sistemas complejos.
- (2) Dinamismo. Los componentes y su forma de interactuar entre sí cambian con el tiempo.
- (3) Falta de anidamiento. Las variables se relacionan con otras que se hallan en otros niveles. No obstante, dicha relación no puede pensarse como «jerárquica» ya que la influencia entre las variables puede funcionar en cualquier dirección.
- (4) Proceso iterativo. El sistema es afectado por flujos provenientes del ambiente y por el estado previo del sistema.
- (5) Falta de linealidad. En un sistema lineal la relación entre los elementos está fija y el cambio es constante y proporcional al input recibido. Por otra parte, un escenario no lineal es aquel donde las relaciones entre elementos del sistema se mueven constantemente y el cambio no es proporcional.
- (6) Interacción con el ambiente. El ambiente 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» quiere decir que el sistema permite la salida o entrada de energía; con lo cual, 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; o sea, se adapta a transformaciones del contexto.

(7) Auto-organización. Se produce un incremento de complejidad por medio de la creación de nuevas formas. Dichas formas 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 energía del ambiente (porque es sistema es abierto). Además, se depende de la disponibilidad de información en el ambiente.

(8) Fractalidad. Es un patrón que existe en un nivel de escala del sistema y también para cualquier otro nivel de escala. El sistema se vuelve similar a sí mismo.

Varios artículos han explorado la relación entre los sistemas dinámicos complejos y la adquisición de lenguas extranjeras: Larsen Freeman (1997, 2011, 2015), Lowie (2013); De Bot et al. (2007, 2013); Jiménez-Ramírez (2020).

Un alumno, al aprender una lengua, crea variabilidad, o sea, produce nuevas formas que no están codificadas en el input: "errores". La interacción entre los parámetros de un sistema dinámico eventualmente hace que surja una coordinación entre ellos, emergiendo un patrón coherente que se perpetúa en el tiempo: el atractor; que se representa como puntos recurrentes en una región del espacio de fase. Es decir que los alumnos vuelven a cometer patrones de errores recurrentes; incluso quedando atrapados en un estado de bimodalidad (por ejemplo, un alumno italiano podría cometer el error *los alemanos / alemanes*, aplicando la estrategia de aprendizaje "-i" > "-os": "tedesch-i" > "aleman-os"). O bien, si la adquisición avanza, se acercan a una región del estado de fase cuyos patrones son muy similares a los del hablante nativo en la misma situación de comunicación y con el mismo objetivo pragmático. El sistema también se podría apartar de ciertos estados, los cuales constituyen "repulsores" en el espacio de fase. Sin embargo, los atractores poseen diferentes grados de fuerza; de ahí que los alumnos persistan en la reproducción de determinado patrón antes de saltar a un patrón diferente.

Las condiciones iniciales de cada alumno son distintas, y, por lo tanto, los trayectos de aprendizaje también. Esto es así porque un estado del sistema depende del anterior, o sea, depende del conocimiento (habilidades) adquirido por el estudiante en particular. En otras palabras, las condiciones de los diferentes subsistemas (fonológico, morfológico, sintáctico, etc.) en un punto temporal determinan la producción en el siguiente. El aprendizaje depende asimismo de los recursos del "ambiente cognitivo": memoria de trabajo, atención, motivación, esfuerzo, etc.; y el estado del sistema se adapta a cambios en dichos factores. El sistema puede ser "perturbado" en cualquier momento por transformaciones repentinas del "ambiente cognitivo".

Los alumnos adaptan, pues, su producción al contexto, que es cambiante; lo cual desencadena fluctuaciones y variabilidad. Ello equivale a decir que el aprendizaje no resulta lineal: un estudiante no aprende completamente un patrón para luego pasar al siguiente. En consecuencia, no es plausible pensar el aprendizaje desde la metáfora de una «escalera», en la cual los aprendientes pasan por estadios discretos de menor a mayor complejidad en sentido ascendente, partiendo de un "escalón inicial" y llegando a un "escalón final" o estado nativo (Larsen-Freeman, 2006). No existen estadios discretos sin variabilidad; estos son atractores en los que determinadas formas dominan en frecuencia sobre otras.

Además, se encuentran factores que son decisivos para el comportamiento del sistema. Estos se denominan variables controladoras, ya que, más allá de un cierto valor (punto crítico) el sistema sufre un cambio cualitativo (bifurcación) y se reorganiza hacia una nueva

configuración de atractores. El atractor inicial se vuelve inestable y comienza a fluctuar cada vez más, llevando a saltos y discontinuidades. Como resultado *emergen* nuevos patrones que, eventualmente, se vuelven estables. Cerca del atractor hay linealidad y continuidad; en cambio, cerca de una transición se observan discontinuidad y saltos (variabilidad). En suma, opera un mecanismo acumulativo de competencia entre las formas hasta atravesar el punto crítico, lo cual dispara el cambio de dinámica (Larsen- Freeman y Cameron, 2008). Por ejemplo, en el contexto de un locativo puede dominar la forma *estar*, y en el contexto de los adjetivos psico-físicos la forma de *ser*: *está en su casa porque *es enfermo*; en el cambio de fase observaríamos gran variabilidad entre *ser* y *estar* con dichos adjetivos; luego la dinámica se estabilizaría -con algunas perturbaciones que llevarán a errores, por supuesto- en una nueva configuración *estar + lugar / + adjetivos psico-físicos*. Recuérdese asimismo que los atractores no son irreversibles porque dependen de las condiciones del ambiente. Por ejemplo, luego de años de práctica, las configuraciones de *estar + lugar / + adjetivos psico-físicos / + adjetivos léxicos* (*el libro está ahí / mi amigo está contento; es un chico listo - la cena está lista*) pueden lograr precisión arriba del 85% (atractor más estable) pero las del tipo *está triste porque lo dejó la novia / es un niño triste* seguramente experimentarían mucha más variabilidad (atractor menos estable).

La auto-organización, es decir, el hecho de que el sistema avance hacia nuevas configuraciones estables y no presentes en las condiciones iniciales, implica una reorganización en la información. Aprender significa que la información se vuelva más clara, que haya más certidumbre, ya que los atractores que emergen tienden a la estabilidad, hay menos períodos de intermitencia de error. Para usar el ejemplo anterior, en el contexto del locativo y del adjetivo psico-físico, el aprendiente usará con más probabilidad *estar* que *ser*. Al auto-organizarse, el sistema del aprendiente reduce la incertidumbre (donde tenía más dudas sobre qué forma utilizar ahora tendrá menos) y la complejidad de lo aprendido (lo que era difícil ahora es más fácil).

El aprendiente no es un mero receptor pasivo de input. En lugar de «input», es mejor hablar de *affordances* o percepción de oportunidades de aprendizaje. Se trata de relaciones aprendiente-entorno comunicativo, que el mismo aprendiente percibe como posibilidad o potencialidad de aprendizaje. Una vez detectada la oportunidad en la interacción comunicativa, el aprendiente pasará a la acción adaptando los recursos de expresión de los cuales dispone. Dichas potencialidades para la acción son percibidas por los aprendientes mientras están activos; o sea, participando de actividades y tareas comunicativas. No es que las características del input (frecuencia, saliencia, etc.) por sí solas promuevan la adquisición. Resulta esencial que el aprendiente note que una forma frecuente del input lo ayudará a mejorar. Lo que no se perciba como oportunidad para la acción, aunque sea frecuente, constituye «ruido». En consecuencia, lo más frecuente, saliente, menos ambiguo, etc. no necesariamente se adquiere antes (Larsen- Freeman, 2015).

Por otra parte, los sistemas son iterativos: lo que resulta de una iteración se usa como punto de partida para la siguiente. Lo que promueve la adquisición es la repetición de los mismos eventos de uso del lenguaje a lo largo de tareas comunicativas diferentes. La repetición no es vista como replicación exacta sino como mecanismo de generación de variabilidad (Larsen Freeman, 2012). No repetir el mismo evento de lenguaje de la misma forma crea más opciones como recursos para expresión de significados. Cada actividad comunicativa conlleva sus propios objetivos. La repetición de construcciones en actividades diferentes permite que el aprendiente las adapte a los nuevos objetivos comunicativos. En suma, la iteración, al generar variabilidad, provee de opciones para adaptarse a nuevos contextos y satisfacer diferentes necesidades comunicativas.

3. ANTECEDENTES

Los artículos que han seguido este enfoque se han centrado en producción escrita; generalmente buscando evidencia de que la adquisición de lenguas segundas se rige efectivamente por los principios de los sistemas dinámicos arriba descritos. Varios de ellos han usado también el modelo ecológico de Paul Van Geert (1994, 1995, 2003); en el cual las variables interactúan entre sí por medio de varias relaciones funcionales y compiten por la disponibilidad de recursos limitados (memoria de trabajo, atención, motivación, tiempo invertido en el aprendizaje, conocimiento pre-existente, etc.). Estos recursos ejercen un efecto de curvatura en lo que otra de forma sería un crecimiento exponencial de dichas variables. Las relaciones entre éstas pueden ser: (a) de soporte: ambas se desarrollan al unísono porque se refuerzan mutuamente (correlación positiva); (b) de competición: cuando una crece la otra decrece debido a la competición entre ellas (correlación negativa); (c) condicional: se necesita que una variable llegue a un cierto nivel para que la otra empiece a crecer.

Verspoor, Lowie y Van Dijk (2008) estudiaron las producciones escritas a lo largo de dos años en un estudiante holandés de inglés L2 (nivel superior). La relación entre la longitud media de los SNs y el ratio palabras / verbos finitos (ambas medidas de complejidad) fue en general de apoyo mutuo; sin embargo, al principio de la trayectoria se registró un poco de competición entre las variables. Esto significa que, si bien en general la complejidad de los SNs aumenta junto a la oracional, al principio el aprendiente buscó complejizar produciendo SNs más largos a expensas de oraciones con verbos finitos, una estrategia común en holandés.

Ellis y Ferreira (2009) encontraron, en estudiantes de inglés L2, evidencia de distribuciones de Zipft; las cuales siguen una ley de potencia, que constituye un posible índice de fractalidad. Examinaron la adquisición de tres tipos de construcciones: (a) Verbo + locativo: *come here!*; (b) Verbo + objeto + locativo: *put it over there*; (c) Verbo + objeto1 + oobjeto2 (ditransitiva): *give the book to me / give me the book*. Hallaron evidencia de ley de potencia en la distribución de los tipos de verbos usados en cada construcción tanto en el input como en el output. Es decir que la frecuencia de cada tipo de verbo era inversamente proporcional a su posición en la tabla de frecuencias para cada construcción en particular. Además, comparando los verbos pertenecientes a cada construcción, el tipo de verbo de más frecuencia en el input era siempre mucho más frecuente que los otros verbos, resultaba ser además altamente prototípico y de significado genérico. O sea, un input que sigue la distribución de Zipft ayuda a optimizar el aprendizaje ofreciendo ejemplares prototípicos, frecuentes y ampliamente aplicables a varios contextos. Con dichos verbos el aprendiente se forma «una construcción marco», a partir de la cual puede reclutar los otros verbos para cada construcción.

Caspi (2010) siguió durante diez meses a cuatro estudiantes de inglés L2 de nivel avanzado. Consideró cuatro niveles jerarquizados de conocimiento de vocabulario de menor a mayor uso en producción: (1) reconocimiento; (2) recuperación de la memoria; (3) producción controlada; (4) producción libre. Halló que los niveles (1) y (2) más pasivos se apoyaban entre sí; en cambio, los más activos (3) y (4) competían entre sí. También estudió las variables de precisión y complejidad en la sintaxis y el léxico en producción escrita. Encontró que las variables de precisión y complejidad en el dominio léxico evidenciaban competición débil, mientras que en el dominio sintáctico se reforzaban entre sí. Por otro lado, la competición más fuerte se produjo entre la precisión léxica y la complejidad sintáctica.

Spoelman y Verspoor (2010) investigaron el desarrollo del sistema de casos en un estudiante holandés principiante de finés L2. A pesar de que el sistema de casos constituye un desafío de aprendizaje, la precisión se mantuvo entre el 80-100 %. La tasa de error superó el 10 % solamente en el partitivo y el nominativo singular, notorios por su morfología no transparente y complejidad semántica. Hallaron dos etapas de adquisición, con aumento de variabilidad en la vecindad de una transición de fase de una etapa a la otra. Por un lado, la complejidad en el nivel de la palabra evidenció relaciones de mutuo refuerzo con la complejidad tanto a nivel del SN como de la oración. Por otro lado, la complejidad en la palabra estuvo en competición con la de la oración, estabilizándose al final.

Eskildsen (2012) estudió a dos aprendientes hispano hablantes de inglés L2 (Carlos y Valerio). Eran datos longitudinales de entre dos años y medio y tres años. Se focalizó en la adquisición de la negación. Las trayectorias de aprendizaje de ambos fueron diferentes. En el caso de Carlos, se caracterizó por: alta frecuencia de la fórmula «I don't know»; una competencia entre los patrones «SUJETO don't VERBO» (correcto) y «SUJETO no VERBO» (incorrecto). Valerio también exhibió alta frecuencia de «I don't know». No obstante, ello y a diferencia de Carlos, al principio dominó el patrón «SUJETO don't VERBO por sobre «SUJETO no VERBO». Luego la dominancia se dio a la inversa; y, al final, el patrón «SUJETO no VERBO» bajó su frecuencia hasta casi desaparecer. En consecuencia, las trayectorias fueron específicas de los aprendientes y no se conformaban a estadios «universales» de adquisición.

Baba y Nitta (2013) estudiaron la fluidez (cantidad de palabras) en producciones escritas con tiempo limitado (10 minutos). Siguió a dos estudiantes japoneses de inglés L2 durante un año académico. Hallaron evidencia de al menos una transición / bifurcación en cada estudiante según cuatro índices: (a) saltos repentinos de frecuencia en los datos; (b) aumento de la variabilidad cerca de la transición de fase; (c) presencia de un elemento de «catálisis» que empuja al sistema hacia un nuevo estadio; (d) cambios cualitativos en el modo de escribir. Los índices (a) y (b) se analizaron cuantitativamente y los otros cualitativamente. Por ejemplo, se observaron (a partir de comentarios de los alumnos para cada composición) como elementos catalizadores la auto-confianza y la facilidad del tópico elegido. En cuanto a cambios cualitativos, se notó, por ejemplo, que, en la segunda etapa, se producían escritos con más párrafos y más sintéticos; o bien subdividiéndolos con subtópicos.

Lowie et al. (2014) también encontraron evidencia de fractalidad en el aprendizaje, detectando «ruido rosa». Por medio de un análisis espectral, la variabilidad fue analizada relacionando la magnitud de los cambios en el sistema con la frecuencia en la que éstos ocurren. Si todas las magnitudes de cambio ocurren con igual frecuencia, entonces el patrón es totalmente estocástico e impredecible, asociado al «ruido blanco». En el extremo opuesto, si la magnitud del cambio es completamente (inversamente) proporcional a la frecuencia de ocurrencia, se observa un comportamiento hiper-regulado o «ruido marrón». El «ruido rosa» es la situación óptima de organización u coordinación entre subsistemas (ya que permite la adaptación rápida), que se encuentra entre ambos extremos. A lo largo de 6 años se estudiaron los tiempos de reacción de un individuo en una tarea para nombrar palabras en holandés (L1) e inglés (L2). Los resultados arrojaron, previsiblemente, más coordinación («ruido rosa») en L1 que en L2, pero también un aumento de la coordinación en L2 a medida que se incrementaba el nivel de competencia del aprendiente. Además, luego de un período de práctica intensiva en la L2, los datos evidenciaban incluso una tendencia hacia el «ruido rosa», interpretado como incremento en la auto-organización del sistema de aprendizaje.

4. CORPUS Y VARIABLES CREADAS

Se analizaron datos de cuatro casos de estudiantes de español como lengua extranjera. Se trató de cuatro alumnos adultos, de lengua nativa italiana, estudiantes del Instituto Cervantes de Milán en el año académico 2008/09. Cada alumno poseía un nivel distinto de competencia lingüística (según el Marco Común Europeo de Referencia). Se hicieron entrevistas de 30 minutos entre el alumno y el investigador (autor de este trabajo). La tarea consistió en una conversación no estructurada, sobre temas acordes al nivel de competencia del sujeto. Dichas entrevistas tuvieron lugar aproximadamente cada 20 días, según la disponibilidad de los alumnos. Cada alumno realizaba simultáneamente el curso de español. Hubo entre doce y catorce entrevistas por alumno. El *corpus* estuvo 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. El cuadro que sigue muestra el perfil de cada alumno.

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

Tabla 1. Perfil de los sujetos

La codificación y transcripción de los datos se hizo mediante el formato CHAT, siguiendo a Mac Whinney (2020). El siguiente constituye un ejemplo de transcripción del aprendiente SONIA (sesión 2):

```

1 @Begin
2 @Languages: spa
3 @Participants: STU Sonia Participant, INV Pablo Investigator
4 @ID: spa | sonia | | |female| | |Participant| |
5 @ID: spa | sonia | | |male| | |Investigator| |
6 *INV: háblame de tus amigos .
7 *STU: yo tengo muchos amigos [*0] .
8 *STU: Marina es fotógrafa .
9 *STU: ella quiere hacer fotos .
10 *STU: y para hacer fotos ella va a Londres (.) a París (.) a Madrid (.) a
    Berlino@s:ita a Roma .
11 *STU: por le@s:ita grande ciudades [*3] de Europa .
[...]
```

Cada concordancia se codificó con dos términos, pero pudiendo haber más términos “objetivo”: por ejemplo, en *los libros azules* se codificaron dos instancias: *los libros* y *libros*

azules. Se anotaron a continuación etiquetas en el corpus para realizar el conteo posterior. Son las siguientes: (i) [*0] = ausencia de error; (ii) [*1] = errores en el género; (iii) [*2] = errores debidos al uso de la terminación “(-e-)s”: (a) por no tomar en cuenta la última consonante de la raíz léxica, que exige un plural con “e” epentética en -(e)-s; (b) por uso en contexto incorrecto u omisión en correcto; (iv) [*3] = errores de plural, o sea ausencia de -s; (v) [*4] = errores mixtos por acumulación de los anteriores. Por ejemplo:

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

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

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

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

El conteo se hizo con el programa CLAN (Mac Whinney, 2020). Se codificaron 1857 concordancias en total. Se crearon dos tipos de series con niveles idénticos (0, 1, 2, 3, 4) a las etiquetas descritas. La primera serie calificaba cada instancia de concordancia con dichas etiquetas. En cambio, en la segunda se calificaba el error de cada término de la concordancia; o sea, en una serie bi-variada. Así, por ejemplo, las primeras cuatro instancias de SONIA son: “muchos personas”, “muchos pueblos”, “mucho pensadores”, “otros autores”. En la primera serie se codifica como [1 0 3 0]; en cambio en la segunda es: SERIE_1 = [1 0 3 0] (primer término); SERIE_2 = [0 0 0 0] (segundo término).

5. OBJETIVOS

Los objetivos del presente trabajo son: (i) Aplicar a un corpus oral longitudinal técnicas de análisis no lineales; (ii) Identificar sesiones donde se promueve la adquisición; (iii) Descubrir posibles saltos cualitativos (bifurcaciones que reorganizan el aprendizaje); esto implica detectar posibles transiciones entre regímenes de errores. Se responderán las siguientes preguntas: (i) ¿Qué regímenes de dinámica diferente es posible identificar?; (ii) En particular, ¿cuáles sesiones destacan frente al resto por tener una mejor dinámica del error?; (iii) si hay un cambio en el aprendizaje, ¿Dónde se encuentra la transición?

Se llevó a cabo el análisis en R (R Core Team, 2017). El código empleado se encuentra en: https://github.com/pablomarafioti/PabloMarafioti/tree/master/dinamica_de_error/R%20scripts

6. MÉTODOS

Para analizar los datos se utilizaron las técnicas cuantitativas no lineales que se describen en los apartados que siguen.

6.1. El plano entropía – complejidad

En el campo de la teoría de la información (Cover y Thomas, 1991), la información de un evento en particular (“errores”, “ítems correctos”, etc.) está definida como el logaritmo (en base 2, si la unidad es el *bit*) de la probabilidad de dicho evento: $\log_2[p(x)]$. Los eventos improbables nos dan mayor información que los probables (nos “sorprenden” más). La entropía de *Shannon* sobre una función de masa de probabilidad (FMP) de la serie de los eventos¹: $\vec{\pi} = \{\pi_1, \dots, \pi_n\}$ reza: $-\sum_{i=1}^N p(\pi_i) \log_2 p(\pi_i)$. La entropía de permutación [*permutation entropy*] fue originariamente propuesta por Bandt & Pompe (2002). En lugar de tomar como base para el cálculo de la entropía de *Shannon* palabras W de longitud L , ellos propusieron asignar a cada palabra un patrón ordinal obtenido por comparación de los valores (símbolos) adyacentes que contiene la palabra W . Más formalmente (ver: Zanin et al., 2012; Ribeiro et al., 2012), considérese una serie de tiempo $\{x_t: t = 1, \dots, N\}$ y en cada punto s de dicha serie, un vector (palabra W) compuesto por D valores subsiguientes: $s \mapsto (x_s, x_{s+1}, \dots, x_{s+(D-2)}, x_{s+(D-1)})$. D se denomina *dimensión* y define el largo del vector/palabra. A cada palabra se asocia luego un patrón ordinal en términos de la permutación $\pi = (r_0, r_1, \dots, r_{D-1})$ del vector $(0, 1, \dots, D-1)$ de forma que se cumpla el orden creciente: $x_{s+r_0} \leq x_{s+r_1} \leq \dots \leq x_{s+r_{(D-2)}} \leq x_{s+r_{(D-1)}}$. Por ejemplo, para la serie $\{3,6,8,9,5,10,2\}$ y con $D = 2$ tenemos para $s = 1$ la palabra (3,6) que ya se encuentra en orden creciente, produciendo el patrón $\pi = (01)$. Para $s = 2$, la palabra (6,8) también produce $\pi = (01)$; en cambio para $s = 4$ obtenemos la palabra (9,5); que puesta en orden creciente es (5,9), y se le asigna el patrón $\pi = (10)$. Se habrá notado entonces que para $D = 2$, solo es posible un patrón creciente o uno decreciente, es decir, $\vec{\pi} = \{\pi_1 = (01), \pi_2 = (10)\}$; en general los posibles patrones para cada D siguen al factorial $D!$ si los valores de la serie son diferentes entre sí (no se repiten). El siguiente paso consiste en generar las probabilidades para cada patrón, siguiendo: $p(\pi_i) = \frac{\#\{s: s < N-D+1; s \in \pi_i\}}{N-D+1}$, donde $\#$ representa las frecuencias absolutas de cada patrón π posible. Se forma entonces la FMP $\vec{\pi} = \{p(\pi_1), \dots, p(\pi_{D!})\}$. Se define pues la entropía de permutación $H_s[P]$ en términos de la entropía de *Shannon* para los patrones π normalizada sobre el logaritmo de los patrones π posibles como sigue:

$$H_s[P] = -\frac{1}{\log_2 D!} \sum_{i=1}^{D!} p(\pi_i) \log_2 p(\pi_i)$$

Como se usa el logaritmo en base dos, está expresada en bits. La FMP para el ejemplo es: $\vec{\pi} = \left\{p(\pi_1 = 01) = \frac{4}{6}, p(\pi_2 = 10) = \frac{2}{6}\right\}$. Con lo cual se tiene: $H_s[P] = -\frac{1}{\log_2(2)} \left(\frac{4}{6} \log_2 \frac{4}{6} + \frac{2}{6} \log_2 \frac{2}{6}\right) = 0,9183$.

La medida de complejidad asociada (Lamberti et al., 2004; Ribeiro et al., 2012) está definida como el producto de la entropía de Shannon normalizada $H_s[P]$ y una medida de desequilibrio $Q_j[P, P_e]$, la denominada distancia de *Jensen - Shannon*, y que mide el grado de separación entre la distribución efectiva $P = \vec{\pi}$ de los patrones ordinales y una hipotética distribución uniforme de éstos: $P_e = \left\{\frac{1}{D!}\right\}$. Formalmente:

$$C_{sj}[P] = Q_j[P, P_e] H_s[P]$$

$$Q_j[P, P_e] = \frac{S\left[\frac{P + P_e}{2}\right] - \frac{S[P]}{2} - \frac{S[P_e]}{2}}{Q_{max}}$$

$$Q_{max} = -\frac{1}{2} \left[\frac{D! + 1}{D!} \log_2(D! + 1) - 2 \log_2(2D!) + \log_2(D!) \right]$$

La última expresión es la distancia máxima entre P y P_e cuando uno de los componentes de P tiene probabilidad 1 y los demás, probabilidad 0. Con esto es posible investigar la dinámica del sistema de aprendizaje en el plano $\vec{H}_s \times \vec{C}_{js}$. Una situación de máximo / mínimo conocimiento-incertidumbre no es compleja; o sea, sencillamente existe completo orden o desorden; por ello muchas medidas de complejidad (C) asignan $C = 0$ cuando H es mínima o máxima. A través del plano de incertidumbre / complejidad $H \times C$, podemos crear el espacio de fase del sistema de aprendizaje de un aprendiente, en el que cada estado (sesión) es un punto en dicho plano; e investigar el movimiento / trayectoria hacia, por ejemplo, un punto atractor que representa al "hablante nativo" (ausencia de error) o bien una región de atracción que represente al nivel superior, una vez cruzado un punto crítico.

Cabe preguntarse por la significancia de las medidas. Si se calculan las medidas H_s^{ruido} y C_{js}^{ruido} para un proceso de ruido estas deberían dar cercanas a uno para la entropía y cercanas a cero para la complejidad. Se genera una muestra de ruido para cada sesión (de tamaño n igual al de la sesión). Luego, se comparan los vectores con los valores de entropía H_s con H_s^{ruido} y C_{js} con C_{js}^{ruido} para cada aprendiente mediante un test de permutación para tendencia central con dos muestras *independientes* [Boninni et al., 2014]. Si se rechaza la hipótesis nula de igualdad de medias (diferencia de medias igual a cero) entre los datos observados y los que provienen de datos sin estructura, entonces se concluye que los datos observados poseen información relevante.

6.2. Análisis de cambio de punto

"Change Point Analysis" es una metodología que sirve para estimar el punto de una serie en el cual cambian las propiedades estadísticas de dicha serie. Se usará el método no paramétrico propuesto por Dehling et al. (2018), útil para series de datos no independientes y asimétricos. Se supone que las observaciones $(X_n)_{n \geq 1}$ provienen del modelo $X_n = \mu_n + Y_n$, donde $(\mu_n)_{n \geq 1}$ es una secuencia de constantes desconocidas y $(Y_n)_{n \geq 1}$ es un proceso estacionario con media cero. Dadas las observaciones X_1, \dots, X_n , se desea testear la hipótesis nula de que el proceso es estacionario: $H_0: \mu_1 = \dots = \mu_n$ contra la alternativa H_1 de que existe un punto k en el tiempo en el cual la locación (mediana) cambia: $\exists k: k \in \{1, \dots, n-1\}$ tal que $\mu_1 = \dots = \mu_k \neq \mu_{k+1} = \dots = \mu_n$. Para cada punto k se calcula el siguiente estadístico basado en el estimador de Hodges-Lehmann² entre la serie segmentada entre dos grupos $[1, k]$ y $[k+1, n]$ y se elige el máximo en valor absoluto:

$$T_n = \sqrt{n} \max_{1 \leq k \leq n} \frac{k}{n} \left(1 - \frac{k}{n} \right) | \text{med}\{(X_j - X_i): 1 \leq i \leq k, k+1 \leq j \leq n\} |$$

El estimador rechaza la hipótesis nula para valores (negativos o positivos) grandes del estadístico. Se calculó la entropía de permutación H_s ($D = 3$) por medio de una ventana móvil de largo $w = 80$ para cada aprendiente. Ya que los datos de la serie creada por la

ventana móvil no son independientes, se aplicó el método explicado para descubrir un punto de cambio de locación³.

6.3. Análisis cuantitativo de recurrencias

Recurrence Quantification Analysis o *RQA*, por sus siglas en inglés (Eckmann et al., 1987; Webber y Zbilut, 2005; Marwan et al., 2007; Webber y Marwan, 2015; Webber, Ioana y Marwan, 2016; Huffaker et al., 2017) es una técnica que permite medir propiedades de atractores en el espacio de fase [phase space] reconstruido. Los atractores se definen en términos de regiones en el espacio (multi-)dimensional que son visitadas una y otra vez por el sistema dinámico. *RQA* mide el grado de nuevas visitas del sistema en dichas regiones y computa ciertas medidas basadas en dichas recurrencias. El método fue introducido por Eckmann et al. (1987). Supóngase una trayectoria de vectores en el espacio de fase. Los componentes de estos vectores son parámetros. La evolución del sistema está representada por una serie de estos vectores, que representan su trayectoria. Sin embargo, en los datos reales por lo general no todos los parámetros necesarios para construir el espacio de fase pueden ser medidos; es decir que a menudo se tiene a disposición solamente un observable medido de modo discreto. Pero, si el sistema depende de más parámetros, o sea, si existe en un espacio multi-dimensional, ¿cómo reconstruirlo a partir de una sola variable observada? La solución yace en el hecho de que todas las variables de un sistema dinámico determinístico están relacionadas. Por ende, si en el tiempo τ solamente se conoce el valor de una variable x , entonces otra medida de dicha variable en $t + \tau$ (τ representa el *embedding delay*) portará implícitamente información sobre otras variables del sistema. Entonces, conociendo τ , si se deben obtener, por ejemplo, tres variables que actúen en el sistema, es posible reconstruir un vector tridimensional $(x(t), x(t + \tau), x(t + 2\tau))$. En general tenemos: $y_t = (x(t), x(t + \tau), x(t + 2\tau), \dots, x(t + (m - 1)\tau))$; donde m es la dimensión y τ , el retraso. Cada y_t es un punto en el espacio de fase reconstruido. Luego se calculan las distancias entre dichos puntos (entre los puntos en $t = i$ y $t = j$): $\delta_{ij} = \|y_i - y_j\|$. Se pueden usar diferentes normas para la distancia: *Euclídea*, *Manhattan*, etc. Dos puntos se declaran recurrentes si la distancia calculada es menor a un cierto radio ϵ : $\|y_i - y_j\| < \epsilon$. Entonces un gráfico de recurrencia [*RP*, *recurrence plot*, por sus siglas en inglés] se basa en las siguientes condiciones en una matriz de recurrencias:

$$\mathbf{R}_{i,j} = \theta(\epsilon - \|y_i - y_j\|); y_i \in \mathbb{R}^m; i, j = 1, \dots, N.$$

En donde: N es el número de estados de la trayectoria y_i , ϵ es el radio de recurrencia, $\| \cdot \|$ es una norma para calcular la distancia y θ es la función *Heaviside*:

$$\begin{cases} \theta(x) = 0 & x > \epsilon \\ \theta(x) = 1 & x < \epsilon \end{cases}$$

Por lo tanto, para todos los estados recurrentes (todos los y_i que se encuentran en el vecindario definido por el radio ϵ), $\mathbf{R}_{ij} = 1$, y se representa mediante un punto oscuro; alternativamente el estado no recurre y $\mathbf{R}_{ij} = 0$, representado por un punto blanco. Ya que un estado siempre recurre con sí mismo porque la distancia es cero, $\mathbf{R}_{ii} = 1$, la diagonal de la matriz se llama línea de identidad [*line of identity*, *LOI*, por sus siglas en inglés]. La matriz es simétrica respecto a dicha diagonal: $\mathbf{R}_{ij} = \mathbf{R}_{ji}$. Supongamos ahora que se tienen dos sistemas dinámicos, cada uno representado por las trayectorias x_i e y_i en el mismo espacio m -dimensional. El gráfico de cros-recurrencias [*CRP*, *cross-recurrence plot*] es una

extensión bi-variada del *recurrence plot* que permite cuantificar comportamientos similares en la evolución de dichas trayectorias. En este caso las distancias se computan entre ambas trayectorias. Luego el gráfico se construye en modo análogo al anterior:

$$R_{ij}^{x,y} = \theta(\epsilon - \|x_i - y_j\|); x_i, y_i \in \mathbb{R}^m; i = 1, \dots, N; j = 1, \dots, M.$$

El gráfico de recurrencia ilustra ahora cuándo un estado de un sistema "recurre" con un estado del otro sistema. La diagonal principal se llama "línea de sincronización" [*LOS, line of synchronization*], en la cual no necesariamente los sistemas visitan los mismos estados. A continuación, se presentan algunas de las medidas que se han propuesto para cuantificar la matriz de recurrencias.

(1) Recurrencia [Recurrence Rate; RR]. Porcentaje de puntos que recurren en la matriz R, incluyendo o no a la *LOI* (diagonal principal). En el caso del CRP debe estar la diagonal principal si se desea investigar el grado de sincronización entre ambos sistemas.

$$RR = \frac{1}{N^2} \sum_{i,j=1}^N R_{i,j}$$

(2) Determinismo [Determinism, DET]. Porcentaje de puntos recurrentes que forman líneas diagonales, incluyendo o no a la *LOI*:

$$DET = \frac{\sum_{l=d_{min}}^N l P_D(l)}{\sum_{l=1}^N l P_D(l)}$$

Donde: d_{min} es la longitud mínima para que la secuencia sea considerada como línea diagonal en el conteo, l es la longitud de cada línea y $P(l)$ es el *histograma* de frecuencias absolutas de las líneas diagonales. Para $d_{min} = 1$, $DET = 1$. DET mide la predictibilidad del sistema: cuanto más alto, más se parece a un sistema periódico. En sistemas caóticos las diagonales son cortas; en sistemas periódicos, largas; y en el ruido aleatorio no existirían en principio, salvo algunas de corta longitud que aparecen por azar. Observar que: $\sum_{l=1}^N l P_D(l) = \sum_{i,j=1}^N R_{i,j}$.

(3) Línea diagonal máxima [maximal diagonal line length; D_{max}]. Es simplemente la longitud máxima de las líneas diagonales: $D_{max} = \arg \max_l P_D(l)$. Una línea diagonal de largo l indica que un segmento de una trayectoria (o sistema) está cerca durante l pasos de tiempo de otro segmento de la misma trayectoria (o de la trayectoria del otro sistema en CRP) que se encuentra en un tiempo diferente. Por ende, cuánta más divergencia, más cortas serán las líneas diagonales. Su inversa ($\frac{1}{L_{max}}$) es la medida llamada *DIV*, la cual cuantifica dicha divergencia.

(4) Línea diagonal promedio [average diagonal line length; $\langle D \rangle$]. Es la longitud media de las líneas diagonales. Es el tiempo promedio en el cual dos segmentos de una trayectoria están cerca uno del otro. Se interpreta como tiempo promedio de predicción.

$$\langle D \rangle = \frac{\sum_{l=d_{min}}^N l P_D(l)}{\sum_{l=d_{min}}^N P_D(l)}$$

(5) LAM [Laminarity]. Análogamente a *DET*, es el ratio entre los puntos recurrentes que forman parte de líneas verticales u horizontales y el número total de puntos recurrentes. El análisis separado de líneas horizontales y verticales cobra especial relevancia en el *CRP* ya que se observan comportamientos en dos sistemas diferentes. Se define como:

$$LAM_V = \frac{\sum_{l=v_{min}}^N lP_V(l)}{\sum_{l=1}^N lP_V(l)} ; LAM_H = \frac{\sum_{l=h_{min}}^N lP_H(l)}{\sum_{l=1}^N lP_H(l)}$$

Donde: $v_{min}; h_{min}$ es la longitud mínima para que la secuencia sea considerada como línea vertical / horizontal en el conteo, l es la longitud de cada línea y $P_V(l); P_H(l)$ es el *histograma* de frecuencias absolutas de las líneas verticales / horizontales. Está relacionada con periodos de *laminaridad* del sistema, es decir, con segmentos de tiempo en los cuales no hay intermitencia. La laminaridad indica por cuánto tiempo el sistema está "atrapado" expresando recurrencias, o sea en un mismo estado, desde el punto de vista del (sub)sistema vertical (en el eje vertical) o del (sub)sistema horizontal (en el eje horizontal).

(6) Línea horizontal (vertical) promedio [trapping time; TT]. Es la longitud media de las líneas verticales u horizontales. Indican en tiempo promedio en el cual el (sub)sistema vertical u horizontal se encuentra "atrapado" en un mismo estado de recurrencias. Si TT_V es alto, el sistema del eje vertical tiende a estar atrapado por tiempos prolongados en recurrencias con puntos del sistema en el eje horizontal en tiempos diferentes. Análogamente, para TT_H alto, el sistema en el eje horizontal está atrapado en el mismo comportamiento recurrente que coincide con el otro sistema en algún punto diferente de tiempo.

$$TT_V = \frac{\sum_{l=v_{min}}^N lP_V(l)}{\sum_{l=v_{min}}^N P_V(l)} ; TT_H = \frac{\sum_{l=h_{min}}^N lP_H(l)}{\sum_{l=h_{min}}^N P_H(l)}$$

La Figura 1 ilustra el gráfico de recurrencias para los datos de las sesiones de SONIA 2 y 12 (Tabla 2). En el gráfico se traza una línea donde hay un error. El "Array1" (serie del primer término de concordancia) está en el eje horizontal y el "Array2" (serie del segundo término de concordancia), en el vertical. Las franjas negras son "1" en la matriz de recurrencia; en cambio las blancas son "0". Dado que en el "Array1" hay tres errores, hay tres franjas blancas de ceros y en el "Array 2" hay cuatro. Obsérvese que en el primer gráfico hacia arriba hay una "franja doble" en el "Array2" ya que se trata de dos errores seguidos. Las líneas horizontales (verticales) de ceros significan que hay un error (falta de recurrencia) en el término i -ésimo del *Array* en cuestión, y que dicho término i -ésimo tampoco recurre con todos los términos j -ésimos del otro *Array*. La línea diagonal es la llamada "línea de sincronización". En la sesión 12 solamente se comete un error en el *Array1*. Por eso solamente hay una franja vertical. Es fácil ver que un alumno sin errores tendría el gráfico completamente pintado de negro. Para datos categóricos los gráficos del tipo "mosaico", como en el ejemplo, aparecen comúnmente.

Sesión 2	Array1: 0 3 0 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0
	Array2: 0 0 0 0 0 0 0 4.1 0 0 0 0 0 3.1 2.1 0 3.1
Sesión 12	Array1: 0 0 0 0 0 0 0 0 0 0 3 0 0 0 0 0 0 0
	Array2: 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0

Tabla 2. Datos de las sesiones 2 y 12 de SONIA

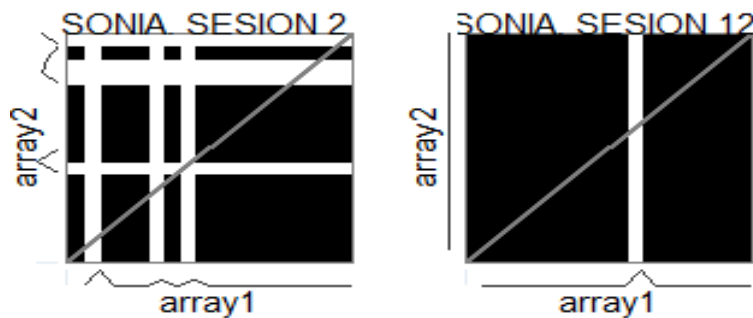


Figura 1. Gráficos de recurrencia para SONIA, sesiones 2 y 12

Con el objetivo de establecer la confianza con la cual una determinada medida logra detectar una transición en la dinámica del sistema o un desvío de esta, se hizo correr una ventana móvil (con o sin superposición) y se calcularon las medidas para cada ventana; precisamente para ser capaz de detectar dichos cambios. Marwan et al. (2013) sugirieron el siguiente método, para rechazar la hipótesis nula H_0 : "la dinámica del sistema no cambia en el tiempo". La estructura de la matriz de recurrencias se mantiene relativamente constante. Lo cual implica que la medida derivada de recurrencia M tiene una distribución con media $\mu(M)$ y varianza $\sigma(M)$ no nulas, aunque desconocidas. Solamente es aplicable a medidas que dependan de líneas (diagonales, horizontales, verticales). Se parte de los histogramas de frecuencias de las líneas $P(l)$. Habrá uno para cada ventana. Se los combina a todos para formar un histograma global de líneas $P_{global}(l)$. Luego se lleva a cabo un remuestreo a partir de dicho histograma global para crear un nuevo histograma $P^*(l)$ y se calcula la medida deseada M^* . Se repite el procedimiento B veces con el fin de obtener una distribución empírica de la medida y derivar los percentiles los percentiles $\frac{\alpha}{2}$ y $1 - \frac{\alpha}{2}$. Se define un intervalo de confianza del $100 - \alpha$ % como el rango entre dichos percentiles. La ventana se declara de dinámica diferente si el valor de la medida se ubica *fuera* del intervalo de confianza. Notar que la correspondiente medida se asigna a la ventana $w[i]$; con lo cual, si se detecta allí un cambio de dinámica, este habrá ocurrido presumiblemente en los datos originales entre los tiempos i e $i + w - (m - 1)\tau$ (si $m = 1$, $\tau = 1$ el intervalo es $[i, i + w]$).

7. ANÁLISIS DE DATOS

La Tabla 3 muestra las frecuencias absolutas y relativas (redondeadas al segundo decimal) de error, y el número de casos de concordancias por sesión y totales, por alumno. En las secciones siguientes se analizan los datos según cada método.

Aprendiente	SONIA			NATI			JAKO			MIRKA		
	Fa	C	Fr	Fa	C	Fr	Fa	C	Fr	Fa	C	Fr
1	2	4	0.5	12	30	0.4	6	26	0.23	12	80	0.15
2	6	18	0.33	10	16	0.62	10	52	0.19	13	39	0.33
3	4	19	0.21	5	21	0.24	1	30	0.03	21	64	0.33
4	14	42	0.33	2	13	0.15	6	28	0.21	15	62	0.24
5	11	42	0.26	6	18	0.33	8	41	0.2	8	34	0.24
6	7	25	0.28	18	47	0.38	2	30	0.07	10	47	0.21
7	3	28	0.11	13	37	0.35	9	37	0.24	30	81	0.37

8	7	37	0.19	19	38	0.5	8	32	0.25	7	35	0.2
9	5	26	0.19	7	27	0.26	8	51	0.16	27	74	0.36
10	1	19	0.05	14	37	0.38	5	18	0.28	17	52	0.33
11	6	34	0.18	7	29	0.24	7	47	0.15	22	48	0.46
12	1	18	0.06	9	28	0.32	1	32	0.03	19	75	0.25
13	--	--	--	6	22	0.27	5	27	0.19	--	--	--
14	--	--	--	3	27	0.11	0	13	0	--	--	--
TOTAL	67	312	0.21	131	390	0.33	76	464	0.16	201	691	0.29

Tabla 3. Frecuencias de error (absolutas y relativas) por alumno y sesiones. Fa: frecuencia absoluta; Fr: frecuencia relativa; C: casos

7.1. Plano entropía – complejidad

Se utilizó la respuesta categórica con los valores $\{0,1,2,3,4\}$ y se la consideró como cuantitativa discreta a los efectos de aplicar ambas medidas. Se calcularon H_s y C_{js} para cada sesión de cada aprendiente. Se estableció $D = 2$ ya que hay sesiones con pocos datos⁴. Se evaluó la significatividad de las medidas mediante el procedimiento detallado en párrafo anterior.

En la Figura 2 se muestra el plano entropía-complejidad para todos los aprendientes. Se ha ubicado en cada sesión su proporción de errores y se establecieron las siguientes franjas de frecuencia relativa de error: (i) $f_{error} \leq 0.20$; (ii) $0.21 \leq f_{error} \leq 0.35$; (iii) $f_{error} > 0.35$. Se ubicó el atractor "sin error" en el punto ($H_s = 0, C_{js} = 0$) y el del "error" en el punto ($H_s = 1, C_{js} = 0$). Se observa que para $H_s < 0.55$ la mayoría de los puntos coincide con la primera franja. En contraste, para $H_s > 0.80$ la mayoría de los puntos está en la tercera franja. Por lo tanto, se estableció el punto crítico (arbitrario) $f_{error} = 0.21$ para designar aquel punto más allá del cual se ingresa en el "cuenca de atracción" del atractor "no error"; y el punto crítico (arbitrario) $f_{error} = 0.35$ para designar aquel punto más allá del cual se ingresa en el "cuenca de atracción" del atractor "error". La región intermedia se designó de "transición" entre ambos regímenes. La Tabla 5 muestra los resultados para cada aprendiente. Se ilustra en la Figura 3 la ubicación de los valores para el aprendiente SONIA. En la tabla se muestran también los p-valores del test de permutación para dos muestras independientes entre los datos obtenidos y aquellos generados con ruido (primera línea): se rechaza la hipótesis nula de igualdad de medias para todos los aprendientes, con lo cual se considera que las medidas son informativas.

La Tabla 4 ubica las sesiones de los aprendientes en cada régimen. JAKO es el aprendiente con más sesiones en el régimen "NO_ERROR"; además, la sesión 14 corresponde al punto del atractor ("no error"); y posee más sesiones cercanas a dicho punto. No hay sesiones en la región "ERROR". Nótese además que hay tanto sesiones iniciales como finales en la región "NO_ERROR". Por lo tanto, JAKO solamente consolida su nivel de competencia, que es el más alto de los cuatro. SONIA, a pesar de haber hecho un curso de nivel A, evidencia solamente un par de sesiones en la región "ERROR"; además son las dos sesiones iniciales de la serie en dicha región. Por ende, SONIA está dejando el atractor del "ERROR". MIRKA se encuentra todavía en la región del "ERROR" en tres sesiones, tanto del principio, del medio y del final de la serie. Por lo tanto, la atracción al error es fuerte. NATI parece estar en una situación similar: con sesiones del principio, medio y final de la serie en la región del "ERROR", pero con una sesión más en la región "NO_ERROR", incluida la última. En resumen, el plano entropía-complejidad confirma el orden de competencia: JAKO > SONIA > NATI > SONIA.

	NO_ERROR	TRANSICIÓN	ERROR
SONIA	3,7,10,12	4,5,6,8,9,11	1,2
NATI	4,5,14	1,3,6,8,9,11,13	2,7,10,12
JAKO	1,2,3,6,9,12,14	4,5,7,8,10,11,13	
MIRKA	1,8	2,4,5,6,9,10,12	3,7,11

Tabla 4. Sesiones en cada región del plano entropía-complejidad

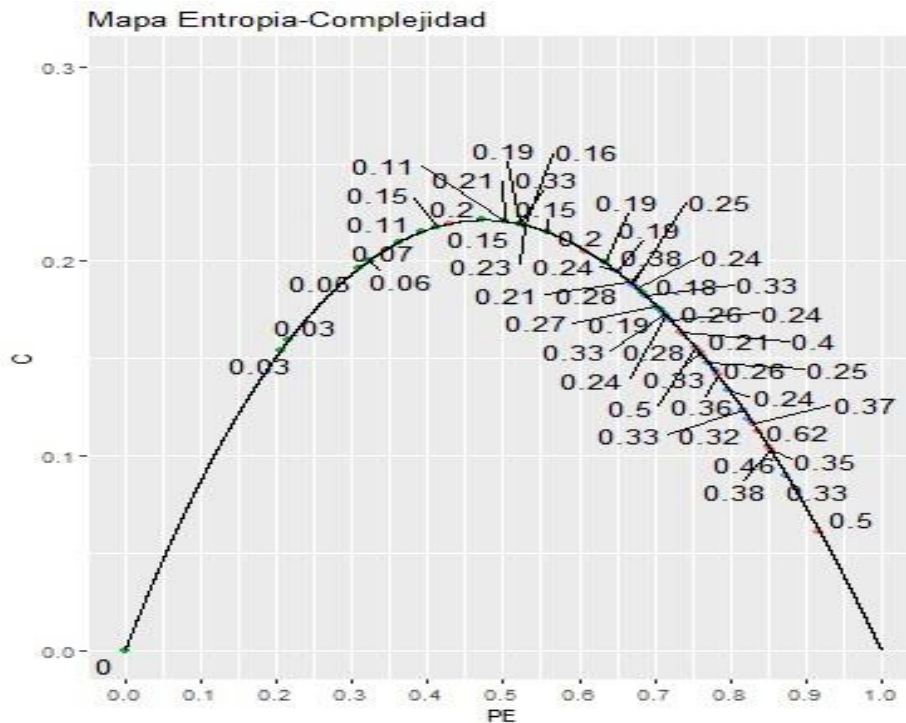


Figura 2. El plano entropía-complejidad para todos los aprendientes. Se señaló la proporción de error correspondiente a la sesión; EP: entropía de permutación H_s ; C: complejidad C_{js}

PE.Sonia	C.Sonia	PE.Nati	C.Nati	PE.Jako	C.Jako	PE.Mirka	C.Mirka
(0.011)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)	(< 0.001)
0.92	0.06	0.74	0.16	0.53	0.22	0.47	0.22
0.87	0.09	0.84	0.11	0.52	0.22	0.69	0.18
0.50	0.22	0.72	0.17	0.22	0.16	0.82	0.12
0.76	0.15	0.41	0.22	0.76	0.15	0.71	0.17
0.71	0.17	0.52	0.22	0.61	0.20	0.80	0.13
0.74	0.16	0.67	0.19	0.36	0.21	0.67	0.19
0.50	0.22	0.85	0.10	0.65	0.20	0.85	0.11
0.65	0.19	0.75	0.16	0.77	0.15	0.43	0.22

0.63	0.20	0.78	0.14	0.53	0.22	0.78	0.14
0.31	0.20	0.85	0.10	0.67	0.19	0.71	0.17
0.68	0.18	0.68	0.19	0.56	0.21	0.88	0.09
0.32	0.20	0.83	0.12	0.21	0.15	0.67	0.19
		0.70	0.18	0.71	0.18		
		0.39	0.21	0.00	0.00		

Tabla 5. Medidas de Entropía de Permutación (PE) y Complejidad (C) para cada aprendiente

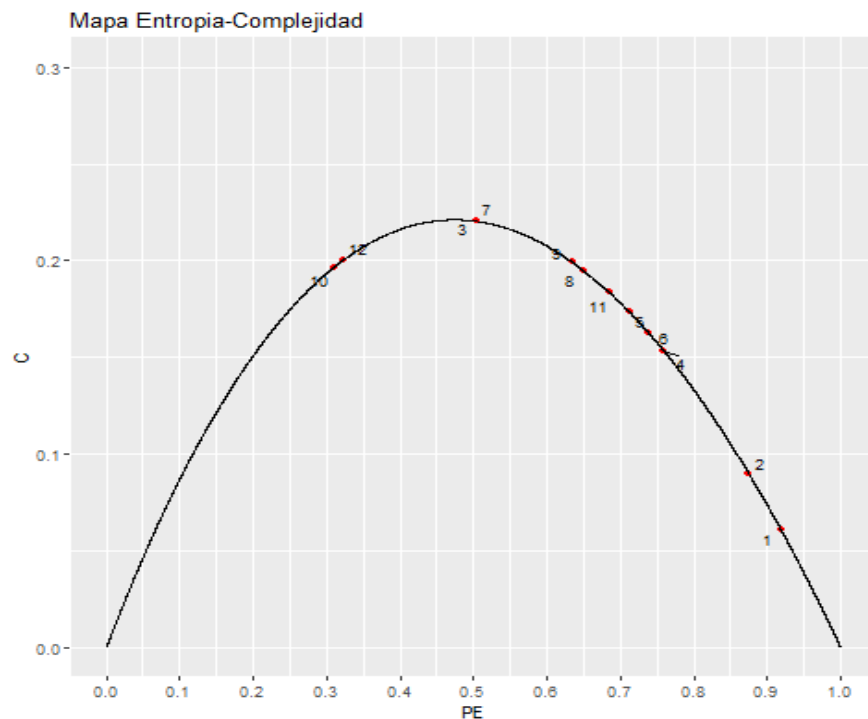


Figura 3. El plano entropía-complejidad para el aprendiente SONIA. EP: entropía de permutación H_s ; C: complejidad C_{j_s}

7.2. Análisis de cambio de punto

Las gráficas que siguen [Figura 4] muestran la ubicación del cambio de locación para cada aprendiente. Las líneas punteadas marcan la separación entre las sesiones. Notar que cada punto i representa la entropía de la serie temporal original en los puntos $[i, i + w]$. Los cambios en locación indican cambios en la dinámica (línea sólida vertical). Para SONIA el cambio se produce entre finales de la sesión 5 y mediados de la sesión 8 (ventana de puntos [112,192]); para NATI, durante las sesiones 5 y mediados de la 7 (ventana de puntos [82,162]); en JAKO, durante el periodo entre las sesiones 9 y principios de la 11 (ventana de puntos [280,360]). Por último, para MIRKA; entre las sesiones 2 y principios de la 3 (ventana de puntos [61,141]). Sin embargo, nótese que el p-valor del estadístico en MIRKA es $p = 0.047$. Recuérdese también que cuanto más baja H_s menos errores hay en la

ventana. Por lo tanto, se observa que SONIA y JAKO cambian a una mejor dinámica; pero NATI y MIRKA, a una peor.

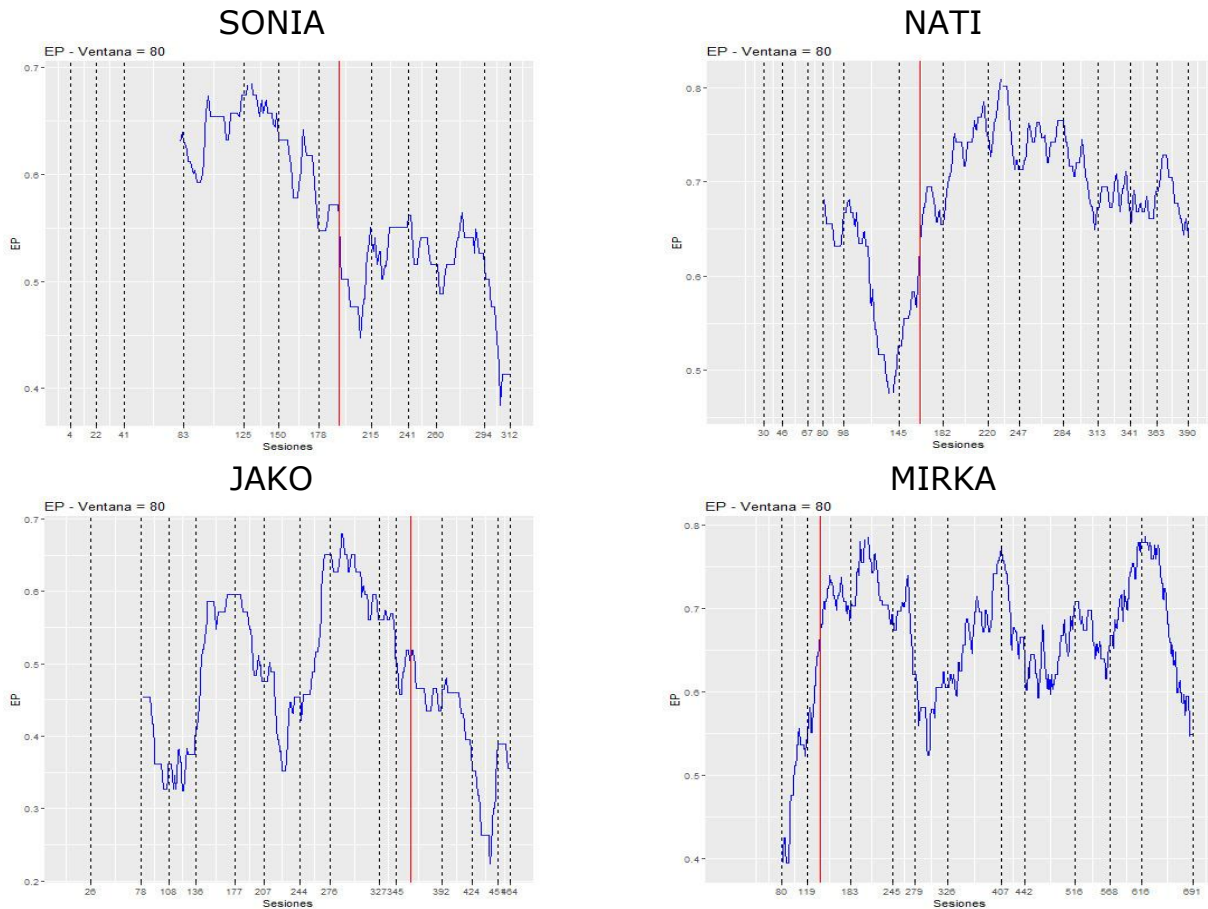


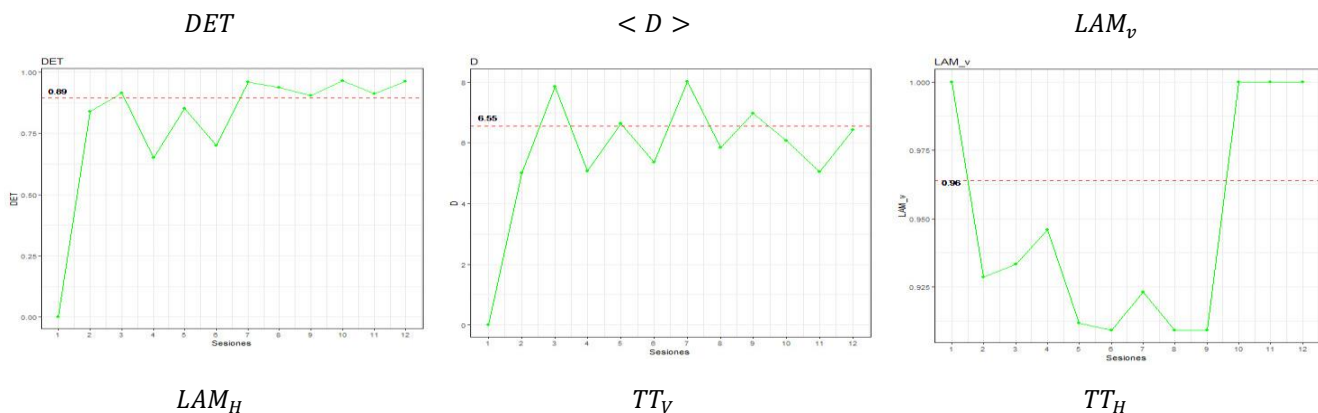
Figura 4. Ventana móvil de $H_s(D = 3)$ para cada aprendiz. Las líneas punteadas marcan separación entre sesiones y la roja el lugar de la serie donde se estima el cambio en la locación

7.3. Análisis cuantitativo de recurrencias

Para cada alumno, se creó una serie temporal bi-variada con categorías por tipo de error en el alfabeto usual $A = \{0, 1, 2, 3, 4\}$, discriminado por término de concordancia. Por ejemplo, la instancia *grande ciudades* [SONIA, sesión 2, línea 12] le corresponde el error $[3, 0]$ ya que hay un error de plural en el primer término y ningún error en el segundo término. Es decir que se consideraron dos sub-sistemas dinámicos en proceso de sincronización. Como el interés radicó en evaluar la sincronización entre instancias correctas en ambos términos, se sumó una pequeña cantidad constante de 0.1 a los datos del segundo término que no tuvieran la categoría "0". De este modo solamente se considerarían recurrentes los pares $[0, 0]$ y no los pares de repetición de error (por ejemplo $[3, 3]$). Se llevó a cabo un análisis de cros-recurrencias por alumno y para cada sesión, con los siguientes parámetros: $m = 1$, $\tau = 1$, $\epsilon = 0.0001$. El largo mínimo para considerar como línea (diagonal, vertical, horizontal) fue $l = 3$. Para detectar posibles cambios de dinámica se utilizó el método descrito en Marwan et al. (2013), el cual se detalló más arriba. Se consideraron las medidas: *DET*

[determinismo], $<D>$ [línea diagonal promedio], LAM_V [laminaridad vertical], LAM_H [laminaridad horizontal], TT_V ["trapping time" vertical], TT_H ["trapping time" horizontal]⁵.

Las Figuras 5 a 8 muestran los resultados de las medidas de recurrencia según las sesiones para los cuatro aprendientes. La línea punteada marca el límite superior del intervalo puntual de confianza del 95 % [el número en negrita especifica el cuantil correspondiente en la distribución de *bootstrap*], más allá del cual se considera un cambio en la dinámica. Lo primero que salta a la vista es que las medidas no necesariamente coinciden en cuanto al cambio de dinámica en una determinada sesión. Por lo tanto, se adoptó una estrategia de "consenso". Se discretizaron las medidas en: (i) "1" = "cambia la dinámica" (arriba del límite superior del I.C. 95 %), (ii) "0" = "no cambia la dinámica". Luego se sumaron las medidas para cada sesión. Se declaró un cambio de dinámica si dicha suma resultaba ≥ 4 . Por ejemplo, en la sesión 1 de SONIA solamente la medida LAM_V supera el límite superior, por ende, allí no habría cambio de dinámica. Entonces los conteos por sujeto quedan: (1) SONIA = [1 0 4 1 2 0 5 4 4 4 3 4]; NATI = [1 0 1 5 3 3 2 0 3 1 1 3 6]; JAKO = [3 3 6 1 1 6 1 0 3 2 3 5 2 4]; MIRKA = [6 4 1 3 2 5 0 5 0 2 0 4]. Según este criterio, SONIA cambia la dinámica en las sesiones 3,7,8,9,10,12; NATI, en las sesiones 4,14; JAKO, en las sesiones 3,6,12,14; y MIRKA, en las sesiones 1,2,6,8,12. Las medidas tienen valores más bajos para NATI y MIRKA, como es esperable. Un cambio en la dinámica del error *no* implica necesariamente que el alumno haga menos errores. Por ejemplo, en las sesiones 2 y 3 de MIRKA se comete un 33 % de error; sin embargo, la laminaridad y el "trapping time" verticales son mayores en la segunda sesión. Esto significa que el sistema del primer término de la concordancia se encuentra "atrapado" más tiempo en el atractor de lo "correcto". Además, es más alta la línea diagonal promedio, lo cual indica que en media segmentos de ambos términos de la concordancia están más cerca por más tiempo que en la sesión tres. En general, la suma de las medidas posee una alta correlación negativa con las proporciones de error. Según el test no paramétrico de *Spearman*: (i) SONIA: $\rho_s = -0.803$ ($p < 0.01$); (ii) NATI: $\rho_s = -0.651$ ($p < 0.05$); (iii) JAKO: $\rho_s = -0.824$ ($p < 0.001$); (iv) MIRKA: $\rho_s = -0.894$ ($p < 0.0001$).



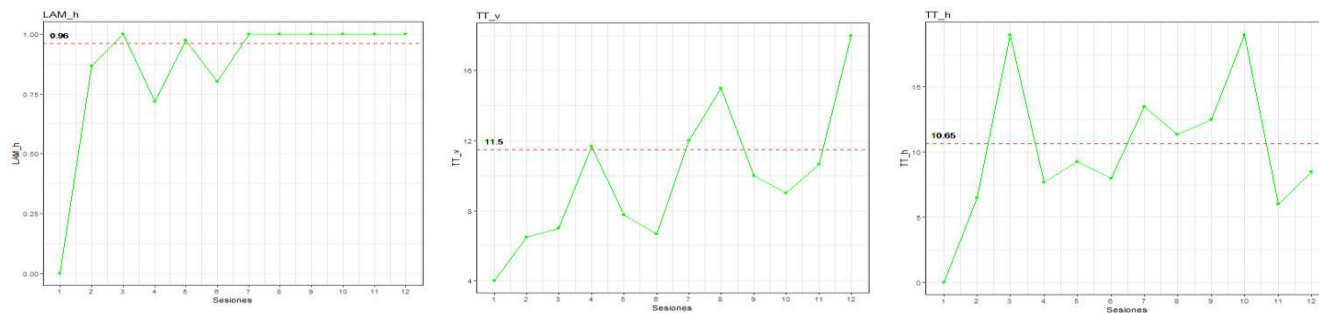


Figura 5. Medidas de recurrencia (SONIA)

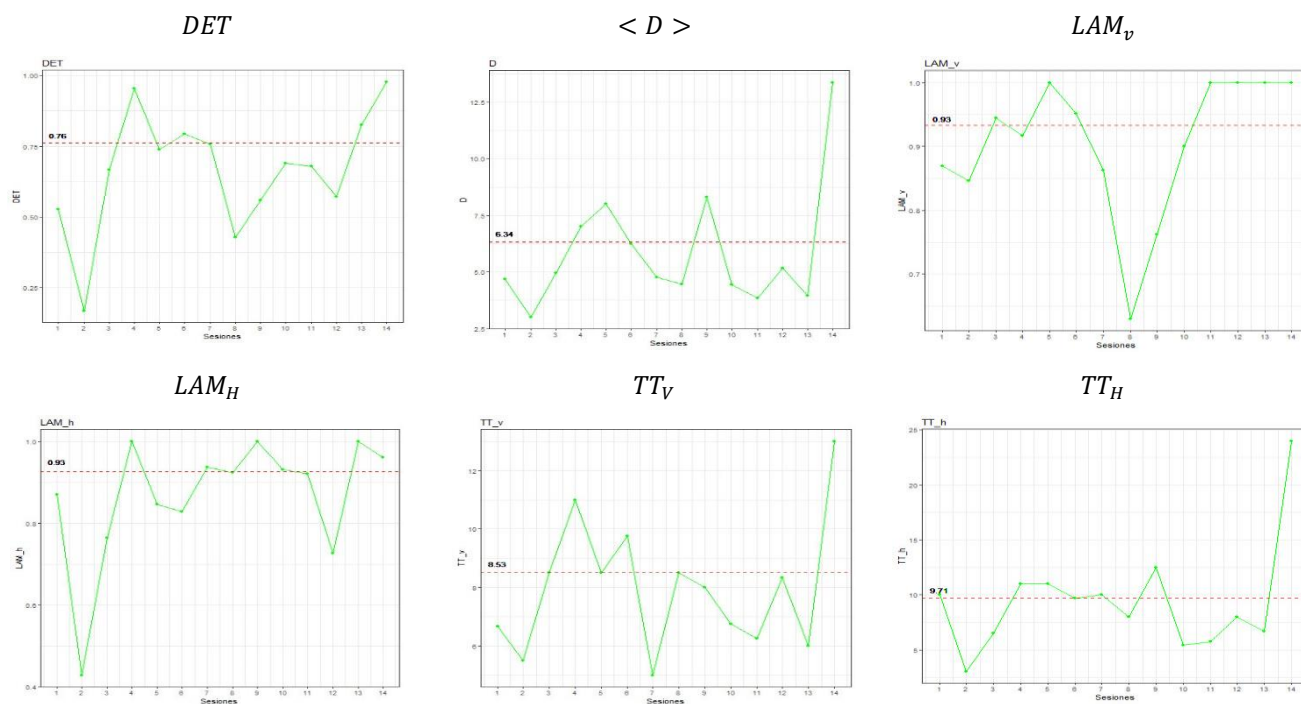
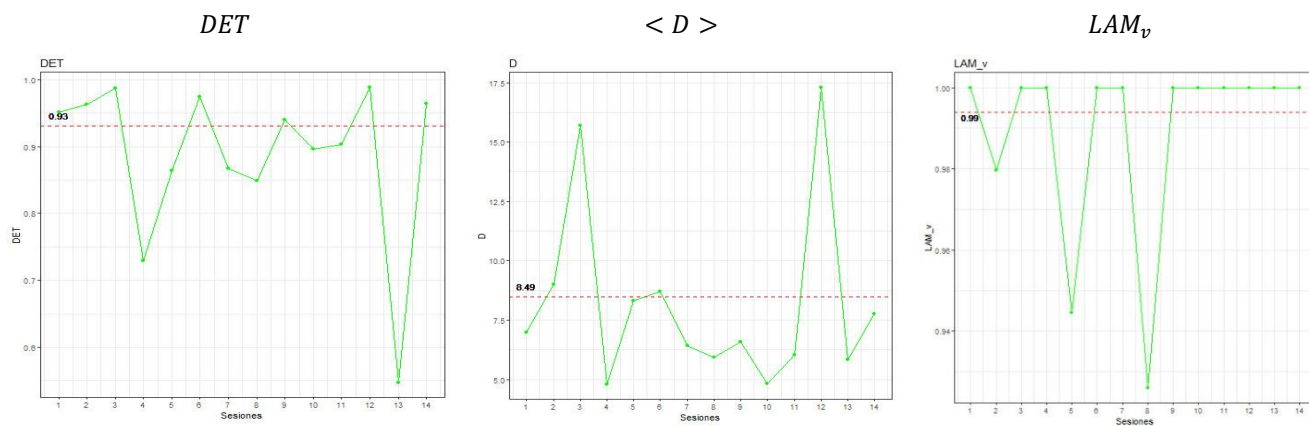


Figura 6. Medidas de recurrencia (NATI)



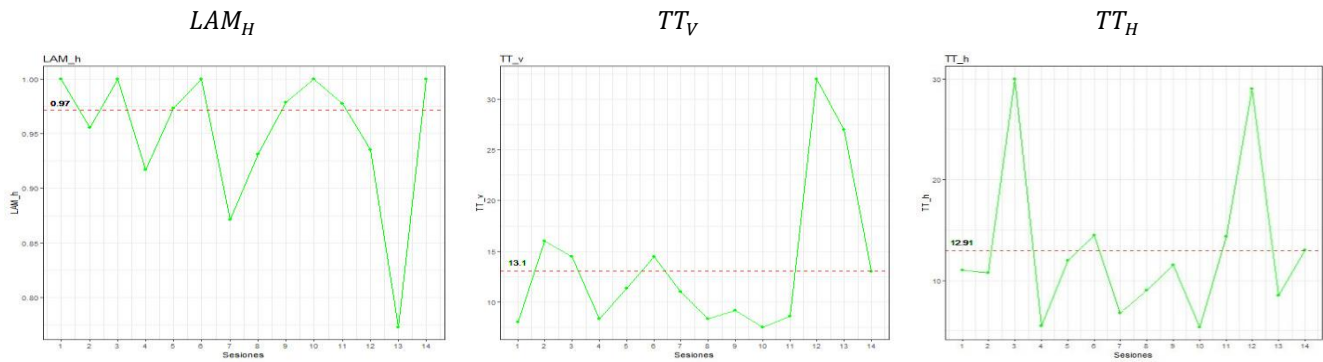


Figura 7. Medidas de recurrencia (JAKO)

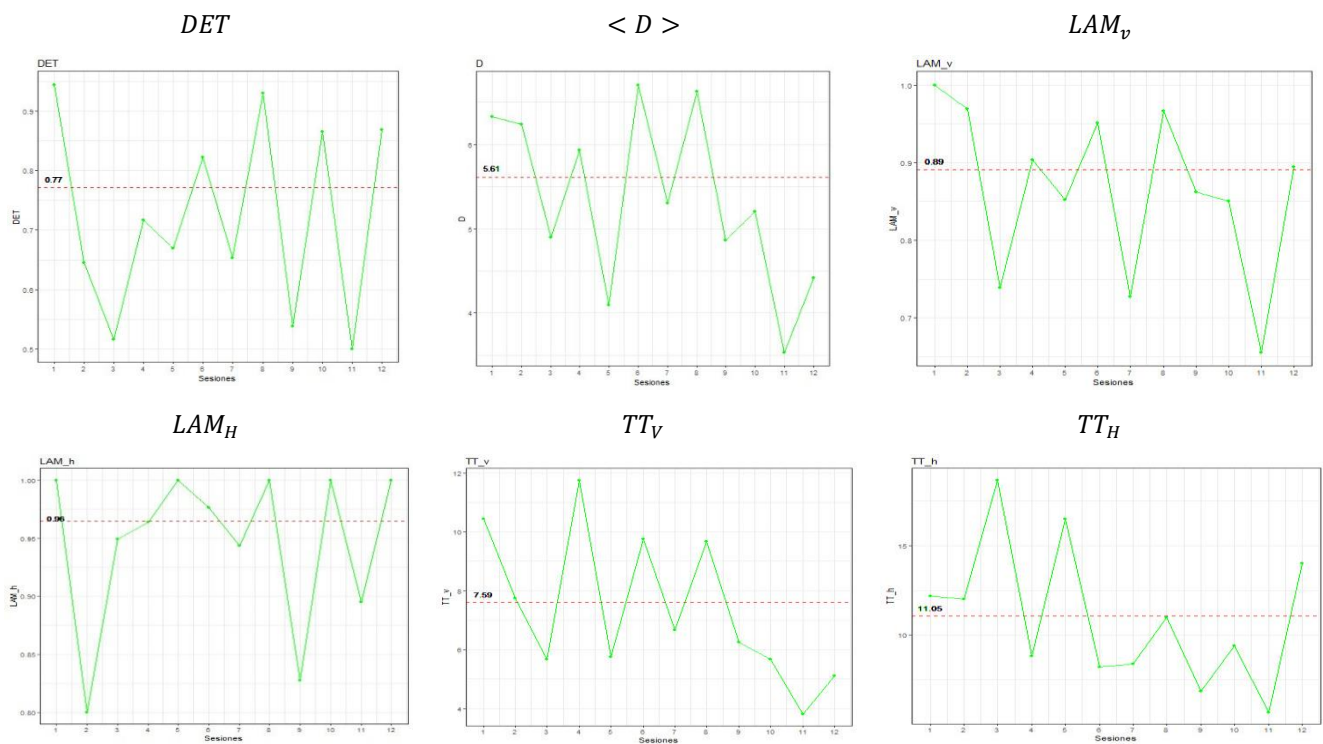


Figura 8. Medidas de recurrencia (MIRKA)

8. CONCLUSIONES

El siguiente Cuadro muestra los resultados surgidos de la aplicación de las técnicas, que responden a las preguntas hechas al principio del trabajo. En el plano *Entropía-Complejidad*, tanto JAKO como SONIA están pasando del régimen de TRANSICIÓN al de NO_ERROR, con JAKO más avanzado. En contraste, para NATI y MIRKA se hallan más sesiones en el tránsito del régimen ERROR al de TRANSICIÓN. El análisis con RQA reveló más sesiones con cambio de dinámica (por ejemplo, más períodos de laminaridad) en SONIA y MIRKA. O sea que MIRKA, aunque comete más errores, estos están más lejos en el tiempo durante más

sesiones que en NATI. El análisis de *change point* arrojó un cambio hacia la sesión 8 en SONIA, y hacia la 11 en JAKO, dirigiéndose ambos hacia un nivel superior de competencia. En NATI parece haber un cambio entre las sesiones 5-7, y en MIRKA en 2-3, pero en ambas hacia un peor nivel de competencia. El principal problema del análisis de este trabajo fue que las técnicas utilizadas requieren gran cantidad de datos, con lo cual se tuvo que recurrir a ventanas móviles de largo ancho. Esto dificultó a su vez la identificación más precisa de particulares sesiones donde se hallen las transiciones posibles.

Pregunta	Técnica	SONIA	NATI	JAKO	MIRKA
¿Regímenes?	Plano E-C	TRANS. NO_ERROR	→ ERROR → TRANS.	TRANS. NO_ERROR	→ ERROR → TRANS.
¿Sesiones con cambio de dinámica?	RQA	3,7,8,9,10,12	4,14	3,6,12,14	1,2,6,8,12
¿Transiciones?	Entropía	5-8	5-7	9-11	2-3

Tabla 6. Resultados de las técnicas

A diferencia de la mayoría de los estudios del área, en este trabajo se han aplicado técnicas cuantitativas de análisis del ámbito de los sistemas complejos a un corpus oral longitudinal de errores de concordancias plurales en español como lengua extranjera. La teoría de los sistemas dinámicos nos invita a considerar al lenguaje como fenómeno complejo y emergente. La no linealidad del aprendizaje no predice etapas fijas de adquisición. SONIA y JAKO realizan una transición hacia un atractor de error débil, pero se encuentra evidencia de cambio de dinámica tanto hacia las primeras sesiones como hacia las últimas, es decir que no hay un cambio lineal “de menor a mayor”. A través del lenguaje creativo el aprendiente crea variación (errores). Al principio dicho aprendiente puede no estar consciente de encontrarse atrapado en un atractor del error. Si el profesor provee *feedback*, puede ayudar a elevar el nivel de atención respecto del atractor del error. El primer paso es darse cuenta de que se está en la “cuenca de atracción” de un atractor del error. Si al principio, el estudiante no se auto-corrige; es preciso dejarlo continuar con la creación de lenguaje y ayudarlo con más *feedback*. Eventualmente los ciclos de repetición del error van a durar cada vez menos. Es decir, el atractor se debilitará. Se observará mayor nivel de laminaridad en instancias correctas y recurrencias de error más alejadas en el tiempo. Véase como ejemplo la sesión octava de SONIA respecto del plural de “profesor”. En este caso el aprendiente se auto-corrige inmediatamente. De hecho, las instancias en las que se usa “profesores” a lo largo de la sesión son: *los profesores [línea 170]* > *los profesores [línea 287]* > *los profesores [línea 294]* > *los profesor [línea 386]* > *los profesores [línea 390]*. O sea que la instancia incorrecta es simplemente una pequeña *perturbación*: el sistema vuelve de inmediato al atractor estable. Introducir *feedback* aquí no serviría de nada. Pero obsérvese en cambio la trayectoria de “alemanes” en SONIA: *los alemanos [sesión 3, línea 178]* > *alemanes ingenuos [sesión 3, línea 197]* > *los alemanos [sesión 4, línea 186]* > *los estudiantes alemanes [sesión 4, línea 193]* > *los alemanos [sesión 8, línea 40]* > *los alemanios [sesión 9, línea 67]* > *los alemanos [sesión 9, línea 72]* > *los alemanos [sesión 9, línea 127]*. En este caso parece haber inclusive bimodalidad: se salta del atractor correcto al del error con cuasi-periodicidad. Aquí el atractor del error es fuerte y se requeriría un *feedback*. No se sabe cuándo se producirá una bifurcación que produzca un cambio cualitativo en el sistema y ayude al aprendiente a avanzar en la

adquisición. Sin embargo, la variabilidad del error es condición para que dicha bifurcación se produzca. Las bifurcaciones están precedidas de un aumento de la incertidumbre (disminución de la entropía), como se observa en el análisis de cambio de punto. El profesor debería dejar que el aprendiente se equivoque; no corregir inmediatamente, e intervenir cuando se observen periodos cortos de laminaridad para los errores.

NOTAS

1 Una función de masa de probabilidad (FMP) asocia a cada tipo de símbolo del alfabeto $\mathcal{A} = \{A_1, \dots, A_n\}$ presente en una serie, una probabilidad $\vec{\pi} = \{p(\pi_1), \dots, p(\pi_n)\}$, de manera que sus valores estén en el rango $0 \leq p(\pi_i) \leq 1$ y sumen 1.

2 Dados dos grupos independientes de observaciones X_1, \dots, X_{n_1} y Y_1, \dots, Y_{n_2} el estimador de *Hodges-Lehmann* se define como la mediana de sus diferencias: $med\{(Y_j - X_i): 1 \leq i \leq n_1, 1 \leq j \leq n_2\}$.

3 Se utilizó el paquete *robts* de *R*.

4 Amigó y et al., (2008) proponen la relación $\mathcal{L} \geq 5D!$ (\mathcal{L} es el largo de la serie), con $D = 2$ se tiene $\mathcal{L} \geq 5(2!) = 10$. La única sesión con $n \leq 10$ es la primera de SONIA.

5 Se utilizaron las librerías de *R* "casnet" y "crqa" para los cálculos.

REFERENCIAS BIBLIOGRÁFICAS

- Alarcón, I. (2011). Spanish grammatical gender under complete and incomplete acquisition: early and late Bilinguals' linguistic behavior within the noun phrase. *Bilingualism: Language and Cognition*, 14(3), pp. 332-350.
- Ambadiang, T. (1999). La Flexión Nominal. Género y Número. En: I. Bosque y V. Demonte, eds., *Gramática descriptiva de la lengua española: entre la oración y el discurso, morfología*. Madrid: Espasa Calpe, pp. 4843-4914.
- Amigó, J., Zambrano, S. y Sanjuán, M. (2008). Combinatorial detection of determinism in noisy time series. *Europhysics Letters*, 83(6), 60005.
- Baba, K. y Nitta, R. (2013). Phase Transitions in Development of Writing Fluency from a Complex Dynamic Systems Perspective. *Language Learning*, 64(1), pp. 1-35.
- Bandt, Christoph y Bernd Pompe (2002). Permutation Entropy: A Natural Complexity Measure for Time Series. *Physical review letters*, 88(17), 174102.
- Caspi, T. (2010). *A Dynamic Perspective on Second Language Development*. Tesis doctoral, Universidad de Groningen.
- Corbett, G. (1991). *Gender*. Cambridge: Cambridge University Press.
- Corbett, G. (2006). *Agreement*. Cambridge: Cambridge University Press.
- Cover, T. y Thomas, J. (1991). *Elements of Information Theory*. New York: Wiley.
- De Bot, K., Lowie, W., Thorne, S. y Verspoor, M. (2013). Dynamic Systems Theory as a comprehensive theory of second language development. En: M^a del Pilar García Mayo, M. Junkal Gutierrez Mangado y M. Martínez Adrián, eds., *Contemporary Approaches to Second Language Acquisition*. Benjamins, pp. 199-220.
- De Bot, K., Lowie, W. y Verspoor, M. (2007). A Dynamic Systems Theory approach to second language acquisition. *Bilingualism: Language and Cognition*, 10(1), pp. 7-21.
- Dehling, H., Fried, R. y Wendler, M. (2018). *A Robust Method for Shift Detection in Time Series*. <http://martin-wendler.de/publications/>.
- Eckmann, J. P., Kamphorst, S. O. y Ruelle, D. (1987). Recurrence plots of dynamical systems. *Europhys. Lett. (EPL)*, 4(9), pp. 973-977.

- Ellis, N. y Ferreira-Jr., F. (2009). Construction Learning as a Function of Frequency, Frequency, Distribution, and Function. *The Modern Language Journal*, 93(3), pp. 370-385.
- Eskildsen, S. (2012). L2 negation constructions at work. *Language Learning*, 62, pp. 335-372.
- Huffaker, R., Bittelli, M. y Rosa, R. (2017). *Nonlinear Time Series Analysis with R*. Oxford University Press-
- Jiménez-Ramírez, J. (2020). El español es un caos: apuntes sobre las implicaciones de los sistemas complejos para el aprendizaje de L2 más allá de las redes lingüísticas. *Revista Nebrija de Lingüística Aplicada a la Enseñanza de Lenguas (RNAEL)*, 14(28), pp. 132-150.
- Lamberti, P. W., Martin, M. T., Plastino, A. y Rosso, O. A. (2004). Intensive entropic non-triviality measure. *Physica A: Statistical Mechanics and its Applications*, 334(1), pp. 119-131.
- Larsen Freeman, D. (2006). The Emergence of Complexity, Fluency, and Accuracy in the Oral and Written Production of Five Chinese Learners of English. *Applied Linguistics*, 27(4), pp. 590-619.
- Larsen Freeman, D. (2011). A Complexity Theory Approach to Second Language Development/Acquisition. En: D. Atkinson, ed., *Alternative approaches to second language acquisition*. New York: Routledge, pp. 48-72.
- Larsen Freeman, D. (2012). On the roles of repetition in language teaching and learning. *Applied Linguistics Review*, 3.
- Larsen Freeman, D. (2015). Complexity Theory. En: B. Van Patten y J. Williams, eds., *Theories in second language acquisition: an introduction*, 2da ed. Routledge, pp. 227-244.
- Larsen Freeman, D. y Cameron, L. (2008). *Complex Systems and Applied linguistics*. Oxford: Oxford University Press.
- Lowie, W. (2013). Dynamic Systems Theory Approaches to Second Language Acquisition. En: C. A. Chapelle, ed., *The Encyclopedia of Applied Linguistics*, Oxford: Wiley-Blackwell.
- Lowie, W., Plat, R. y De Bot, K. (2014). Pink Noise in Language Production: A NonLinear Approach to the Multilingual Lexicon. *Ecological Psychology*, 26(3), pp. 216-228.
- Mac Whinney, B. (2020). *The Childes Project: Tools for Analyzing Talk, Part 1 & 2* [pdf]. <https://talkbank.org/manuals>
- Marwan, N., Romano, M. C., Thiel, M. y Kurths, J. (2007). Recurrence plots for the analysis of complex systems. *Physics Reports*, 438, pp. 237-329.
- Marwan, N., Schinkel, S. y Kurths, J. (2013). Recurrence Plots 25 years later, gaining confidence in dynamical transitions. *Europhys. Lett.*, 101.
- R Development Core Team. (2017). R: a language and environment for statistical computing. Vienna, Austria. URL: <http://www.R-project.org>.
- Ribeiro, H. V., Zunino, L, Mendes, R. S. y Lenzi, E. K. (2012). Complexity-entropy causality plane: a useful approach for distinguishing songs. *Physica A: Statistical Mechanics and its Applications*, 391(7), pp. 2421-2428.
- Spoelman, M. y Verspoor, M. (2010). Dynamic patterns in development of accuracy and complexity: a longitudinal case study in the acquisition of Finnish. *Applied Linguistics*, 31, pp. 532-553.
- Van Geert, P. (1994). *Dynamic Systems of development: change between complexity and chaos*. London: Harvester Wheatsheaf.
- Van Geert, P. (1995). Growth dynamics in development. En: R. F. Port y T. J. Van Gelder, eds., *Mind as motion: Explorations in the dynamics of cognition*, pp. 313-338.
- Van Geert, P. (2003). Dynamic systems approach and modeling of developmental processes. En: J. Valsiner y K. J. Conolly, eds., *Handbook of developmental psychology*. Sage, pp. 640-672.
- Verspoor, M., Lowie, W. y Van Dijk, M. (2008). Variability in second language from a dynamic systems perspective. *The Modern Language Journal*, 92, pp. 214-231.
- Webber, C. L. (Jr.) y Zbilut, J. P. (2005). Recurrence quantification analysis of nonlinear dynamical systems. En: M. A. Riley y G. C. Van Orden, eds., *Tutorials in contemporary nonlinear methods for the behavioral sciences*. pp 26-94.
- Webber, C. L. (Jr.) y Marwan, N. (2015). *Recurrence Quantification Analysis*. Springer.

- Webber, C. L. (Jr.), Ioana, C. y Marwan, N. (eds.) (2016). *Recurrence Plots and Their Quantifications: Expanding Horizons*. Springer.
- Zanin, M., Zunino, L., Rosso, O. A. y Papo, D. (2012). Permutation Entropy and Its Main Biomedical and Econophysics Applications: A Review. *Entropy*, 14(8), pp. 1553-1577.

Estudio comparativo de métodos de transcripción para corpus orales: el caso del español

Comparative study of transcription methods for spoken corpus: the case of Spanish

Marimar Rufino Morales

Université de Montréal, Canadá

Mdm.rufino.morales@umontreal.ca

RESUMEN

Los avances tecnológicos han propulsado la metodología de investigación en transcripción. Los programas para corpus lingüísticos basados en modelos estadísticos y de aprendizaje profundo han mejorado las fases de alineación y anotación. En cambio, cuando se trata de transcribir el material, la carga interpretativa y la propia naturaleza de las conversaciones obstaculizan la automatización del proceso. De esta manera, la transcripción de entrevistas destinadas al estudio de la lengua oral se sigue haciendo con un reproductor y un teclado, y puede convertirse en uno de los aspectos más largos del procesamiento de datos. Sin embargo, en otros contextos profesionales, el reconocimiento automático del habla se emplea para transcribir de forma eficaz gracias a la colaboración humano-computadora. Las técnicas y estrategias difieren, pero todas tienen en común que estabilizan las fluctuaciones de las herramientas informáticas y son más rápidas que otros métodos. En este estudio se ha utilizado una de ellas, el reablado *off-line* con las entrevistas del Corpus oral de la lengua española en Montreal. Se ha medido el tiempo empleado, así como la precisión y se ha comparado con el reconocimiento automático del habla y con la mecanografía. El reablado *off-line* ha permitido el uso de un programa automático de dictado en su estado actual como herramienta para potenciar la transcripción de entrevistas en menos tiempo y con menos errores.

Palabras clave: reablado, transcripción, reconocimiento automático del habla, programa de dictado, corpus oral

ABSTRACT

Technological advances have propelled the research methodology in transcription. Language corpus tools based on statistical models and deep learning have improved the alignment and annotation phases. However, when it comes to transcribing the material, the conversation's interpretive load and nature themselves hinder automation of the process. That is why interviews used for studying spoken language are still transcribed with a player and keyboard, which can constitute one of the most time-consuming aspects of data processing. In other professional contexts, automatic speech recognition is used to transcribe effectively through human-computer collaboration. The techniques

and strategies may differ, but they all stabilize fluctuations in computing tools and are faster than other methods. In this study, the off-line respeaking method was used to transcribe the interviews of the Spoken Corpus of the Spanish Language in Montreal. Transcription times and accuracy were measured and compared with automatic speech recognition and typing. Off-line respeaking, using automatic speech-to-text software in its current state, proved to be the fastest and most error-free method for transcribing interviews.

Keywords: respeaking, transcription, automatic speech recognition, speech-to-text software, spoken corpus

1. INTRODUCCIÓN

Los avances tecnológicos de las últimas décadas han propulsado la metodología de investigación en todos los ámbitos. También en lingüística, los corpus orales, herramienta esencial para el estudio del lenguaje humano, han aligerado las ineludibles fases previas de recopilación, transcripción y almacenaje de material. Alentados por el rendimiento que ofrecen numerosos ingenios de uso diario, no pocos investigadores han sometido las grabaciones de las entrevistas a uno o varios programas automáticos de dictado para llevar a cabo su transcripción. La gran oferta de programas para corpus lingüísticos donde se amalgaman transcripción, alineación, anotación (Kreuz y Riordan, 2018), identificación y análisis de patrones fomenta la expectativa de automatizar la transcripción ortográfica del material.

La Inteligencia Artificial y el aprendizaje profundo han mejorado, de manera insólita, la precisión de aquellos primeros programas de dictado que comenzaron a inundar nuestros mercados a principios de los 90. Su aplicación en el reconocimiento automático del habla (RAH) propulsó, a finales de 2016, un punto de ruptura: la paridad con el humano (Xiong et al., 2016). En este proceso de reconocimiento automático, basado en las redes neuronales artificiales, la voz se convierte en texto de forma casi instantánea y, sin embargo, intervienen innumerables operaciones de tratamiento del lenguaje natural. Es preciso analizar la señal, procesarla, codificarla, sintetizarla (Li Deng y Yang Liu, 2018). Sin el desarrollo de la microelectrónica (que aumentó la potencia y miniaturización de procesadores y memorias) y de la informática (que favoreció la expansión de las técnicas de comunicación y la interconectividad de las redes), nada hubiera sido posible (Mariani, 2002). Casi un siglo de numerosos experimentos e ingentes esfuerzos de colaboración internacional separan a los asistentes virtuales como Siri o Cortana del vocoder, aquel sintetizador de voz creado en los laboratorios de Bell (Dudley, 1958) y presentado en la Exposición Universal de Nueva York en 1939.

Unos datos tan prometedores merecen ser aclamados, pero también han de ponerse en perspectiva; de hecho, las transcripciones destinadas al estudio de la lengua hablada las siguen haciendo los humanos (Durand, 2017). A las numerosas decisiones que hay que tomar durante el proceso de transcripción, y que lo alejan de un acto puramente mecánico (Ochs, 1979), se suman las limitaciones propias de esta tecnología de RAH. Por una parte, el entrenar modelos acústicos con aprendizaje profundo nos proporciona datos falsamente reales (Saon et al., 2017). Por ejemplo, si en el laboratorio 'enseñamos' a un programa de

RAH abasteciéndolo con cuarenta entrevistas semidirigidas extraídas de un corpus que contiene sesenta, realizadas todas ellas por la misma persona, quien a su vez ha formulado siempre más o menos las mismas preguntas, probablemente que la transcripción de las veinte entrevistas restantes con ese mismo programa una vez entrenado, sea excelente. Por otra, a pesar de que muchos científicos trabajan intensamente para mejorar las tecnologías del habla (Zweigenbaum et al., 2020), ante la conversación espontánea o la variación, los programas de RAH aún no son estables (Ravanelli et al., 2018). En definitiva, la transcripción automática de la lengua hablada sigue enfrentándose a las limitaciones de la tecnología. Programas tales como *Transana*, *Soundsciber*, *Transcriber* o *Audacity* funcionan bien como herramientas para etiquetar la transcripción, pero para transcribir, todavía no han suplantado al humano (Revuelta Domínguez y Sánchez Gómez 2005). Los de mayor utilidad son aquellos que, basados en modelos estadísticos y de aprendizaje profundo, ayudan en las fases de alineación y anotación de corpus (Yadav et al. 2018). También en español estos programas se ocupan menos del primer nivel de representación, la transcripción ortográfica, si bien automatizan con éxito la codificación, por ejemplo, de características suprasegmentales de la prosodia, tales como la entonación para la transcripción fonética (Elvira-García et al., 2015).

Y, sin embargo, tenemos constancia del uso eficiente y eficaz de la tecnología de RAH y de los programas de dictado automático, para transcribir, en varios contextos profesionales donde, desde hace dos décadas, se han venido desarrollando métodos basados en la colaboración humano-computadora. Las técnicas y estrategias de dictado difieren según el contexto, pero todas tienen en común que estabilizan las fluctuaciones de las herramientas informáticas empleadas. Nos hemos centrado en el reablado, que se usa en televisión (Romero-Fresco, 2011), por el paralelismo entre la oralidad que está presente en la mayoría de los escenarios o situaciones de los medios audiovisuales y la que caracteriza los datos que se recopilan para crear un corpus de la lengua hablada. Hemos empleado los programas de dictado con una adaptación del reablado, logrando, al igual que en televisión, optimizar el proceso de transcripción (Rufino Morales, 2020). Para abordar la cuestión, hemos confrontado el reablado con los otros métodos de transcripción de corpus orales para la investigación lingüística. Presentamos aquí los resultados de los datos recopilados durante el primer taller de reablado *off-line* que se organizó el pasado mes de enero en la Sección de estudios hispánicos de la Universidad de Montreal. Se ha comparado el reablado con el reconocimiento automático del habla (RAH) y con la mecanografía, en cuanto al tiempo empleado, así como la precisión resultante al transcribir entrevistas del Corpus oral de la lengua española en Montreal (COLEM), (Pato dir.). Es un corpus fruto de una serie de entrevistas semidirigidas grabadas en entorno natural, de una hora de duración aproximadamente y estructuradas en función de un protocolo de encuesta común. El estilo que recoge el COLEM es el conversacional y familiar (habla espontánea). Refleja la situación del español en la Región Metropolitana Montreal, única en el mundo, por el contacto mantenido con el francés y el inglés, pero también por el contacto de todas las variedades del español, sin que una se imponga sobre las demás. En este contexto se encuentra, además, la mayor concentración de latinoamericanos de todo Canadá (Pato, 2017).

Este trabajo forma parte de una investigación cuyo objetivo es describir el perfil *ad hoc* del reablador *off-line* para optimizar la transcripción de grabaciones, al menos hasta que los programas informáticos de reconocimiento automático del habla permitan cederles el testigo. Primero revisaremos en qué ámbitos se transcribe de forma eficaz con ayuda de programas de dictado; comprobaremos que la investigación cualitativa también ha intentado usarlos sin llegar a adoptarlos; finalmente, ofreceremos una prueba de que es

posible utilizarlos para potenciar la transcripción de corpus orales de la lengua hablada siempre y cuando se empleen las estrategias adecuadas. De seguro, las aproximaciones de este manuscrito retendrán la atención de otros investigadores que trabajan con textos orales.

2. TRANSCRIPCIÓN CON UN PROGRAMA DE DICTADO

Convertir la voz a texto es una necesidad presente en todas las épocas y en distintas esferas de la actividad humana, ya sea de forma intralingüística o interlingüística. No obstante, desde el punto de vista diatópico, las herramientas (o la forma de usarlas) difieren. Si optamos por un método al alcance de todos, un reproductor y un teclado, nos enfrentaremos al desfase entre la velocidad del habla y la de la escritura. “La velocidad de habla o tempo de elocución es la rapidez con que una persona articula las palabras a lo largo de su discurso. Para determinar la velocidad, se computa la cantidad de palabras que emite en un periodo de tiempo. El resultado se expresa, generalmente, en palabras por minuto; en español se calcula que una velocidad normal oscila entre las ciento cincuenta y las doscientas palabras por minuto” (AA.VV. 2008).

La velocidad óptima de producción oral se sitúa entre ciento setenta y ciento noventa palabras por minuto (Rodero Antón, 2016), pero el habla espontánea no se ciñe a este canon, pudiendo ser más elevada. En cuanto a la velocidad de escritura, alguien sin formación teclea entre quince y veinticinco palabras por minuto (Ainsworth, 1988); la media de una persona experimentada se sitúa entre ochenta y noventa palabras por minuto (Moro Vallina, 2010: 7). Además de esta diferencia entre la velocidad a la que hablamos y la velocidad a la que escribimos, al transcribir una grabación habrá que detenerla, retroceder, volver a ponerla en marcha, escuchar de nuevo, etc.

Podemos sustituir el teclado por el RAH, lo que además restará subjetividad a la transcripción (Tatham y Morton, 2005: 372). Pero entonces, tendremos que enfrentarnos a los errores provocados por las variantes geográficas (Winata et al., 2020) o sociodemográficas; también habrá que borrar y reescribir una representación más cercana a la actuación del hablante cada vez que el programa de dictado le haya atribuido una correspondencia que figura en su léxico a palabras fragmentadas, mal pronunciadas, solapadas o inexistentes. De manera que la automatización del proceso termina siendo más tediosa que la clásica transcripción mecanografiada, sobre todo si sumamos el tiempo que puede tomar aprender las nociones básicas de la herramienta que haya decidido emplearse. La comunidad científica es unánime: fuera del laboratorio, la tecnología de RAH aún no está a punto para convertir con exactitud la voz en texto en situaciones reales, donde, por ende, los ruidos y la reverberación afectan a la calidad de la señal de voz emitida por el hablante (Lu et al., 2020).

Una forma de estabilizar los resultados del RAH consiste en utilizar estrategias de repetición que, emulando el modo de corrección privilegiado en el diálogo humano-humano (Brinton et al., 1986), pero aplicado a la comunicación humano-computadora, consiguen optimizar la transcripción en tiempo real con ayuda de un programa de dictado. Es lo que acometen los siguientes métodos.

2.1 El reablado

El reablado surgió para paliar las limitaciones de la tecnología de RAH a la hora de subtitar en tiempo real en televisión (Utray Delgado et al., 2015; Lambourne, 2007). El reablador, instalado en un entorno exento de ruido, a la vez que escucha la emisión a subtitar, la repite o parafrasea verbalmente a través de un micrófono conectado a un programa de dictado, de manera que genere una transcripción legible de forma automática. El texto resultante es enviado al codificador de subtítulos (Brousseau et al., 2003) y la emisión se hace accesible, para aquellas personas que no pueden oírlo, casi simultáneamente.

La subtitulación en directo consiste en emitir la transcripción del contenido sonoro de forma simultánea a la difusión de un programa audiovisual. Los subtítulos enviados en directo pueden haber sido preparados con antelación o en tiempo real (ACR/CAB 2012). Deben incluir además de las palabras, la prosodia (entonación, acento, ritmos), los efectos sonoros, las señales musicales y cualquier otra información de la banda sonora pertinente, de modo que la lectura de los subtítulos y la escucha del audio proporcionen una comprensión análoga de dicho programa (Ivarsson, 1992).

Los subtítulos y la audiodescripción son instrumentos vitales para garantizar el derecho de acceso a la comunicación audiovisual de todas las personas por igual (ONU, 1994); la aplicación de dicho principio de igualdad queda asegurada a través de medidas legales a nivel de cada país. Así, por ejemplo, desde 1995, el Gobierno de Canadá comenzó instando a los teledifusores tanto de lengua francesa como de lengua inglesa a aumentar el contenido subtulado, hasta que en 2007, los obligó a subtitar el 100% de la programación (CRTC, 2007).

Los primeros subtítulos en directo para televisión se enviaron por teletexto; para su confección, se probaron distintas maneras y también varios teclados de estenotipia (Hawkins y Robinson, 1979; Tanton, 1979). Por aquel entonces, se trataba del método más extendido para transcribir en tiempo real. Pero la estenotipia necesita una instrucción intensiva de más de dos años. Por otra parte, existen distintos sistemas, máquinas y programas informáticos con sendos métodos, difícilmente intercambiables (Manrique Fuero, 2016). Por ejemplo, la compañía americana Stenograph (www.stenograph.com) comercializa varias máquinas de estenotipia: Diamante, Élan Mira, Stentura. Familiarizarse con un teclado toma mucho tiempo. En Quebec, la formación profesional de transcritores se da en l'École de sténographie judiciaire de Québec. Los estudiantes deben procurarse una de las dos máquinas sugeridas con la que podrán aprender el programa *Case CATalyst*. Pero al cabo de doscientas setenta horas de clase y noventa más de prácticas laborales, los diplomados solo habrán utilizado un teclado y un programa, y difícilmente podrán trabajar con otro.

De manera que, para subtitar en directo en televisión, desde el primer momento se buscaron alternativas más rentables que la estenotipia. Fue así como surgió la idea de usar los programas de RAH potenciados por la interacción humano-computadora (Damper et al., 1985; McCoy y Shumway, 1979). El éxito del reablado on-line a la hora de subtitar en directo y en tiempo real en televisión reside en que este método requiere menos personal (un reablador subtita un programa de treinta minutos; dos reabladores por turnos hacen accesibles ocho horas continuas de programación). Además, es una técnica que se domina en poco tiempo, salvando la brecha entre la formación en empresa (de dos a cuatro

semanas) y los módulos y cursos académicos disponibles (hasta seis meses) (Bernabé Caro et al., 2019).

Las ventajas que ofrece el reablado *on-line* a la hora de producir subtítulos en directo en televisión han extendido su uso a muchos otros ámbitos. En la actualidad, tanto en universidades, colegios, actos públicos, eventos de masas (conferencias, fórums, seminarios), reuniones, teleconferencias, podcasts, programas de radio, teatros, museos, iglesias (Moore, 2016; Romero-Fresco, 2012), el reablado *on-line* proporciona accesibilidad en tiempo real del contenido sonoro emitido en directo, de forma presencial o por internet (portátil, tableta, teléfono).

Asimismo, muchas de las subtituladoras que cuentan con la infraestructura de reablado *on-line* aprovechan sus recursos para acelerar la transcripción de programas 'enlatados' o en diferido (FCC, 2014). Para transcribir en tiempo real documentos audiovisuales grabados con antelación se emplean estrategias particulares. El objetivo no es mejorar la inteligibilidad del enunciado obtenido por el programa de dictado (que no va a ser leído por los televidentes de forma simultánea), sino reducir el trabajo de edición posterior de los subtítulos.

Las estrategias empleadas en el reablado *off-line* sirven para crear un documento lo más cercano posible a su versión definitiva. Este es el método que estoy usando en mi investigación, donde he adoptado una nomenclatura que distingue dos formas de reablado, según las circunstancias en las que se produce la transcripción: *on-line* y *off-line* (Lindsay y O'Connell, 1995; Ferber, 1991).

2.2. La escritura de voz

En otros contextos donde también se requiere la transcripción de un discurso hablado, como en el campo jurídico y en la administración pública, el primer método eficaz que se empleó para transcribir al mismo tiempo que se produce el discurso oral fue la taquigrafía. De él derivan la estenotipia y la escritura de voz. Mientras que la taquigrafía (también llamada estenografía) emplea trazos breves, abreviaturas y caracteres especiales para representar las letras, palabras y frases, la estenotipia utiliza un teclado reducido basado en las sílabas (Núñez Hidalgo y Ramos Villajos, 2010).

La escritura de voz (*voice writing*) es otra forma de transcribir; pero, en este caso, como su nombre indica, mediante la voz. Para que al dictar, el escritor de voz (*voice writer*) no moleste a las otras personas presentes en la reunión –y para atenuar el ruido ambiente–, se emplea un silenciador de voz. La idea le vino a Horace Webb, un transcriptor en el tribunal de Chicago que, en los años 40, insertó un micrófono en una caja de cigarrillos, y después en una lata de café. Amortiguaba el sonido mientras repetía y grababa lo que oía en el tribunal y conseguía transcribir palabra por palabra el discurso original. Los escritores de voz adoptaron los programas automáticos de dictado y un nuevo método de transcripción, que se extendió rápidamente en los años 2000. El oficio de transcriptor oficial (*verbatim reporter*) en Estados Unidos tiene dos especializaciones: la estenotipia y la escritura de voz. La escritura de voz cuenta con dos ramas con formación y acreditación propia: para transcribir en juzgados (*court and real time reporting*) o en televisión (*captioning*) (NVRA, 2008).

Así como el reablador, el escritor de voz también transcribe a partir de grabaciones, ya sea la defensa de un abogado, el veredicto de un juez o un testimonio. Por la naturaleza oficial de la tarea, el resultado final debe corresponderse palabra por palabra con el discurso original; ahora bien, ese mismo carácter resta espontaneidad a la interacción discursiva.

2.3. El dictado médico

Numerosas actividades en el ámbito de la salud requieren transcripción (Pollard et al. 2013). Hasta el giro tecnológico de los 90, los médicos y especialistas que necesitaban transcribir se lo encargaban a taquígrafos o administrativos. Desde entonces, cada vez más profesionales han ido integrando los programas de dictado, ya sea para elaborar la historia digital del paciente (Johnson et al., 2014), redactar un informe en radiología, endocrinología, psiquiatría o en patología quirúrgica, entre otros.

Por lo general, los médicos que recurren al dictado envían el texto generado a un administrativo para su revisión, antes de darle el visto bueno final. Pero en otros casos, el propio médico edita su transcripción generada con el programa de RAH. Sea cual sea el método empleado, la revisión es sumamente importante (Zhou et al., 2018) porque, a pesar de que los programas de dictado constituyen un buen punto de partida para transcribir (Edwards et al., 2017), en el ámbito médico también se constatan las mismas deficiencias del RAH que en otros contextos (Blackley et al., 2019). Por ende, la automatización del proceso plantea un problema ético: altas tasas de precisión no son sinónimo de resultados clínicamente seguros; en ocasiones han llegado a poner en peligro la seguridad del paciente (Hodgson y Coiera, 2016).

Estudios recientes centrados en la tipología de los errores durante el uso de aplicaciones de reconocimiento automático del habla en el campo médico confirman que: i) los mayores errores se producen en las conversaciones casuales (Chiu et al., 2018); ii) los resultados mejoran considerablemente con programas de vocabulario especializado y si se aprovechan bien los atajos (*shortcuts*) que ofrece el programa (Edwards et al., 2017) y iii) la precisión aumenta con una metodología estandarizada (Blackley et al., 2019).

Y, sin embargo, el uso de estrategias para mejorar los resultados de los programas de dictado no parece haber pasado de iniciativas aisladas. MacLean, Meyer et al. (2004: 115) emplearon “*an oral transcriptionist to act as an intermediary*” en la transcripción de entrevistas con *Dragon NaturallySpeaking*. El método *leasen and repeat* se usó en un experimento en la universidad de Auckland para entrenar a sus participantes a servirse de *Dragon* y a repetir verbalmente las grabaciones (Park y Zeanah 2005). Sirvió de inspiración, primero, para la técnica de transcripción vocal o *Voice Transcription Technique* (VTT) (Matheson, 2007) y, posteriormente, para otra llamada *Embodied Transcription* (ET), porque se encarna a los entrevistados al repetir sus palabras para acelerar el proceso de transcripción con un programa de dictado (Brooks, 2010).

3. LA TRANSCRIPCIÓN EN LA INVESTIGACIÓN LINGÜÍSTICA Y CUALITATIVA

La revisión de las actas de eventos internacionales como los organizados por la European Language Resources Association (www.elra.info) o la International Speech Communication Association (www.isca-speech.org), nos confirma que la transcripción de la lengua hablada, larga y costosa, es más difícil cuanto más detallado sea el nivel de transcripción (Adolphs y Knight, 2010) y que los grandes corpus se siguen transcribiendo de forma manual (Niemants, 2018; Gadet et al., 2012) o se completan de forma manual (www.isip.piconepress.com). Transcribir una hora de entrevista puede tomar entre cuatro y sesenta horas (Markle et al., 2011). Como en los demás contextos donde se requiere una transcripción *off-line*, también aquí se han puesto a prueba los programas de dictado para acelerar la tarea de transcripción de las grabaciones, laboriosa y lenta *per se*. Necesitan ser

más rápidos que las otras formas de transcribir al alcance; pero también, en su adopción como herramienta, juegan un papel decisivo la precisión de los resultados obtenidos.

Aunque la investigación cualitativa se ha ocupado más de la metodología para producir una transcripción que de la mecánica empleada durante el proceso, varios investigadores han estudiado el uso de programas de dictado para transcribir datos grabados de forma automática. Los resultados obtenidos contenían tantos errores que nadie ha llegado a considerar el RAH como una alternativa seria a la transcripción manual. Argumentan que, dependiendo de la velocidad a la que la persona escriba, el tiempo de aprendizaje y entrenamiento del programa no compensa, a menos que se explore un método de optimización de las herramientas informáticas (Evers, 2011). En este sentido, dos trabajos han retenido nuestra atención.

Tilley investigó el principio de co-construcción de la transcripción a través de la experiencia de personas contratadas en universidades canadienses para transcribir grabaciones académicas (2003). Reportó que uno de los transcritores utilizaba un programa de RAH y documentó cómo este se fue habilitando, sobre la marcha, con estrategias para mejorar la precisión. Según Tilley, el transcriptor reproducía, sin saberlo, el efecto fantasma (*ghosting*) que habían acuñado Frogg y Wightman (2000).

Por su parte, Johnson dictó una entrevista mientras la iba escuchando, luego de haber entrenado el programa *MacSpeech Dictate* (2011). El dictado de la entrevista (17:38 minutos) le tomó treinta minutos cincuenta y seis segundos, ya que había disminuido la velocidad para lograr repetir todo el contenido de la grabación. La transcripción obtenida tenía 96,4% de precisión y necesitó veintinueve minutos para corregirla. Después llevó a cabo la transcripción de la misma entrevista de forma manual, es decir, escuchando y tecleando al ordenador. Le tomó treinta y nueve minutos y siete segundos, y obtuvo 98% de precisión. Necesitó menos de doce minutos para corregirla. Concluyó que los programas de dictado no son más rápidos ni más precisos que el método manual. La validez de la prueba puede cuestionarse (entre otras cosas porque cuando tecleó la entrevista ya la había oído al menos tres veces), pero también apoya nuestro postulado: no habrá optimización del RAH si no se emplean las estrategias adecuadas durante la repetición.

Lo mismo ocurre con otros experimentos que se han llevado a cabo para medir el potencial de los programas de dictado utilizados de forma automática o potenciados con distintas técnicas de repetición a la hora de transcribir grabaciones *off-line*. Comparan transcripciones obtenidas, entre otras, de forma manual (mecnografía y/o estenotipia) con transcripciones obtenidas con la técnica de *leasen and repeat* (Johnson, 2011; Matheson, 2007; Park y Zeanah, 2005) o con reablado *on-line* (Matamala et al., 2017; D'Arcangelo y Cellini, 2013). Ninguna de las técnicas de dictado empleadas ofrece una neta ventaja con respecto a los demás métodos de transcripción.

Teniendo en cuenta que la transcripción de entrevistas para la investigación, basada en la literalidad (Davidson 2009), debe cumplir con unos requisitos, reflejados en un protocolo, dependiendo del uso que se le vaya a dar, hemos optado por aplicar nuestras propias estrategias. Fuera de los trabajos que estamos realizando, no tenemos constancia de ningún otro que mida estrategias específicas de reablado para optimizar la transcripción *off-line* del habla espontánea con fines lingüísticos.

4. METODOLOGÍA

Los datos que analizo provienen del primer taller de reablado *off-line* para transcribir corpus orales de la lengua hablada organizado en la Universidad de Montreal. Se trata de una actividad vinculada al proyecto El español en Montreal y el COLEM (CERAS-2014-15-159D) que, por estar dirigida a estudiantes de segundo y tercer ciclo de la Sección de estudios hispánicos, nos garantizaba el nivel de lengua necesario a la hora de transcribir y revisar entrevistas con las variaciones diatópicas expresadas por los hispanohablantes de la Región metropolitana de Montreal.

Los diez participantes (siete hombres y tres mujeres) estaban familiarizados con los corpus orales de la lengua, pero ninguno había reablado antes. En este trabajo, se emplean los datos de los seis que completaron todas las actividades.

Se organizaron dos jornadas completas. De la primera se han extraído los datos sobre la mecanografía. Se ha medido tanto la cantidad como la calidad de la transcripción obtenida por cada participante durante treinta minutos: primero, al revisar una entrevista reablada por un profesional; segundo, al mecanografiar la continuación de la entrevista. Somos conscientes de la ventaja que puede suponer, en el momento de mecanografiar, el haber tenido ya contacto con parte de la entrevista reablada, pero en nuestra elección primó mantener el nivel de lengua. Se utilizó una entrevista del COLEM realizada en 2018 con un informante procedente de Moca, República Dominicana de treinta y siete años y treinta y tres en Montreal (E32) (60:54 minutos). Al inicio del taller, se dedicó un tiempo para que los participantes se familiarizaran con las pautas de transcripción del COLEM y con los atajos de teclado del reproductor multimedia. Seguidamente se les suministró un texto (E32b) correspondiente a la primera parte de la entrevista E32 (29:50 minutos) reablada por un profesional. Disponían de treinta minutos para revisarlo respetando el protocolo del COLEM. También se midió la velocidad a la que los participantes son capaces de teclear (10fastfingers.com).

Con objeto de obtener datos precisos sobre el reablado, en el segundo día, proporcionamos la grabación de una entrevista distinta del COLEM a cada participante. Debían reablarla y editar la transcripción resultante, luego de una iniciación a las principales estrategias del reablado *off-line*. Aparecen aquí los datos de los dos primeros participantes que completaron la tarea. L1 reabló parte de una entrevista realizada en 2019 con una informante de Santiago de Chile de cuarenta y siete años, y nueve en Montreal (E5) (30:27 minutos). L2 reabló la entrevista completa realizada en 2019 con un informante de La Estrella, Chiriquí, Panamá, de cincuenta y ocho años, y treinta en Montreal (E27) (72:32 minutos).

Para determinar si las variantes del español y la experiencia del reablador pueden repercutir en los resultados, hemos cotejado los datos obtenidos con L1 y L2 con dos entrevistas reabladas por un profesional (L0). La entrevista E4 se realizó en 2019 con un informante procedente de Santiago de Chile de cincuenta años y once en Montreal (106:41 minutos); E26 se hizo en 2019 con un informante de la Ciudad de Panamá de cincuenta y un años, y treinta en Montreal (69:20 minutos).

En el marco de mi investigación, las entrevistas del COLEM se han transcrito con distintos programas de dictado, y se han comparado las transcripciones automáticas con las reabladas. Para el presente artículo, las transcripciones se han realizado con el programa *Dragon Naturally Speaking 13*, por ser representativo de la media de los programas comparados. Presento los resultados obtenidos de la transcripción automática de las entrevistas para cotejarlos con los otros dos métodos.

Todas las transcripciones han sido tratadas en el Centre de Recherche Informatique de Montréal con la herramienta *Align-text* de Kaldi (kaldi-asr.org). Se ha medido el índice de precisión, expresado en porcentaje, según la fórmula del National Institute of Standards and Technology que alinea cada texto con su correspondiente versión final (T0). El porcentaje de precisión se ha calculado dividiendo la suma de borrados (B) (cuando falta algo que figura en el T0), sustituciones (S) (por ejemplo, donde en lugar de adonde) e inserciones (I) (cuando se añade algo que no figura en el T0), por el número de unidades de referencia (N). A su vez, (N) corresponde a la suma de (B + S + C), siendo (C) las unidades correctas.

La versión final (T0) hace referencia a una transcripción corregida según el protocolo del COLEM; esto es, el T0 corresponde a un texto que, sea cual sea el método empleado para obtenerlo, antes de ser sometido a una lectura final por el director del proyecto, debe ser comparado con su audio por un revisor. En este sentido, la revisión de transcripciones de entrevistas puede compararse a la revisión en traducción, ya que ambas requieren revisión y relectura. La revisión comparada del texto meta con el de origen (en nuestro caso, la grabación de la entrevista) es necesaria, entre otras cosas, para comprobar que no ha habido omisiones durante la transferencia (AENOR, 2006).

Por otra parte, no existe un conjunto de reglas y criterios universales para establecer la calidad de una transcripción, dependerá del propósito al que vaya destinada y siempre será una herramienta interpretativa (Lapadat, 2000). Para esta investigación, el índice de precisión no solo tiene en cuenta caracteres o palabras. Le hemos otorgado el mismo valor a palabras, signos de puntuación y otras convecciones del COLEM que forman parte del protocolo: omisión de todos los nombres propios representados con [NP]; identificación del informante con [I:] y del entrevistador con [E:]; inclusión de rasgos suprasegmentales: entonación, ritmo, tono, pausas; elementos paralingüísticos como aplauso, beso y demás recursos cinéticos audibles; reproducción exacta de las alteraciones morfosintácticas. Se trata de una forma de cuantificar el hecho de que sea cual sea la acción a realizar para corregir, siempre habrá que detener el audio y luego editar el texto; en definitiva, siempre supone tiempo.

5. RESULTADOS Y ANÁLISIS

5.1. La mecanografía frente al rehablado

Los datos de la Figura 1 comparan la mecanografía frente al rehablado. Consta de tres bloques. Primero se midió la velocidad a la que cada participante es capaz de teclear durante un minuto, expresada en número de palabras por minuto (ppm), y el porcentaje de precisión. Después, a partir de la entrevista E32b, medimos los minutos y el número de palabras mecanografiados (T5) durante treinta minutos, así como el número de unidades contabilizadas y la precisión. En el tercer bloque figuran los minutos, las palabras, el número de unidades contabilizadas y la precisión de la transcripción T5, que revisó cada participante, durante treinta minutos; corresponde a la entrevista E32a rehablada por un profesional. En la última línea se han calculado los promedios correspondientes a un minuto.

Participante	L1	L2	L3	L4	L5	L6	Media 1 minuto
Prueba de mecanografía: ppm	79	64	43	71	71	52	63,33
Prueba de mecanografía: precisión (%)	99,50	92,80	99,09	97,51	95,16	89,90	95,66
T5-E32b: minutos de grabación	02:07	02:02	01:38	02:30	02:32	04:09	02:29
T5-E32b: palabras	342	362	250	445	272	608	12,66
T5-E32b: unidades	557	508	393	635	392	1003	19,38
T5-E32b: precisión (%)	58,28	76,06	66,49	63,80	68,38	60,16	65,53
T6-E32a: minutos de grabación	13:46	06:23	07:47	04:52	23:35	15:46	12:01
T6-E32a: palabras	1809	1187	1387	980	4355	2643	68,67
T6-E32a: unidades	2735	1756	2059	1434	6394	3939	101,76
T6-E32a: precisión (%)	78,54	80,15	78,19	73,40	81,73	79,61	78,60

Figura 1. Prueba de mecanografía / transcripción mecanografiada (T5-E32b) / transcripción rehablada (T6-E32a)

Los participantes teclean una media de 63,33 ppm, cercana a la de un mecanógrafo experimentado (Moro Vallina, 2010). Sin embargo, al mecanografiar la entrevista E32, obtienen una media de 12,66 ppm. Esta diferencia puede explicarse, en primer lugar, porque en las pruebas de mecanografía como 10fastfingers.com, las palabras se suceden separadas por espacios, sin ningún otro signo de puntuación ni mayúsculas; además, a la hora de mecanografiar la entrevista hay que considerar el tiempo empleado en manipular el audio y el tempo del habla. Con la entrevista rehablada los participantes no necesitan teclear cada palabra; el número de palabras revisadas asciende a una media de 68,67 ppm.

Mecanografiando la entrevista E32b durante treinta minutos, los participantes solo avanzan una media de 2:29 minutos; mientras que, en el mismo tiempo, con la entrevista E32a rehablada —que además contiene ya una revisión—, la media de transcripción obtenida asciende a 12:01 minutos. Esto es, aun añadiendo el tiempo de reablado, que equivale a 1X1 (un minuto de grabación se reabla en un minuto), esta forma de transcribir sigue siendo más rápida. Podríamos deducir que para obtener la transcripción completa de la entrevista E32 (60:54 minutos) mecanografiándola se necesitarían unas trece horas, mientras que a partir del reablado, la media sería de menos de tres horas. Aun sumando el tiempo empleado al reablarla, la transcripción T6 tomó menos tiempo que la mecanografiada; pero, además, es más precisa.

Porque si la rapidez a la que obtengamos la transcripción de la entrevista cuenta, la calidad de su contenido es primordial. En el ámbito del COLEM, para que una transcripción sea de calidad, además de ser fiel al contenido, tendrá que responder rigurosamente al estándar de su protocolo de transcripción. Por eso, aunque durante la prueba de mecanografía se midió la precisión, la media, que solo cuenta palabras, no nos sirve. Nos fijaremos en los porcentajes de precisión de las columnas 7 y 11, resultantes de la comparación de las transcripciones mecanografiadas (T5) y reabladas (T6) con sus correspondientes T0. Estos porcentajes sí representan un baremo de calidad porque al

comparar unidades en lugar palabras tendrán en cuenta todos los elementos que debe reflejar la transcripción del COLEM.

La precisión media al mecanografiar fue de 65,53%, inferior a la de la transcripción rehablada: 78,60%. He aquí algunos ejemplos de los errores y cómo se han contabilizado: (I) por inserción, (S) por sustitución.

T0	T5 / T6
Para identificar entrevistador se emplea la etiqueta [E:] seguida de tabulador (sin espacio)	<ul style="list-style-type: none"> • [E] (S) • seguida de espacio (I), • seguido de dos puntos (I) • y uno o varios espacios (I)
Para identificar al informante se emplea la etiqueta [I:] seguida de tabulador (sin espacio)	<ul style="list-style-type: none"> • [I] (S) • seguida de espacio (I) • seguido de dos puntos (I) • y uno o varios espacios (I)
Todos los nombres propios se sustituyen por [NP] [RISAS]	<ul style="list-style-type: none"> • Aparece el nombre (S) • (RISAS) (S) • [RISA] (S)
[SOLAPAMIENTO]	<ul style="list-style-type: none"> • [Soplamiento] (S) • (SOLOPAMIENTO) (S)
[EXPIRACIÓN]	<ul style="list-style-type: none"> • [RESOPLIDO] (S) • [RESOPLIDO] (S)
[RUIDO]	<ul style="list-style-type: none"> • [toca] (S)
La aspiración de vocales y consonantes no se representa: dieciséis trabajadores	<ul style="list-style-type: none"> • dieciséi (S) • trabajadore (S)
Representación completa de palabras cortadas: conserv- llega-	<ul style="list-style-type: none"> • conservaron (S) • llegaste (S)

Figura 2. Errores de formato en las transcripciones

Los errores se han detectado en ambas versiones, pero en mayor proporción en la mecanografiada; por tanto, la automatización del proceso reduce la parcialidad y la fuerte carga interpretativa inherentes a la transcripción.

Por último, la observación combinada de tiempo y precisión no permite pensar que exista correlación entre la velocidad y la calidad. L5, fue quien más avanzó (23:35 minutos); pero también fue quien obtuvo la mayor precisión (81,73%). Con el siguiente mejor porcentaje (80,15%), L2 solo avanzó 6:23 minutos.

5.2. El RAH frente al rehablado (experto / principiante)

En la figura 3 se comparan transcripciones obtenidas con el RAH (T1) con transcripciones rehabladas (T2). Se han medido los índices de precisión de cuatro entrevistas realizadas a dos chilenos (E4 y E5) y dos panameños (E26 y E27). El rehablado de E4 (Chile) y E26 (Panamá) los hizo un experto (L0). L1 rehabló la entrevista E5 (Chile) y L2 rehabló la E27 (Panamá); para ambos fue su primer rehablado. Las ocho transcripciones se han obtenido con el programa de dictado *Dragon NaturallySpeaking 13*.

Identificación	Transcripción	Precisión	Duración
T1-E4	RAH	12,24%	106:41:00
T2-E4	L0 (experto)	87,75%	106:41:00
T1-E5	RAH	36,93%	90:59:00
T2-E5	L1	62,10%	30:27:00
T1-E26	RAH	25,54%	69:20:00
T2-E26	L0 (experto)	74,46%	69:20:00
T1-E27	RAH	5,53%	72:32:00
T2-E27	L2	67,77%	72:32:00

Figura 3. Transcripción automática / Transcripción rehablada

Ninguna de las transcripciones automáticas (T1) consigue la paridad con el humano (Xiong et al. 2016). La precisión de todas las transcripciones automáticas está por debajo de las rehabladas (T2), confirmándose dos premisas iniciales: por una parte, para transcribir la lengua hablada, el RAH aún no permite la completa automatización del proceso con un programa de dictado automático (Ravanelli et al. 2018); por otra, el rehablado mejora los resultados del RAH de los programas de los programas de dictado (Utray Delgado et al. 2015; Romero-Fresco 2011; Lambourne 2007).

Los errores léxicos reflejan cómo afecta la variación del español en los resultados: inserciones (I), borrados (B), sustituciones (S) y equivalencias correctas (C). En los siguientes ejemplos aparece la palabra tal y como se pronuncia en el texto de origen que es la grabación (T0) con su correspondiente representación en la transcripción T1 (RAH) y T2 (rehablado).

T0	T1	T2
amerindiez	abrir indios (S) + (I)	(B)
citación	situación (S)	citación (C)
outsourcing	(B)	outsourcing (C)
intalando	instalando (S)	instalando (S)
candenciosa	(B)	ambiciosa (S)
quartier	catión (S)	cartilla (S)
marché	marche (S)	marche (S)
incibían	vivía (S)	inscribían (S)
úneco	único (S)	único (S)
tranporte	transporte (S)	transporte (S)
afordable	(B)	razonable (S)
contrucción	confusión (S)	construcción (S)
indostanes	en dos canes (I) + (S) + (I)	(B)
bombardios	un (S)	bombardeos (S)
distes	tristes (S)	diste (S)
cargastes	(B)	cargas tests (S) + (I)
meritaba	meditaba (S)	meditaba (S)
noventicinco	(B)	95 (S)
quebecuá	quebecuá (C)	quebecuá (C)
folcloro	(B)	folclor (S)

Figura 4. Errores léxicos en las transcripciones

Tanto el reabrador como el programa de dictado se enfrentan a la variación a nivel léxico y fonológico. Para que el programa de dictado pueda representar correctamente, en cualquiera de las transcripciones, una palabra pronunciada en la grabación original ya sea en español, incluyendo sus variantes, como en otro idioma (también las originadas por contacto), dicha palabra debe figurar en el léxico del programa. Si la variante *quebecuá* fue reconocida correctamente tanto en la transcripción automática (T1) como en la reabrada (T2) fue porque yo la había añadido al léxico del programa de dictado *Dragon*. Sin embargo, las voces francesas *quartier* y *marchés*, que no formaban parte del léxico de *Dragon*, no podían aparecer en ninguna de las transcripciones; en su lugar, el programa de dictado les atribuyó equivalencias probables, originando errores (S). Por su parte, el reabrador tiende a no repetir una palabra que no forme parte de su propio vocabulario; además, si es experto, tampoco repetirá ninguna que crea que no figura en el diccionario del programa, como veremos más adelante.

La experiencia del reabrador ha permitido marcar el acento prosódico para desambiguar algunos monosílabos (dé, más, tú, él, mí, té, sí, sé), pronombres interrogativos y exclamativos (qué, quién, quiénes, cuál, cuáles, cuánto, cuánta, cuántos, cuántas, cómo), pero, sobre todo, se ha logrado un mejor reconocimiento ante solapamientos, asentimientos y vacilaciones, muy característicos de la conversación. Por ejemplo, en la entrevista E4 el entrevistador asiente doscientas veinticuatro veces. Veamos, en la figura 5, un fragmento de la transcripción de la entrevista E24 y el rendimiento de *Dragon* con la estrategia aplicada por el reabrador cuando el entrevistador asiente. Las líneas en blanco no aparecen en T2, las hemos añadido para ayudar a la lectura.

TO	T2-L0
mejoramiento de lo que es la calidad de vida.	mejoramiento de lo que es la calidad de vida.
[E:] [ASERT] ¡Hum!	[E:]
[I:] Porque en... en... en Panamá a pesar de que la moneda, ah...	[I:] Porque en Panamá a pesar de que la moneda
[E:] [ASERT] ¡Hum!	
[I:] ...local es el balboa...	[E:]
[E:] [ASERT] ¡Hum!	[I:] Local es el balboa,
[I:] ...a raíz del... del canal...	[E:]
[E:] [ASERT] ¡Hum!	[I:] A raíz del, del canal
[I:] ...el balboa, que solamente se... se da en moneda...	[E:]
[E:] [ASERT] ¡Hum!	[I:] El balboa que solamente se moneda
[I:] ...no se... no existe en papel, tiene el valor... el valor equivalente monetario al dólar americano.	[E:]
	[I:] No existe en papel, tiene el valor el valor equivalente monetario al dólar americano.

Figura 5. Fragmento 1 (E4): estrategias de reabrado

Cada vez que el entrevistador asiente, el reabrador ha optado por utilizar un atajo verbal que le sirve para que el programa de dictado identifique al entrevistador seguido de un atajo verbal que le sirve para que el programa de dictado identifique al informante. El texto resultante (T2) aparece con la etiqueta [E:] seguida de tabulador, nueva línea, [I:], tabulador. Antes de revisar la transcripción (T2), una búsqueda rápida le permite añadir [ASERT] ¡Hum! en un par de clics. A modo de comparación, este operador de asentimiento aparece ciento dieciocho veces en la entrevista E5, ciento setenta y ocho en la E26 y doscientas cincuenta y cinco en la E27.

En la figura 6 puede verse un ejemplo de las vacilaciones y solapamientos que caracterizan la comunicación oral. De nuevo, las líneas en blanco no aparecen en T2, las hemos añadido para ayudar a la lectura.

T0	T2-L0
[I:] Eh... es... es que uno de sus... eh... bisabuelos...	[I:] Es que uno de sus... bisabuelos
[E:] [ASERT] Ajá.	
[I:] ...había sido un gran... eh gram-... eh... gramático...	[I:] había sido un gran gran dramático.
[E:] [ASERT] Ajá.	
[I:] ...después literato. Eh... la familia tenía como un lado de musical y otro lado por... eh... por la lengua.	[I:] Después literato. La familia tenía acumulado de musical y otro lado por por la lengua.
[E:] [ASERT] ¡Hum!	
[I:] ¿Ya?	[E:]
[E:] [ASERT] ¡Hum!	[I:] Ya
[I:] Y... y fue muy interesante, eh... porque... eh, eh... en tu país ocurre una cosa que en mi país no pasa y es que eh... eh... ustedes son más abiertos que nosotros.	[E:]
[E:] Sí.	[I:] Y y fue muy interesante porque una cosa que mi país no pasa es que ustedes son más abiertos que nosotros.
[I:] Ya. O sea, en Chile...	[E:]
[E:] [SOLAPAMIENTO] No he estado en Chile, pero digamos que sí... sí... sí es... [ASERT] mhm...	[I:] Ya
[I:] [CHASQUIDO] Y cuando tú llegas a un sector social determinado es casi impenetrable.	[E:] Y agregamos que sí sí
[E:] [ASERT] ¡Hum!	[I:] [CHASQUIDO] y cuando tú llegas a un sector social determinado es casi impenetrable.
	[E:]

Figura 6. Fragmento 2 (E4): estrategias de rehablado

Ante solapamientos y vacilaciones, el rehablador L0 solo repite lo que está seguro de que le permitirá obtener una suerte de documento esqueleto que podrá editar posteriormente rellenando huecos, en lugar de tener que borrar y reescribir.

6. CONCLUSIONES Y PERSPECTIVAS FUTURAS

A la hora de crear un corpus, la transcripción de grabaciones orales puede convertirse en uno de los aspectos más largos del proceso. Para saber qué método al alcance de todos ofrece más ventajas, he comparado la mecanografía, el reconocimiento automático del habla y el rehablado *off-line*. Las estrategias de rehablado *off-line* me han permitido el uso de un programa automático de dictado en su estado actual como herramienta para potenciar la transcripción de las entrevistas del COLEM en menos tiempo y con menos errores.

Con las transcripciones rehabladas hemos reducido la parcialidad y la carga interpretativa. No obstante, sea cual sea el método empleado, el trabajo de transcripción, intenso y difícil, origina muchísima variación (Roulston et al. 2003: 657). Por una parte, el traslado intersemiótico que supone la transcripción de la lengua hablada, al ceñirla dentro de las convenciones escritas, nos obliga a seleccionar qué datos somos capaces de trasladar. Por otra, por muy detallada y completa que sea una transcripción, esta siempre habrá

pasado por el filtro de la interpretación del transcriptor (Kvale 1996) a la vez que, de manera paradójica, el resultado presenta varios grados de heterogeneidad.

La cuestión primordial es definir lo que se considera una transcripción correcta; en su ausencia, los datos que obtendremos serán muy dispares. Esto explica que, por un lado, hayamos encontrado investigadores que afirman que quien teclee rápido puede transcribir una entrevista de una hora entre tres y seis horas (Punch y Oancea 2014; Walford 2001); mientras que, por el otro, hay investigadores que mantienen que la tarea de mecanografía podría estar más cerca de sesenta horas por cada hora registrada de datos, dependiendo del tipo de transcripción que se realice (Evers 2011). El nivel de transcripción, esto es, la cantidad de contenido transcrito es la que está al origen de tales diferencias. El protocolo de transcripción contrarresta la ineludible parcialidad del proceso, minimiza la variabilidad de los resultados, delimita qué debe reflejar cada transcripción y cómo habrá de hacerlo, a la vez que sistematiza la carga interpretativa de signos semióticos propios de una conversación (Cook 1995).

Hay que mencionar la dificultad añadida que supone el hecho de que la representación escrita del discurso oral no es biunívoca. Para ilustrar la complejidad de la tarea, a modo de ejemplo, retomaremos la representación escrita de las unidades léxicas que sirven para construir la interacción. Estos retrocanalizadores lingüísticos (Bravo Cladera, 2009) o *back channel responses* (Yngve, 1970), ratifican la receptividad y atención continuada del interlocutor. Tottie (2014) los llama *vocalizations* y los define como señales o muestras de respuesta que pueden tener varios significados, como asegurarle a nuestro interlocutor que lo estamos siguiendo sin que haya necesidad de retomar el turno de palabra; también pueden tener varias funciones como expresar vacilación, duda, reparo, asombro, asentimiento, negación, etc. En el discurso se caracterizan por tener muchas variantes fonológicas y suelen acompañarse de respuestas no verbales (mover la cabeza de delante a atrás o de un lado a otro, encogerse de hombros), las cuales, en la transcripción, se pierden. Tampoco todos tienen una ortografía normalizada, además, están sometidos al fenómeno de la variación ya que en varias lenguas poseen similitudes sonoras con sus propias variantes entonativas y de contenido semántico. Estos retrocanalizadores lingüísticos aparecen con tal frecuencia en las entrevistas que han ocasionado mucha inversión de tiempo por parte de transcriptores y revisores. En los corpus, la variación en su representación "dificulta una consulta comparable y, en consecuencia, la descripción de sus usos" (Solís García y León Gómez, 2018: 332). Para nuestra investigación se ha propuesto una etiqueta única acompañada de tres codificaciones: la interjección *hum*, que es la representación más generalizada (Bravo Cladera, 2009), *mhm.*, cuando el alargamiento es más consonántico y *ehm* cuando el contenido fónico es más vocálico. La etiqueta debe facilitar la automatización del lenguaje y el rehablado, así como la transcripción y la revisión, dejando el corpus abierto a futuros estudios de otra índole: lingüísticos, de análisis del discurso, de pragmática.

A modo de reflexión final, señalamos la apremiante necesidad de encontrar un sistema que integre todas las herramientas para la elaboración de materiales de estudio de corpus: transcripción, revisión, etiquetado y análisis de datos. Entendemos que la clave está en una colaboración eficaz humano-computadora. Centrándonos en nuestro objeto de estudio, la transcripción, los creadores de interfaces contemplan la corrección de errores del RAH desde la perspectiva de la transcripción asistida por ordenador (Revuelta-Martínez et al., 2012). Esto es, permitir la intervención del humano en el momento en que el programa está produciendo el reconocimiento automático para corregir, ya sea con teclado, ratón, voz, gestos o atajos, en tiempo real, a imagen de los modelos predictivos usados en traducción

asistida por ordenador (de Souza et al., 2015). Dentro de este enfoque, el reablado *off-line* debería poder usarse para editar los segmentos que contienen más errores sobre la marcha (Sperber et al., 2017; Sperber et al., 2016). En definitiva, gracias a la combinación de herramientas que hasta ahora se han empleado de forma aislada se lograrán mejores resultados en menor tiempo y la metodología de transcripción dará un gran paso adelante.

AGRADECIMIENTOS

Nuestro agradecimiento a los participantes en el taller de reablado *off-line*, al equipo del CRIM, especialmente a Simon Desrochers, por la colaboración con nuestro proyecto para el análisis de datos y a Enrique Pato por la lectura del manuscrito.

REFERENCIAS BIBLIOGRÁFICAS

- AA.VV. (2008). Diccionario de términos clave de ELE. Diccionario de términos clave de ELE. <https://cvc.cervantes.es>. [Consultado el 29 de junio de 2020].
- ACR/CAB (2012). *Normes universelles du sous-titrage codé à l'intention des télédiffuseurs canadiens de langue française*. Document révisé en fonction de CRTC 2011-741 et CRTC 2011-741-1. <http://www.cab-acr.ca>. [Consultado el 25 de junio de 2020].
- Adolphs, S., Knight, D. (2010). Building a spoken corpus. What are the basics? In: A. O'keeffe, M. McCarthy (eds.), *The Routledge Handbook of Corpus Linguistics*. London: Routledge, 38-52.
- AENOR (2006). *Norma Española UNE-EN 15038. Servicios de traducción. Requisitos para la prestación del servicio*. Madrid: AENOR.
- Ainsworth, W. A. (1988). *Speech recognition by machine*. London, UK: Institution of Electrical Engineers.
- Bernabé Caro, R., Eugeni, C., Masia, V., Oncins, E. (2019). Bridging the gap between training and profession in real-time intralingual subtitling. *Media for All 8*, Stockholms universitet, Stockholm, 17-19 de junio de 2019.
- Blackley, S. V., Huynh, J., Wang, L., Korach, Z., Zhou, L. (2019). Speech recognition for clinical documentation from 1990 to 2018: a systematic review. *Journal of the American Medical Informatics Association*, 26(4). 324-338. doi: 10.1093/jamia/ocy179.
- Bravo Cladera, N. (2009). La retrocanalización como una realización de la interacción: algunos usos de mm y mhm en español. In: *Actas del II Congreso de Hispanistas y Lusitanistas Nórdicos: Actas do II Congresso de Hispanistas e Lusitanistas Nórdicos*, Estocolmo, 25-27 de octubre de 2007: Acta Universitatis Stockholmiensis, Romanica Stockholmiensia 26, 25-42.
- Brinton, B., Fujiki, M., Winkler, E., Loeb, D. F. (1986). Responses to requests for clarification in linguistically normal and language-impaired children. *Journal of Speech & Hearing Disorders*, 51(4). 370-378. doi: 10.1044/jshd.5104.370.
- Brooks, C. (2010). Embodied Transcription: A Creative Method for Using Voice-Recognition Software. *Qualitative Report*, 15(5), pp. 1227-1241.
- Brousseau, J., Beaumont, J.-F., Boulianne, G., Cardinal, P., Chapdelaine, C., Comeau, M., Osterrath, F., Ouellet, P. (2003). Automated closed-captioning of live TV broadcast news in French. In: *Eurospeech 2003 Proceedings: 8th European Conference on Speech Communication and Technology*, INTERSPEECH 2003, Geneva, CICG, 1-4 de septiembre de 2003. Bonn: ISCA, 1245-1248.
- Chiu, C.-C., Tripathi, A., Chou, K., Co, C., Jaitly, N., Jaunzeikare, D., Kannan, A., Nguyen, P., Sak, H., Sankar, A., Tansuwan, J., Wan, N., Wu, Y., Xuedong, Z. (2018). Speech recognition for medical conversations. In: *19th International Conference on Acoustics, Speech and Signal*

- Processing (ICASSP)*, INTERSPEECH 2018, Hyderabad, International Convention Centre, 2-6 de septiembre de 2018: ISCA, 2972-2976.
- Cook, G. (1995). Theoretical issues: transcribing the untranscribable. In: G. Leech, G. Myers, J. Thomas (eds.), *Spoken English on computer: Transcription, mark-up and application*. Harlow, UK: Longman, 35-53.
- CRTC, C. d. l. r. e. d. t. c. (2007). *Nouvelle politique de sous-titrage codé pour malentendants*. 2007-54. Ottawa: Gouvernement du Canada. <https://crtc.gc.ca>. [Consultado el 20 de junio de 2020].
- D'Arcangelo, R., Cellini, F. (2013). Metodi e tempi di una verbalizzazione-prove tecniche. In: C. Eugeni, L. Zambelli (eds.), *Specializzazione on-line. Numero monografico sul Respeaking* (Maggio 2013), pp. 81-95. <http://www.accademia-aliprandi.it>.
- Damper, R. I., Lambourne, A. D., Guy, D. P. (1985). Speech input as an adjunct to keyboard entry in television subtitling. In: B. Shackel (ed.), *Human-Computer Interaction-INTERACT'84*. Amsterdam: Elsevier, 203-208.
- Davidson, C. (2009). Transcription: Imperatives for Qualitative Research. *International Journal of Qualitative Methods*, 8(2). 35-52. doi: 10.1177/160940690900800206.
- de Souza, J. G., Negri, M., Ricci, E., Turchi, M. (2015). Online multitask learning for machine translation quality estimation. In: C. Zong, M. Strube (eds.), *Proceedings of the 53rd Annual Meeting of the Association for Computational Linguistics and the 7th International Joint Conference on Natural Language Processing* (Vol. 1: Long papers), IJCNLP 2015, Beijing, China National Convention Center, 26-31 de julio de 2015. Red Hook, NY: ACL, 219-228.
- Dudley, H. (1958). Phonetic Pattern Recognition Vocoder for Narrow-Band Speech Transmission. *The Journal of the Acoustical Society of America*, 30(8), pp. 733-739.
- Durand, J. (2017). Corpus Phonology. *Oxford Research Encyclopedia of Linguistics*.
- Edwards, E., Salloum, W., Finley, G. P., Fone, J., Cardiff, G., Miller, M., Suendermann-Oeft, D. (2017). Medical Speech Recognition: Reaching Parity with Humans. In: A. Karpov, R. Potapova, I. Mporas (eds.), *Speech and Computer. 19th International Conference, SPECOM 2017 Hatfield, UK, September 12-16, 2017 Proceedings* (Vol. 10458). Cham, CH: Springer International Publishing, 512-524. doi: 10.1007/978-3-319-66429-3_51.
- El español en Montreal: COLEM (2006). Corpus: Pato, Enrique (dir.). <https://esp-montreal.jimdo.com>. [Consultado el 17 de abril de 2020].
- Elvira-García, W., Roseano, P., Fernández Planas, A., Martínez Celdrán, E. (2015). Una herramienta para la transcripción prosódica automática con etiquetas Sp_ToBI en Praat. In: A. Cabedo Nebot (ed.), *Perspectivas actuales en el análisis fónico del habla: tradición y avances en la fonética experimental*. València: Universitat de València, 455-464.
- Evers, J. C. (2011). From the past into the future. How technological developments change our ways of data collection, transcription and analysis. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, 12(1). Art. 38.
- FCC (2014). *Report and Order, Declaratory Ruling, and Further notice of proposed rulemaking*. FCC-14 12A1. Federal Communications Commission. <https://www.fcc.gov/>. [Consultado el 25 de junio de 2020].
- Ferber, R. (1991). Slip of the tongue or slip of the ear? On the perception and transcription of naturalistic slips of the tongue. *Journal of Psycholinguistic Research*, 20(2), pp. 105-122.
- Fogg, T., Wightman, C. W. (2000). Improving Transcription of Qualitative Research Interviews with Speech Recognition Technology. *Creating Knowledge in the 21st Century: Insights From Multiple Perspectives* (American Educational Research Association Annual Meeting), New Orleans, LA, 24-26 de abril de 2000.
- Gadet, F., Ludwig, R., Mondada, L., Pfänder, S., Simon, A. C. (2012). Un grand corpus de français parlé: le CIEL-F. *Revue française de linguistique appliquée*, 17(1). 39-54.
- Hawkins, W. R., Robinson, R. N. (1979). The development and use of an electric keyboard for television subtitling by Palantype. *International Journal of Man-Machine Studies*, 11(6), pp. 701-710.

- Hodgson, T., Coiera, E. (2016). Risks and benefits of speech recognition for clinical documentation: a systematic review. *Journal of the American Medical Informatics Association*, 23(e1), pp. 169-179. doi: 10.1093/jamia/ocv152.
- Ivarsson, J. (1992). *Subtitling for the media: A handbook of an art*. Stockholm: Transedit.
- Johnson, B. E. (2011). The speed and accuracy of voice recognition software-assisted transcription versus the listen-and-type method: a research note. *Qualitative Research*, 11(1), pp. 91-97. doi: 10.1177/1468794110385966.
- Johnson, M., Lapkin, S., Long, V., Sanchez, P., Suominen, H., Basilakis, J., Dawson, L. (2014). A systematic review of speech recognition technology in health care. *BMC Medical Informatics and Decision Making*, 14(1). doi: 10.1186/1472-6947-14-94.
- Kreuz, R. J., Riordan, M. A. (2018). The art of transcription: Systems and methodological issues. In: A. H. Jucker, K. P. Schneider, W. Bublitz (eds.), *Methods in Pragmatics* (Vol. 10). Berlin-Boston: The Gruiter Mouton, 1634.
- Kvale, S. (1996). *InterViews: An introduction to qualitative research interviewing*. Thousand Oaks, CA: Sage.
- Lambourne, A. D. (2007). Re-speaking the truth. *International Broadcast Engineer*, IBE JULY/AUGUST 2007: 44-46.
- Lapadat, J. C. (2000). Problematizing transcription: Purpose, paradigm and quality. *International Journal of Social Research Methodology*, 3(3), pp. 203-219. doi: 10.1080/13645570050083698.
- Li Deng, Yang Liu, ed. (2018). *Deep learning in natural language processing*. Singapore: Springer Singapore. doi: 10.1007/978-981-10-5209-5.
- Lindsay, J., O'Connell, D. C. (1995). How do transcribers deal with audio recordings of spoken discourse? *Journal of Psycholinguistic Research*, 24(2), pp. 101-115.
- Lu, X., Li, S., Fujimoto, M. (2020). Automatic speech recognition. In: Y. Kidawara, E. Sumita, H. Kawai (eds.), *Speech-to-Speech Translation*. Singapore: Springer Singapore, 21-38. doi: 10.1007/978-981-15-0595-9_2.
- MacLean, L. M., Meyer, M., Estable, A. (2004). Improving accuracy of transcripts in qualitative research. *Qualitative Health Research*, 14(1), pp. 113-123. doi: 10.1177/1049732303259804.
- Manrique Fuero, F. (2016). Innovación en técnicas de estenotipia para subtitulado en directo. In: *VIII Congreso de Accesibilidad a los Medios Audiovisuales para Personas con Discapacidad*, AMADIS'16, Toledo, Museo del Ejército del Alcázar de Toledo, 27-28 de octubre de 2016. Madrid: Real Patronato sobre Discapacidad, 44-58.
- Mariani, J., ed. (2002). *Analyse, synthèse et codage de la parole. Traitement automatique du langage parlé 1* (vol 1). Paris: Hermès Science Publications.
- Markle, D. T., West, R. E., Rich, P. J. (2011). Beyond transcription: technology, change, and refinement of method. *Forum Qualitative Sozialforschung/Forum: Qualitative Social Research*, 12(3). Art. 21.
- Matamala, A., Romero-Fresco, P., Daniluk, L. (2017). The use of respeaking for the transcription of non-fictional genres: An exploratory study. *inTRAlinea: Online Translation Journal*, 19. <http://www.intralinea.org>.
- Matheson, J. L. (2007). The voice transcription technique: Use of voice recognition software to transcribe digital interview data in qualitative research. *The Qualitative Report*, 12(4), pp. 547-560.
- McCoy, E., Shumway, R. (1979). Real-time captioning--promise for the future. *American Annals of the Deaf*, 124(5), pp. 681-690.
- Moores, Z. (2016). Subtitling live events through respeaking-Increasing accessibility for all. Unlimited! International Symposium on Accessible Live Events, University of Antwerp, Antwerp, 29 de abril de 2016.
- Moro Vallina, M. (2010). *Aplicaciones ofimáticas*. Madrid: Editorial Paraninfo.
- Niemants, N. (2018). Des enregistrements aux corpus : transcription et extraction de données d'interprétation en milieu médical. *Meta*, 63(3), pp. 665-694.

- Núñez Hidalgo, J., Ramos Villajos, E. (2010). Doscientos años de taquigrafía parlamentaria: de las Cortes de Cádiz a nuestros días (1810-2010). *Revista de las Cortes Generales*, (80), pp. 145-179.
- NVRA, N. V. R. A. (2008). The Horace Webb history. <https://nvra.org>. [Consultado el 2 de febrero de 2018].
- Ochs, E. (1979). Transcription as theory. In: Elinor Ochs, Bambi Schieffelin (eds.), *Developmental pragmatics*. New York, NY: Academic Press, 43-72.
- ONU, N. U. (1994). *Normas Uniformes sobre la igualdad de oportunidades para las personas con discapacidad*. Resolución 48/96. <https://www.un.org>. [Consultado el 20 de junio de 2020].
- Park, J., Zeanah, A. E. (2005). An evaluation of voice recognition software for use in interview-based research: A research note. *Qualitative Research*, 5(2), pp. 245-251.
- Pato, E. (2017). La realidad lingüística en Canadá y la situación del español en Montreal. *Oltreoceano*, 13: 27-37. doi: 10.1400/253045.
- Pollard, S. E., Neri, P. M., Wilcox, A. R., Volk, L. A., Williams, D. H., Schiff, G. D., Ramelson, H. Z., Bates, D. W. (2013). How physicians document outpatient visit notes in an electronic health record. *International Journal of Medical Informatics*, 82(1). 39-46.
- Punch, K. F., Oancea, A. (2014). *Introduction to research methods in education* (2nd edition). SAGE Publishing. www.sagepublications.com. [Consultado el 23 de junio de 2020].
- Ravanelli, M., Brakel, P., Omologo, M., Bengio, Y. (2018). Light gated recurrent units for speech recognition. *IEEE Transactions on Emerging Topics in Computational Intelligence*, 2(2), pp. 92-102. doi: 10.1109/TETCI.2017.2762739.
- Revuelta-Martínez, A., Rodríguez, L., García-Varea, I. (2012). A computer assisted speech transcription system. In: F. Segond (ed.), *Proceedings of the Demonstrations at the 13th Conference of the European Chapter of the Association for Computational Linguistics*, EACL 2012, Avignon, 23-27 de abril de 2012. Stroudsburg, PA: Association for Computational Linguistics, 41-45. doi: 10.3115/v1/P15-1.
- Revuelta Domínguez, F. I., Sánchez Gómez, M. C. (2005). El proceso de transcripción en el marco de la metodología de investigación cualitativa actual. *Enseñanza & Teaching*, 23: 367-386.
- Rodero Antón, E. (2016). Influence of speech rate and information density on recognition: The moderate dynamic mechanism. *Media Psychology*, 19(2), pp. 224-242. Consultado el 2016/04/02. doi: 10.1080/15213269.2014.1002942.
- Romero-Fresco, P. (2011). *Subtitling through speech recognition: respeaking* (1a). Manchester, UK; Kinderhook, NY: St. Jerome Publishing.
- Romero-Fresco, P. (2012). Respeaking in translator training curricula, present and future prospects. *The Interpreter and Translator Trainer*, 6(1), pp. 91-112. doi: 10.1080/13556509.2012.10798831.
- Roulston, K., DeMarrais, K., Lewis, J. B. (2003). Learning to interview in the social sciences. *Qualitative inquiry*, 9(4), pp. 643-668.
- Rufino Morales, M. (2020). El rehablado *off-line* para optimizar la transcripción de corpus orales en español. In: S. Martínez Martínez (ed.), *"Nuevas tendencias en Traducción e Interpretación"*. Granada: Comares, 17-27.
- Saon, G., Kurata, G., Sercu, T., Audhkhasi, K., Thomas, S., Dimitriadis, D., Cui, X., Ramabhadran, B., Picheny, M., Lim, L.-L. (2017). English conversational telephone speech recognition by humans and machines. *arXiv preprint arXiv:1703.02136*.
- Solís García, I., León Gómez, M. (2018). Hum, operador de la interacción en español: valor y usos. *Círculo de Lingüística Aplicada a la Comunicación*, 74: 323-352.
- Sperber, M., Neubig, G., Nakamura, S., Waibel, A. (2016). Optimizing Computer-Assisted Transcription Quality with Iterative User Interfaces. In: N. Calzolari, K. Choukri, T. Declerck, S. Goggi, M. Grobelnik, B. Maegaard, J. Mariani, H. Mazo, A. Moreno, J. Odiijk, S. Piperidis (eds.), *Proceedings of the 10th International Conference on Language Resources and Evaluation Conference (LREC 2016)*, Portoroz, SI, Grand Hotel Bernardin Conference Center, 23-28 de mayo de 2016: European Language Resources Association (ELRA), 1986-1992.

- Sperber, M., Neubig, G., Niehues, J., Nakamura, S., Waibel, A. (2017). Transcribing against time. *Speech communication*, 93: 20-30.
- Tanton, N. E. (1979). UK Teletext-Evolution and Potential. *IEEE Transactions on Consumer Electronics*, CE-25(3), pp. 246-250. doi: 10.1109/TCE.1979.273220.
- Tatham, M., Morton, K. (2005). *Developments in speech synthesis*. Chichester, WS: John Wiley & Sons.
- Tilley, S. A. (2003). "Challenging" research practices: Turning a critical lens on the work of transcription. *Qualitative inquiry*, 9(5), pp. 750-773.
- Tottie, G. (2014). What does uh-(h) uh mean: American English vocalizations and the Swedish learner. *Functions of Language*, 21(1), pp. 6-29.
- Utray Delgado, F., García Castillejo, Á., Puente Rodríguez, L., Carrero Lea, J. M., Pajares, J. L., González, Y., Ruiz-Mezcua, B., Sánchez Pena, J. M. (2015). Informe de seguimiento del subtítulo y la audiodescripción en la TDT 2014, *Colección Inclusión y Diversidad*. 15. Madrid: Grupo Editorial Cinca.
- Walford, G. (2001). *Doing qualitative educational research. A personal guide to research process*. London, UK; New York, NY: Bloomsbury Publishing.
- Winata, G. I., Cahyawijaya, S., Liu, Z., Lin, Z., Madotto, A., Xu, P., Fung, P. (2020). Learning fast adaptation on cross-accented speech recognition. *arXiv preprint arXiv:2003.01901*.
- Xiong, W., Droppo, J., Huang, X., Seide, F., Seltzer, M., Stolcke, A., Yu, D., Zweig, G. (2016). Achieving human parity in conversational speech recognition. *arXiv:1610.05256*.
- Yadav, D., Gupta, M., Chetlur, M., Singh, P. (2018). Automatic annotation of voice forum content for rural users and evaluation of relevance. In: *Proceedings of the 1st ACM SIGCAS Conference on Computing and Sustainable Societies*, Menlo Park, CA, 20-22 de junio de 2018. doi: 10.1145/3209811.3209875.
- Yngve, V. (1970). On getting a word in edgewise. In: *Papers from the sixth regional meeting, Chicago Linguistic Society*, Chicago, IL, 16-18 de abril de 1970: Chicago Linguistic Society, 567-577.
- Zhou, L., Blackley, S. V., Kowalski, L., Doan, R., Acker, W. W., Landman, A. B., Konriant, E., Mack, D., Meteer, M., Bates, D. W., Goss, F. R. (2018). Analysis of Errors in Dictated Clinical Documents Assisted by Speech Recognition Software and Professional Transcriptionists. *JAMA Network Open*, 1(3). e180530-e180543. doi: 10.1001/jamanetworkopen.2018.0530.
- Zweigenbaum, P., Gauvain, J.-L., Braffort, A., Filhol, M., Ghannay, S., Grouin, C., Hamon, T., Illouz, G., Lavergne, T., Ligozat, A.-L. (2020). Le bulletin n°107 – janvier 2020 – IA et Technologies du Langage Humain. Grenoble: Association française pour l'Intelligence Artificielle (AFIA). <https://afia.asso.fr>.

Miscelánea



Miscellanea

Using Dialogic Talks in EFL Primary Teacher Education: An Experience

Una Experiencia sobre Tertulias Dialógicas en Magisterio en Educación Primaria con la especialidad de Lengua Extranjera

Raquel Fernández Fernández

Universidad de Alcalá

raquel.fernandez@uah.es

ABSTRACT

The present study is part of a long-term research project based on the use of Dialogic Talks as a collaborative instructional strategy in EFL Teacher Education. The study was conducted with a group (n=20) of EFL Primary Teacher Education undergraduates with the aim of measuring the impact of the use of this strategy in the classroom. Data gathering tools included a questionnaire, students' blogs, observation and the transcripts of sessions conducted in the classroom. Results show evidence of dialogic and transactional learning, while also finding traits of the use of higher-order thinking skills and the development of students' communicative competence in English.

Keywords: Teacher Education, Dialogic Talks, EFL, techniques

RESUMEN

El presente estudio es parte de un proyecto de investigación de larga duración basado en el uso de las tertulias dialógicas como una estrategia de instrucción colaborativa en la formación de profesorado de lengua extranjera (inglés). El estudio se llevó a cabo con un grupo (n=20) de estudiantes del Grado de Magisterio de Educación Primaria (especialización en enseñanza del inglés como lengua extranjera). El principal objetivo era medir el impacto del uso de esta estrategia en el aula. Las herramientas de recogida de datos incluyen un cuestionario, los blogs de los estudiantes, observación y las transcripciones de las sesiones que se realizaron en el aula. Los resultados muestran evidencia del aprendizaje dialógico y transaccional, así como del uso de las destrezas de pensamiento de rango superior y desarrollo de las competencias comunicativas en inglés.

Palabras clave: Formación de profesorado, tertulias dialógicas, enseñanza del inglés como lengua extranjera, técnicas

1. INTRODUCCIÓN

Two of the main gaps Communicative Language Teaching did not fill were, first, to reach a high level of authenticity of purpose when using the language (Coyle, Hood and Marsh, 2010) and, second, not having defined the role of literature in the classroom (Paran, 2000). With the advent of CLIL as a methodological approach, there is an opportunity not only to cover those areas but also to connect them, as the need to promote good quality literacy practices in bilingual classrooms has been outlined by research (see Meyer et al. 2015).

Learning contexts where real communication is achieved are generally based on genuine interaction and, therefore, have 'dialogue' at the core of their practices. Dialogue has a pivotal role not only in understanding but also in generating knowledge. This is the main tenet of Flecha's 'Dialogic Learning' (1997), which bolsters the creation of collaborative spaces based on equalitarian dialogue and respectful interaction. Flecha considers Dialogic Learning central in the educational process, ultimately leading to transforming the community.

Creating the appropriate conditions for this to occur is, however, not an easy task, due to at least two main reasons. First, that many students have experienced a transmission-based, unidirectional and focus-on-form education. They expect teachers to follow this methodology and feel uneasy when they are asked to build knowledge with their classmates or think on their own. Second, that dialogic talks are based on reading texts, commonly literature books. The 2017 Spanish reading barometer indicates that the reading habit drastically decreased in population aged more than 14. In the case of people aged from 25-34, almost 50% claim that they do not read because they do not like it, or they are not interested in it. If the book is written in an additional language, the motivation may probably decrease even more.

As some difficulties using literary texts have been foreseen, there is a need to follow a methodological model which helps ease the way to introduce texts in the classroom while also matching with the dialogic principles stated before. After having revised Nøae, Carter and Long (1991), the Cultural, Language and Personal Growth models are disregarded, as they cover just partially students' needs and interests. Instead, the transactional model put forward by L.M. Rosenblatt (1938) is chosen. The model elevates the creation of meaning in the reading process. The focus is not on the text or the reader, but on the interaction between those, and how this building of knowledge is also a shared experience (Rosenblatt terms it 'public sphere' of reading). Therefore, the reading transaction will also be favoured using dialogic talks in the classroom, with the purpose of offering a positive and rewarding reading of literary works.

The present study is part of a long-term research project on the impact of dialogic talks in teacher education. The experience presented in this contribution was conducted with a group (n=20) of primary teacher undergraduates completing their final year. They were specialising in English as a Foreign Language and were enrolled in the bilingual group (completing more than 50% of the degree ECTS in English and through CLIL). The experience was carried out in the subject 'Exploring Children's Literature in English' and revolved around the first book of the series *Mary Poppins*, written by P.L. Travers in 1934, and consisted of a total of twelve dialogic talks developed in October and November 2015. The sessions were recorded and transcribed, and students completed a final questionnaire and a blog. In 2017, students were contacted to complete a questionnaire about their experience with Dialogic Talks and its impact on their personal and professional lives.

2. THEORETICAL FRAMEWORK AND LITERATURE REVIEW

2.1 In search of a new model for literature in the bilingual classroom

Literature has played different roles in the history of ELT, from being at the core of methodologies such as the Grammar-Translation method to being almost completely forgotten with the Direct or the Audiolingual Method. Its role in the last decades has not yet been clearly defined. In 2000, Professor Amos Paran asserted that “for the majority of EFL teachers, literature is still not considered to be an essential element within the overall experience of language learning” (78). The use of literature may then be well researched on as an opportunity to develop students’ language while also contributing to their educational development. Following the Loop Input theory stated by Woodward (1991), if this is done in Teacher Education contexts, it is expected to start working on strategies and techniques undergraduates can later bring into their own classes.

Even if it seems clear that literature deserves another opportunity, it is also true that the models defined by Carter and Long (1991) are not really covering students’ needs and, what is more important, not making a full use of the text which ultimately leads to a true interaction not only with the text, but also with other readers. Fostering an adequate experience of the literary text will encourage students to keep on reading and sharing, thinking critically and finding ways to understand and express their ideas in the additional language.

In this context, the Transactional Theory of reading, put forward by Rosenblatt in her work *Literature as an exploration* (1938), and her subsequent publications (see Rosenblatt, 1978 or 2005) is presented as a possible alternative. Her proposal was based on the idea that the text and the reader are modifying each other in a process she labelled as ‘the transaction’. In other words, when we read a text for the first time and when we do it for the second time, neither the text nor we are the same. Using dialogic terminology, the text is in a constant dialogue with the reader. In this sense, Rosenblatt highlighted the importance of the community of readers and established text discussion and debate as the main methodological tool to reach an appropriate transaction with the text:

An atmosphere of informal, friendly exchange should be created. The student should feel free to reveal emotions and to make judgements. [...] Teachers and pupils should be relaxed enough to face what indeed happened as they interpreted in the printed page. (Rosenblatt, 1995: 67)

Another critical element in Rosenblatt’s theory is her distinction between efferent and aesthetic stances. In her view, most reading processes encouraged in the classroom are inviting students to experience texts from an efferent point of view, that is, searching for and storing information. However, native speakers rarely experience literary texts this way, as this type of reading is naturally encouraging aesthetic stances. In other words, the reader is invited to experience and to feel through the text, with the purpose of engaging him/her in the story. In Rosenblatt’s words, these stances are not opposites, but rather the extremes of a continuum which needs to be balanced in the classroom.

Even if Rosenblatt established the central tenets to consider reading as a meaning-making process where all readers can contribute with their experiences and knowledge to the community of learners, few practical guidelines were given as to put her theory into practice in the classrooms. There is then a need to gather strategies, techniques, resources

and tools to put her theories into practice, as they may be beneficial to create a methodology which works on literary texts from a perspective which is still valid even if it was originated almost a century ago.

2.2 Dialogic Talks as an instructional tool to foster Transactional Reading

Dialogic talks are part of the practices of a group of educational experts, led by Prof. Ramón Flecha (University of Barcelona), who are pursuing dialogic learning. As such, these talks are considered a collaborative strategy, based on the creation of knowledge and meaning through dialogue (Flecha 1997). In the Dialogic Reading, the text becomes a complex and enriching activity where collective dialogue-author and readers produce better learning (Loza 2004: 67).

The mechanics of Dialogic Talks are quite simple. Students have to read the text indicated by the teacher and they highlight one sentence they want to share with others. They meet and share this sentence, give their opinion in a respectful and tolerating classroom atmosphere, where the teacher is another reader. Interpretations of the text are not imposed but rather negotiated among all participants. This strategy is based on a dialogic view of learning, which supports the view that learning occurs through interactions with others.

The nature and procedure of these circles are in line with Rosenblatt's proposal of collaborative reading, both combining the private sphere (the reader interacting with the text alone), and the public sphere (the reading sharing reading experiences and having contact with shared meanings of the text). Also, the text is used in an authentic context, fostering aesthetic views, which balanced the natural use of efferent stances, and encouraging critical thinking skills by analysing, discussing and associating the text with other spheres, such as the sociocultural context, the author's biography, etc.

Dialogic Talks may then fit Rosenblatt's theory, as it has been previously highlighted in the literature of the topic (see Fernández, Garvín and González 2012). Therefore, it is assumed that transactional techniques, which favour communication and interaction in the classroom, will also benefit the creation of a communicative and authentic learning environment focused on meaning, and not on form. The main purpose of this type of exchanges is to build up knowledge individually and together using language as a communication tool. In this sense, the use of literature in the bilingual classroom could be guided by these presumably favouring strategies.

2.3 Literature Review

The use of Dialogic Talks in teacher training has been subject to research in the past years. Alonso, Arandia and Loza (2008) and Fernández, Garvín and González (2012) put dialogic into play to train in-service teachers. They focused on dialogic reading seminars, as in the present study, and were interested in participants' learning and how these experiences were different from those they had in their initial training. The main purpose was to demonstrate how dialogic reading seminars may shape teachers' beliefs and practices about reading texts in their classrooms.

Concerning the benefits obtained with the use of this instructional strategy, Alonso et al. (2008) highlighted how Dialogic Talks favour participants' equalitarian dialogue and reinforce the idea of how we are transformational agents and responsible for the change in our own classrooms. They also claim that teachers reading together may influence not only

the teaching process but our lives from a more general perspective. In this line, Fernández et al. (2012) mention that teachers valued collective reading as motivating and rewarding, and compared it with their university training, which was far more individualistic. Also, using original versions and primary sources was considered interesting and valuable. An interesting finding was that teachers value the need to have a sense of understanding and flexibility in this type of lessons. Besides, they were willing to transfer this experience to their classrooms.

Concerning the use of Dialogic Talks in the Teacher Degrees, Chocarro (2013) and Foncillas and Laorden (2014) present studies to measure the impact of the use of this instructional technique in their classrooms. Chocarro (2013) carried out an experience with students taken 2nd Year of Primary Teacher Degree in the subject of Inclusive Education, where the equalitarian dialogue was fostered. In this study, undergraduates used the strategy to teach at school, and reflected on it after the experience took place. Results show that students highlighted equalitarian dialogue in the classroom, and an atmosphere of respect, fostering students' empathy, the interplay of emotions and thoughts, the promotion of reading habit, encouraging participants' communication, working on interculturality. The negative impact of this practice was also present, as the lack of participation of shy students, which could be mitigated thanks to the teacher. Students were then aware of the important role of the teacher to foster participation. The study does not mention the book used; neither does it provide information about the data gathering tool administered to the participant.

The second study, conducted by Foncillas and Laorden (2014), was developed in the Social Education Degree at the same university in which the present work was conducted. Information was gathered using observation grids and recordings and using three main elements as success indicators: the creation of a cooperative learning environment, the development of a transformative, critical and reflective capacity, and the impact on students' relationship with their affective environment (family and friends). Results show that students recognised this instructional strategy as positive for their learning and learned to work with others, respect their opinions and build their knowledge together. Also, students developed communicative skills in an environment based on egalitarian dialogue and respect.

In light of the literature in the area produced in the last years, it may be concluded that Dialogic Talks are considered to have a positive impact on both Teacher Education undergraduates and in-service teachers. All studies highlight how Dialogic Talks favour respect, understanding and communication in the classroom. Also, participants value the possibility to learn from and with others, fostering a collaborative learning environment. Fernández et al. (2012), who carried out their study with in-service teachers, suggest the need to transfer these dialogic practices to university, claiming that they may help fight the traditional model of teachers owning all knowledge and students receiving it.

A clear gap in the area is that the use of Dialogic Talks in CLIL environments has been not found at the time of developing this study. If the use of collaborative and communicative instructional techniques is to favour students' language abilities and content acquisition, it is necessary to measure their impact and reflect on how we can make better use of those. Dialogic Talks have been proved to favour students' L1 communicative skills and may also contribute to their bilingual literacy development.

3. METHOD

3.1 Context and participants

The context of the study is a university college set in Alcalá de Henares, Madrid. It is a private institution run by the religious congregation of the Marist brothers, and is administratively attached to the Universidad de Alcalá. It has been offering degrees in Teaching for more than 40 years now, and has a student population of around 1000 people. The institution encourages educational innovation using active methodologies in the classroom. One of the methodologies favoured is the use of dialogic learning in both the Teacher Education and Social Education Degrees, as proved with a previous study using Dialogic Talks in two Social Education subjects (see Foncillas and Laorden 2014).

The study was conducted in a class composed of 26 students. For the purpose of the study, only those who attended the lessons regularly account for the information provided in the present work; therefore, 20 of them finally participated in the study. They were 6 men and 14 women aged 21/22. The experience was conducted in the subject "Exploring Children's Literature in English", a compulsory subject for students studying to become Primary Teachers with an EFL (English as a Foreign Language) specialisation. The students participating were also following the bilingual itinerary at the University, which meant that all of them hold a minimum B1 level (CEFR) at the beginning of their studies and had at that time completed almost 50% of their studies in English.

The subject is delivered in three different groupings: large group (all the class together), middle group (half of the class, A or B); or seminars, where the whole group is divided into four groups (A1, A2, B1 and B2). The experience was developed during the seminars, having three meetings with groups of 5 students in November and December in 2015. The researcher was lecturing the subject and, during the seminars where the Dialogic Talks were conducted, a language assistant, an English native-speaker, was also present. He was also invited to record his observations and commented on them with the lecturer after each session.

The subject is focused on the development of four key competences. Students should be able to know resources and strategies to use literature in the primary classroom, they should also know how to develop literacy skills through literature in the EFL/CLIL classroom, and be able to plan effective lessons including texts, stories and literature. Apart from that, they are expected to use English appropriately as the language of communication and instruction in the classroom using a B2+ minimum level.

The implementation of Dialogic Talks is an activity included in the lecturer's didactic plan for the subject. The module it belongs to revolves around a famous children's book, in this case, the first book of the series *Mary Poppins*, written by P.L. Travers in 1934. The main goals of this module are to make students engage in a meaningful reading experience using Dialogic Talks, to raise students' interest in how literary texts may contribute to our learning and personal development, to make students experience techniques and strategies to use literary texts in the classroom, and to raise students critical thinking by comparing the book with other related works, such as their adaptation to the screen as films, or documentaries related to the creation of the text or the life of the authors.

The module was developed from October to December, parallel to modules 1 and 2 in the subject, covering 16 classroom hours plus 20 hours of students' autonomous work. The sessions were directed as follows:

- Session 1 (2 hours): Brainstorm information about *Mary Poppins* and create a Padlet (a virtual wallchart with notes on the data gathered in the session).
- Session 2 (homework+2 hours): Watch the movie *Mary Poppins* at home. In class, we discuss the main events in the movie.
- Session 3 (2 hours): Elicit knowledge about the author of the book, P.L. Travers. Watch *Saving Mr Banks* in class (this movie explained the adaptation of the book to the screen).
- Session 4 (1 hour): Simulated dialogic talk with the whole group and the first chapter of the book. Instructions to understand how Dialogic Talks work.
- Sessions 5, 6 and 7 (3 hours, they were repeated in the 4 seminar groups): Dialogic Talks.
- Session 8 (2 hours): In-role teacher pretending to be P.L. Travers to conduct a press-conference with students. The information delivered was taken from Lawson (2013). Reflection and discussion.
- Session 9 (2 hours): Students were asked to prepare an activity to work on *Mary Poppins'* story with Primary Students in Spain.
- Session 10 (2 hours): A group of 52 children from a local school participated in the workshop prepared by students in groups.
- Session 11 (1 hour): Final questionnaire is administered.

3.2 Aims

The study attempted to measure the impact of the use of a transactional-based tool: Dialogic Talks, as an instructional strategy to enhance students' learning (future EFL Primary Teachers) in a bilingual context. More specifically, its application will be measured in terms of how students respond to literature, and how this response may shape their learning experiences and professional development as future teachers. As students are not taking any other similar subject in the degree, it was not possible to measure this impact against a different experience using another strategy. However, students' perceptions of the experience and their learning development are collected to fill this gap in the best way possible.

3.3 Data Gathering Tools

Research tools used were questionnaires, an observation log, students' blogs and analysis of recorded sessions. It was our purpose to triangulate the data gathering tools used to pursue research validity. Responses obtained after the experience were complemented with information obtained in August 2017, a year and eight months after the experience had been conducted. At that time, the lecturer contacted the group, sent a questionnaire about their view of Dialogic Talks and asked for participation to those who had been teaching recently. Information from 5 participants, out of 20, was obtained.

Concerning questionnaires, these were delivered at the end of the experience with the purpose of gathering data related to students' perceptions and opinions of having experienced dialogic literary circles in the classrooms. The first question included in the questionnaire (see Appendix 1) was asking about their gender. Although no significant differences were expected to be found among sexes, this information was considered to prove this. Then, students were asked about whether they liked to read in Spanish and English. The next section revolved around their participation in the Dialogic Talks, and

students were questioned about whether they had participated in the Dialogic Talks (otherwise the responses were not valid), and if they found the sessions interesting. Later, questions concerning their perceptions of language and communicative development were included. More specifically, they were asked if they had felt comfortable talking in English, and if they had realised about their progress in their oral skills in English. The following questions were more open, and asked students how the Dialogic Talks were helping them, and what they had learnt with them (students were encouraged to provide with information about language, content, attitudes, or other aspects). Last, students were asked to give suggestions for improvement and to suggest the lecturer if she should keep this strategy in the future.

Throughout the dialogic talk sessions, students were asked to keep their subject blogs. These blogs are assessed in the subject, and are based on students' reflections on their lessons. Students have a blog schedule which indicates they should write their posts once every two weeks, and respond to their classmates' entries in the second week. The topics around which students need to write are not imposed by the teacher. However, students need to refer to their experiences and learning in class in the last lessons. Therefore, it was expected that there would be some entries dealing with the Dialogic Talks.

The last data gathering tools were the observation logs and the transcriptions of the recordings (the final sessions of the four groups were recorded with students' permission). The lecturer kept an observation diary she used to take notes throughout the lessons. Information gathered was primarily concerned with students' creating of collaborative learning, contributing to each other's knowledge and expertise and, secondly, to explore if students were engaged in a transactional way with the text.

4. RESULTS AND DISCUSSION

Information gathered from the questionnaires indicates that it was the first time that students experienced dialogic talks with literary works in English. They mention it is the first time they had read a book and commented on it in groups. Also, they recognise the importance of developing communicative competences in a stress-free environment, in which nobody is marking their oral proficiency or judging their points of view. In this sense, most of them mentioned the importance of being respectful and considerate of their classmates' opinions. Regarding the didactic potential of this experience, some students indicated that this experience could well be employed in their future classes, with some adaptations. It was quite revealing to know that all of them without exception recommended the lecturer to keep this activity in the future.

Concerning observation and transcript analysis, information gathered shows that students were contributing to their learning in many ways:

- Linguistic issues. The students completed their classmates' sentences or contributed with a more appropriate word when somebody else was stuck.
- Putting academic language into practice. Students show a good command of language resources to express their opinion, agree and disagree, illustrate with examples and back up their opinions. There are some instances in which they all want to speak, but the moderator acts appropriately and establishes turn-taking. In this last session, the lecturer was not moderating and asked a student to play this role instead.
- Building on previous comments. They generate new knowledge by linking what others have said.

- Going back to the book. They were concerned about not forgetting the book in their talks. There are many instances where students refer back to paragraphs in the book or keep their classmates on track referring to lines or scenes in the book.
- Cultural issues. They contribute with their views on British and American society, comparing Travers' background and the context of the book with Disney's creations. Spanish culture was not mentioned in the talks, but they were aware of the differences when creating a different product using the same story.
- Considering the timeline. Some students were aware of the time the book was written and the movie was launched. They established connections between society in the 30s, 60s, and now.
- Teaching, our profession. Many of the reflection captured in the recordings are based on their ideas about how Mary Poppins educate children and their own role as teachers. This comparison leads to insightful conversations were students are shaping their own views on education.

Regarding the level of participation, all students contributed to the group with the sentences they had highlighted in the text. It was also gathered that the role of the moderator is essential to keep students on track. It is relatively easy to depart from the book and start commenting on issues which are not directly related to the story. The lecturer modelled the role of the moderator in the first seminar, and students took this role later. One volunteer was appointed to be the moderator for half of each session, to allow more students to be in this role.

The character of Mary Poppins in the text and in the movie was also subject to discussion. One student commented on an instance in the book when Mary Poppins is displaying kind behaviour towards the children and said out loud: "I see kindness here" (Teresa S, seminar 2). Her classmates agreed and were happy to find this, somewhat relieved. This, and other instances in the seminars, may prove that in transactional reading, readers answer "not to the text but to their evocation of the text, to their own creation" (Galda 2013, p. 6).

Participants in this experience were also interested in how children's literature has evolved over time. More specifically, they mentioned that the original book could be considered not appropriate for children now. They discussed about the dangers of overprotecting kids. This comparison was also valued as rich and fruitful, as the book "(..) makes us think about society, the way they used to teach values, morals and manners. And maybe it would be very useful for us to reflect on them now" (Laura C., seminar 2).

Concerning blogs, students were given access as authors to the university blog site. Each student was requested to write six posts and comment on their peers' posts throughout the module. In the time the Dialogic Talks were used, nine students commented on this experience¹. In most of the cases they reflected on the use of 'dialogic circles, but also described some of the activities or commented on the situation depicted in the books.

Most commentaries on the Dialogic Talks experience referred to the opportunity to exchange ideas in a peaceful atmosphere, build on their classmates' ideas and perspectives, and consider this instructional strategy as to be adapted for their Primary classrooms in the future. Some students referred to Dialogic Talks as an opportunity to learn together. One example is the comment made by Carmen S. (November): "Then we start sharing ideas and I have to highlight that the conversation was fluent all the time. We were respecting each other and adding something new to other thoughts. I really like this idea because we can implement it in our future classrooms and create a debate in a relaxing atmosphere and

with plenty of meaningful ideas". It is worth noting that students were considering that the creation of a relaxed atmosphere was a crucial point to share and learn as a community. Also, they were considering this technique as a plausible teaching tool in the future.

After one year and eight months, the lecturer contacted the group to know whether they had been using literature and Dialogic Talks in their professional lives. Some former students had not had the opportunity to teach yet, but five valid responses were obtained. These now Primary Teachers all have a positive view of the experience, and they agreed on it having influenced them in considering the potential literary texts may have in the primary classrooms. However, just one of them (Inés F.) had the opportunity to implement Dialogic Talks in their classes. She is working at an English school and used the Dialogic Talks with some adaptations, as she added the use of visuals and realia to enhance students' interaction and understanding. This teacher considers that everything she has learned at University is now useful and she can understand much better why it is important to make children love books and reading.

The five former students, now teachers, agreed on having an interest in learning more about how to make better use of books in their classrooms. They also considered that Dialogic Talks are useful and can be implemented in their teaching context. Also, they all agreed in finding 'the *Mary Poppins* experience' crucial to understand how they can make the most of books in the classrooms.

5. CONCLUSIONS

The present study aimed to measure the impact of the use of a transactional-based tool, Dialogic Talks, in the learning of a group of students taking the Primary Teacher Degree. The study has demonstrated that the impact has been positive in terms of students' learning gains. This has also shown that future literacy pedagogies should consider some key elements highlighted by students as having an influence on their learning.

First, Dialogic Talks favour the creation of a collaborative-cooperative environment in the classroom. Students value the chance to share their reading experiences in a comfortable environment where they did not feel judged or assessed. This has been also pointed out in the theory by experts, such as Rosenblatt, who states that: "The primary criterion should be not whether his reactions or his judgments measure up to critical traditions but, rather, whether the ideas and reactions he expresses are genuine" (1995: 67). It seems clear that these conditions were created thanks to the Dialogic Talks. In addition, the role of the moderator was essential to conduct students' interventions. Leading discussion without imposing our own views is not an easy task; however, there are some guidelines, which are also indicated by research, which were effective, namely: "Listen well, ask contingent questions, seize opportunities to clarify or enlarge concepts and ideas, and be challenging but supportive in interactions with students" (Galda 2013).

Second, the importance of integrating different views and learn how to build our own knowledge with the help of others. This was carefully modelled by the lecturer in the first sessions, when she acted as the moderator. It is very important to let them know that all interventions are valid, but they need to be justified and clear. The teacher is just one more reader, but an informed one.

Third, the relevance of using literature in class with a meaning-making perspective. Rosenblatt insisted on the importance of making readers have a live-through experience of the text, as to evocate their reading and construct it from the private (individual) and the

public (group) sphere. In this case, students were granted with time to read at home, but were also encouraged to bring one sentence to share with the rest. This simple process made students focus on what words meant to them, and they were re-constructing the text together. This view opposes current practices focused on exploiting the text to learn grammar and vocabulary.

Fourth, the use of English for real communication purposes. Even though focus-on-form was not encouraged, some interventions in the seminars were concerned with vocabulary and expressions. Students helped each other to make meaning out of those words and to make the text their own (hence, unobtrusive scaffolding was provided in this way). In their interventions, students were using appropriate language, and were aware of the importance of the expressions they were using (respectful and semiformal), but the message they were conveyed was of the utmost importance.

Fifth, students claimed they had learned much with this experience, as the Dialogic Talks gave them the chance to know about different elements and ideas they could not have come up with alone. This also led to putting higher order thinking skills (HOTS) into play, as participants could not be content with understanding the text, they needed to analyse, synthesise, compare or evaluate.

All in all, research shows that Dialogic Talks may be influential in changing teacher trainees' views on the use of literature in their classes. Also, they contribute to creating an appropriate atmosphere for students to interact with the text, their classmates and the lecturer. Dialogic Talks foster listening skills, and the practice of turn-taking, agreeing and disagreeing, and constructing own views with the help of others.

Concerning future lines of research, this study has made an attempt to gather information about the impact of this experience on students and could gather the views of five of them one year and a half after it was conducted. It would be interesting to contact them in the future to know whether their views have changed if these former students have implemented dialogic-friendly strategies in their classes once they became professionals. Also, it would be interesting to provide students with more Dialogic Talks during their training, so as to test whether their language skills and critical thinking skills are fostered with their use. Last, it would be interesting to classify students' responses to the text, as to know whether dialogic-friendly tools are encouraging any specific type or if direct intervention from the teachers' part is needed to work on some areas which would be otherwise not dealt with.

NOTES

1 Students' quotes are cited literally. No language correction has been made.

REFERENCES

Alonso, M.J., Arandia, M. & Loza, M. (2008). *La tertulia como estrategia metodológica en la formación continua: avanzando en las dinámicas dialógicas*. *Revista Electrónica Interuniversitaria de Formación del Profesorado*, 11 (1), pp. 71-77.
http://www.aufop.com/aufop/uploaded_files/articulos/1240860668.pdf

- Carter, R. & Long, M. (1991). *Teaching Literature*. Harlow: Longman.
- Chocarro de Luis, E. (2013). *Las tertulias dialógicas, un recurso didáctico en la formación de docentes*. *Historia y Comunicación Social*, 13: 219-229.
<https://revistas.ucm.es/index.php/HICS/article/viewFile/44238/41800>
- Coyle, D., Hood, P. & Marsh, D. (2010). *Content and Language Integrated Learning*. Cambridge: Cambridge University Press.
- Fernández González, S., Garvín Fernández, R., & González Manzanero, V. (2012). *Tertulias pedagógicas: con el libro en la mano*. *Revista Interuniversitaria de Formación de Profesorado* 45, pp. 113-118.
http://aufop.com/aufop/uploaded_files/articulos/1364431359.pdf
- Flecha García, R. (1997). *Compartiendo palabras. El aprendizaje de las personas adultas a través del diálogo*. Barcelona: Paidós.
- Foncillas Beamonte, M., & Laorden Gutiérrez, C. (2014). *Tertulias Dialógicas en Educación Social: Transformando el Aprendizaje*. *International Journal of Sociology of Education*, 3(3) , pp. 244-268.
- Galda, L. (2013). *Learning from Children Reading Books: Transactional Theory and the Teaching of Literature*. *Journal of Children's Literature*, 39(2): 5-13. Retrieved from <http://www.childrensliteratureassembly.org/docs/39-2-Galda.pdf>, accessed 27 July, 2018.
- Lawson, V. (2001). *Mary Poppins, she wrote. The life of P.L.Travers*. NY: Simon and Schuster.
- Loza, M. (2004). *Tertulias Literarias*. *Cuadernos de Pedagogía*, 34(1): 66-69.
- Meyer, O., Coyle, D., Halbach, A., Shuck, K & Ting, T. (2015). A Pluriliteracies approach to content and language integrated learning – mapping learner progressions in knowledge construction and meaning making. *Language, Culture and Curriculum*, 28(1): 41-57.
- Paran, A. (2000). *Survey Review: Recent Book son the Teaching of Literature*, *ELT Journal*, 54(1), pp. 75-88.
- Rosenblatt, L. M. (1995). 1938. *Literature as Exploration*. New York: D. Appleton-Century Co. 5th Edition.
- Rosenblatt, L. M. (1978). *The Reader, the Text, the Poem: The Transactional Theory of the Literary Work*. Carbondale: Southern Illinois University Press.
- Rosenblatt, L. M. (2005). *Making meaning with texts: selected essays*. Portsmouth: Heinemann.

Panorama general y sistematización de un conjunto de errores en el aprendizaje escolar del maya en México

General outlook and systematization of a set of errors in the school learning of Mayan language in Mexico

Hamlet Antonio García Zúñiga

Instituto Nacional de Antropología e Historia, México

hamlet_garcia@inah.gob.mx

Didier A. Chan Quijano

Universidad de Oriente, México

didier.chan@uno.edu.mx

RESUMEN

En este artículo se analiza un grupo de errores en el aprendizaje del maya (Península de Yucatán, México), obtenidos en un medio escolarizado. Se establece una clasificación basada en sus características lingüísticas, la modalidad de producción (oralidad, audición, escritura) y su aceptación social. Se propone, igualmente, una comparación entre el nivel de familiaridad que tiene el estudiantado con la lengua, ya que para algunos de los individuos es una lengua de herencia, en estadios diferenciados, tanto fuera como dentro del territorio en el que se habla. Se sugiere que se presenta una interacción jerárquica entre niveles, que si no es atendida, el error no será productivo. Se advierte que este tipo de estudios (centrados en una lengua indígena) son escasos e importantes para los contextos en los que se implementan medidas que contribuyen al mantenimiento y fortalecimiento de lenguas en una condición de minorización.

Palabras clave: Lingüística aplicada, L2, errores, lenguas indígenas.

ABSTRACT

This article analyzes a group of errors in the learning of Maya (Yucatan Peninsula, Mexico), obtained in a school environment. We propose a classification based on their linguistic characteristics, the mode of production (orality, hearing, and writing), and their social perception. A comparison is also proposed between the level of familiarity the student has with the language; some of the individuals has the language as heritage system, in differentiated stages, both outside and within the territory in which it is spoken. We found a hierarchical interaction between levels, that if it is not addressed, the error would not be productive. By another hand, these types of studies (focusing on an indigenous language) are scarce and important for the contexts in which the government implements measures that contribute to the maintenance and strengthening of languages in a condition of minority.

Keywords: Applied linguistics, L2, errors, Indigenous languages.

1. INTRODUCCIÓN

En diciembre de 2019, el Congreso Estatal de Yucatán (sureste de México) aprobó, por unanimidad, una serie de reformas a la constitución de la entidad con el propósito de hacer obligatoria la enseñanza de la lengua maya (familia lingüística mayance) en todas las escuelas del territorio en el nivel básico. Sin embargo, la aplicación de esta determinación no es inmediata y esta se hará de forma paulatina, mientras se publica en el Diario Oficial del Estado, se "armonizan" los cambios con la Ley de Educación local, se definen partidas presupuestales, se diseña la coordinación del programa y se capacita a quienes fungirán como docentes, en el conocimiento gramatical de la lengua, así como en lo estrictamente pedagógico.

La medida, que es ejemplar en el país, posee varias aristas de interés (en algunos casos de crítica o escepticismo). En primer lugar, el objetivo que se persigue es doble; por un lado, promover esta lengua en las comunidades yucatecas en las que su presencia sea baja y, por el otro, fortalecer su uso en aquellas poblaciones en las que el número de hablantes sea alto y su tendencia sea la transmisión, así como la funcionalidad (ver en la sección Generalidades). En segundo lugar, sin duda alguna, la propuesta tiene como eje el orgullo identitario que se extiende entre un amplio sector de la sociedad, en el que la lengua, tanto en una representación de contacto (el español yucateco), como en una manifestación de legado (la antigua civilización maya), juega un papel fundamental. Por último, en contraste con lo anterior, es innegable que, como en buena parte de los lugares en donde conviven dos o más lenguas, es posible encontrar actitudes sociales e históricas que fomentan, entre muchas otras cuestiones, exclusión, presión, vergüenza y pérdida lingüística. En este sentido, quienes promovieron las reformas reconocen que ha sido el mismo Estado, ya sea en el plano federal, el regional o el municipal, el que no ha diseñado, implementado y ejecutado políticas públicas orientadas al rescate y el fortalecimiento lingüístico. Independientemente de este punto, se debe reconocer que estas reformas constituyen un avance hacia el reconocimiento y la justicia que merece gozar el pueblo maya.

Hay que llamar la atención sobre el hecho de que el esfuerzo y la demanda por, respectivamente, enseñar y aprender la *maayáat'aan* (como la nombran quienes hablan la lengua) siempre han existido. En este sentido, la experiencia ya existe y esta se ha dado, tomando en cuenta la iniciativa, desde múltiples perspectivas: la oficial, la institucional, la personal (con apoyo), la individual (sin apoyo). En cualquier caso, los resultados han sido siempre los mismos: falta de continuidad, programación de niveles elementales, deserción, así como carencia de sustento en los contenidos y su organización. Por lo anterior, revisar uno de los muchos aspectos que integran el aprendizaje de una lengua, como lo es la producción y asimilación de errores, sin tocar necesariamente la forma en la que se enseña, puede representar una oportunidad de mejora.

En efecto, el trabajo con errores ha arrojado en otras lenguas (inglés, español, ruso, chino, por citar los ejemplos más referidos en la bibliografía especializada) un perfeccionamiento en el sistema de enseñanza. Vale decir que el presente estudio no constituye, en ningún sentido, un modelo de instrucción, aunque sí contiene elementos que son indispensables para esta, en cuanto progresión en el conocimiento y su dominio.

Sin lugar a duda, en el caso que se presenta son muchas las cuestiones que se tienen que atender previamente; por ejemplo, la homologación de la escritura de la lengua, la formación del profesorado, la asignación de recursos suficientes para sostener los programas de enseñanza, sean públicos o privados, los fines específicos por los cuales se oferta el aprendizaje de la lengua y la discriminación tanto social como lingüística que

padece la sociedad maya actual (la cual ya se aludió líneas arriba). No obstante, se considera que es necesario contar con una exploración de un trabajo inexistente a la fecha, el cual, como todo, requiere de una revisión constante.

La investigación se realizó entre estudiantes de las cinco licenciaturas que se imparten en la Universidad de Oriente (Valladolid, Yucatán): administración pública, bibliotecología y gestión de la información, desarrollo turístico, gastronomía, lingüística y cultura maya, y mercadotecnia. Esta escuela de educación superior está enclavada en la región oriental de la entidad, que es, junto con la parte sur, donde hay mayor concentración de hablantes de la lengua (132 180, en 2005), los municipios que la conforman detentan porcentajes significativos en este rubro, el uso de la lengua es una norma social (en cuanto a intención comunicativa y empleo en todas las situaciones), es relativamente común el monolingüismo y su transmisión intergeneracional es alta. Estas características obligaron junto con los elevados índices de pobreza y la amplia brecha educativa que tiene la población maya, respecto a la sociedad yucateca y mexicana, a analizar la viabilidad de abrir un centro de educación superior en la zona, lo que se logró en 2005.¹

A pesar del panorama descrito, no es raro que soliciten su ingreso a esta universidad estudiantes mayas que no hablan la lengua o, incluso, personas de otras poblaciones de México, donde no se habla maya. Por esto se imparte la asignatura de "Seminario de lengua maya" en todos los programas de estudio, salvo el de lingüística y cultura maya (el perfil de ingreso obliga a ser hablante de la lengua), durante nueve cuatrimestres. La intención es enseñar y homologar el conocimiento del maya. Lo interesante, como se verá, es que, aun cuando se hable maya, por diferentes circunstancias, se cometen errores.

2. OBJETIVO

Este es un estudio que pertenece a un proyecto más amplio sobre los errores en el aprendizaje del maya; su corte es descriptivo y en él se plantea identificar, describir y catalogar una serie de errores detectados e identificados como frecuentes en el interior de una clase formal de lengua maya, tanto en un nivel de expresión (oral y escrita), como en uno de comprensión. Lo anterior permitirá esbozar un análisis que combina un contraste entre los rasgos de quienes producen los errores, la percepción que de ellos tiene el cuerpo docente de la escuela, así como el vínculo de los errores con los diferentes niveles de análisis lingüístico. Esto último corresponde a la distinción entre criterios de dominio; en sentido estricto, el lingüístico, el comunicativo y el estratégico superficial. Por tanto, quedan fuera del análisis los criterios de origen y los pedagógicos.

Con esto lo que se busca es sentar un precedente en los estudios de L2, cuando esta es indígena (o minorizada), que permita idear y planificar la mejor manera en la que se puede trabajar el error en cursos formales (más adelante se brindan más detalles sobre esto), al igual que contribuir en el desarrollo de programas de estudio, técnicas de enseñanza coherentes con las realidades de la propia comunidad lingüística maya, así como con los recursos del medio, además de la elaboración de material didáctico.

3. CONSIDERACIONES METODOLÓGICAS

En este apartado se tratan aspectos que sirven para contextualizar la investigación que aquí se presenta. Primeramente, se anticipa lo que se entiende por un "error" producido durante el aprendizaje de una segunda lengua. Posteriormente, se explicitan y justifican las

variables que entran en juego en el estudio, así como la postura que se toma respecto al binomio lengua-escuela en la situación que se atiende. Por último, se detalla el procedimiento que se siguió para coleccionar y ordenar la información.

3.1. Fundamentos teóricos de los que se parte

En general, se puede afirmar que el error en el aprendizaje de lenguas adicionales a la materna cuenta con un doble papel en el salón de clases. Uno orientado a quien enseña y otro dirigido a quien aprende. Esto es, en principio, el error evidencia tanto la visualización del nivel que se ha alcanzado en la comprensión y asimilación de los principios que orientan la comunicación en sus distintas modalidades (gramatical, social, cultural, pragmática) teniendo en consideración un objetivo concreto, como la manifestación del control y la eficiencia que se tiene en el uso lingüístico. Al respecto, conviene especificar que el sustento teórico de la presente investigación dista de aquellos primeros análisis en el campo, como el de Corder (1967), en el que se distingue entre "equivocación", "falta" y "error" para resaltar que este se refiere a una desviación en la producción verbal como producto del desconocimiento de las reglas.

Este distanciamiento entre propuestas posibilita que exista un mayor entendimiento de lo que es y significa un error. La idea que aquí se defiende se orienta en este sentido, por lo que, necesariamente, implica una actitud activa en el salón de clases, así como un compromiso. Lo primero tendría el propósito de alcanzar la competencia de "aprender a aprender" (capacidad para proseguir, persistir y organizar el aprendizaje propio),² en tanto que lo segundo perseguiría la realización constante de análisis sistemáticos de las sesiones de trabajo. En consecuencia, un error tiene el potencial de ser una muestra del aprendizaje de una lengua (Cf. Richards, 1974; Larsen-Freeman & Long, 1995; Torrijano, 2004; Alba, 2009; Vázquez, 2009). Con esto en mente, la finalidad de esta investigación es proporcionar pistas sobre la manera en la que se atiende, reflexiona, descubre, aprende y emplea el maya.

En la bibliografía consultada se establece una distinción entre los errores de actuación y los errores de competencia. El material que se estudia en esta investigación es del segundo tipo, debido a su sistematicidad. Esta clase de error es la que permite reconstruir el conocimiento que posee y alcanza el alumnado de una lengua.

3.2. Postura

Históricamente las relaciones entre escuela y lengua en la región estudiada no han sido buenas. Efectivamente, este representa uno de los ámbitos sociales en los que se ha excluido el uso del maya. Por esta razón, al hablar de cursos escolarizados en Yucatán, aunque estos sean de una segunda lengua, se remite, por lo menos, a tres reflexiones que hay que tener en consideración.

Para quienes hablan la lengua: se debe de aceptar que la escuela constituye un contexto más en una comunidad de habla, no es el único ni el más importante, ya que la lengua se seguirá adquiriendo y reproduciendo en el hogar y la calle. Esto no quiere decir que se tenga que dejar de lado la aspiración de que la educación sea en y sobre la lengua.

Para quienes no hablan la lengua, pero están inmersos en el espacio y los dominios en los cuales se emplea cotidianamente: la escuela no representa una opción de transmisión lingüística cuando se está cerca de gente *mayera*³ y cuando, en este ámbito, se estima más provechoso ejercer un aprendizaje activo; es decir, un aprendizaje en el que se requiere

tomar la iniciativa y no culpar a nadie porque no se benefició con la práctica de la lengua desde la infancia.

Para quienes no hablan la lengua, pero se encuentran fuera del espacio y los dominios en los cuales se emplea día a día, inclusive, pueden ser personas totalmente extrañas al universo maya, los cursos escolarizados se convierten en una opción (tal vez la única) asequible. Este punto ofrece la oportunidad para vivir la otra cara de la interculturalidad que se ha impuesto en México (Cf. Dietz, 2017); a saber, la que no responde a intereses de castellanización ni integración de minorías, sino aquella que se acerca a quien se ha definido, desde múltiples plataformas, como "otro(a)", como "diferente", como "ajeno(a)" y que, como consecuencia, deconstruye su pensamiento, su percepción, su expresión y su actuar.

3.3. Variables consideradas

De lo expuesto en el apartado anterior se justifica la necesidad de establecer variables que respondan a las circunstancias que irremediamente se encuentran en Yucatán. Aunque en un sentido estricto se debe contemplar la inclusión de la naturaleza de los cursos (formal: escolar con contenidos, planificación y evaluación; informal: sin los rasgos mencionados), se tomó la determinación de que en este trabajo solo se abordaría la temática de los errores a partir del tipo de estudiante.

Los contextos que, en este sentido, se generaron están fundamentados en el origen del alumnado. Por tanto, se contemplan estudiantes, en un primer plano, yucatecos(as), no yucatecos(as) residentes, así como extranjeros(as) y, en un segundo plano, que están inmersos(as) en el contexto maya, o bien, que no tienen este contacto comunitario. La incorporación de las actitudes y los comportamientos que las personas (interés u orgullo por las manifestaciones culturales) muestran hacia la lengua se deja para otro estudio.⁴ En la Tabla 1 se proporciona un resumen y una combinación de estas características (la información no corresponde al universo del cual se extrajeron los datos, se trata, más bien, de una proyección de lo que se podría encontrar en el aula).

Origen del alumnado	Con inmersión en un contexto maya		Sin inmersión en un contexto maya	
	Habla maya	No habla maya	Habla maya	No habla maya
Yucatecos(as)	X	X	X	X
No yucatecos(as)	X	X	X	X
Extranjeros(as)	X	X	X	X

Tabla 1. Rasgos de los estudiantes del estudio

3.4. Metodología

La totalidad de la información que aquí se examina procede de observaciones en grupo; para estas se contó con la autorización por parte de las autoridades de la institución educativa y de quienes durante la fase de investigación fungieron como docentes. El trabajo de observación se realizó en 2018 y duró ocho meses (dos periodos escolares); tiempo en el cual se visitaron y se dio seguimiento a cinco aulas, correspondientes a cuatro licenciaturas (dos grupos de gastronomía, uno de administración pública, uno de mercadotecnia y uno de bibliotecología) de los cuatrimestres intermedios (6º) y últimos (8º).

Se puso atención en la dinámica de la clase, en especial en los ejercicios y las participaciones, así como en la revisión de portafolios de evidencias. También se consideraron las notas y las bitácoras del cuerpo docente con la finalidad de cotejar los objetivos de enseñanza y el tipo de error que se produjo. Lo anterior se complementó con entrevistas a docentes; a través de estas se pudieron conocer las opiniones o posturas que este núcleo de profesionales tiene respecto a la detección del error, su significado e importancia en el aprendizaje, así como su valoración. Por cuestiones de espacio, este material no se consideró para este artículo.

Las visitas a los grupos en los que se trabajó se realizaron una vez por semana (10 por grupo, excluyendo los días de asueto y los de examen, al igual que las actividades extracurriculares) y duraron toda la sesión; esto es, una hora. Al final se logró reunir evidencia procedente de cien visitas.

La atención se centró en tres formas de emergencia del error: la oralidad (saludos, preguntas, respuestas, exposiciones temáticas, comentarios a lo mencionado por otras personas, presentaciones de las actividades realizadas en el día), la expresión escrita y la comprensión auditiva. Para la primera modalidad, los errores se anotaron en el modelo de ficha que se presenta en la Tabla 2 tomando en cuenta únicamente la vía por la cual se externó el error, la posible corrección (en este caso también se colocó la forma correcta) y el sujeto que lo detectó, así como su naturaleza. Al concluir cada sesión se tuvieron reuniones con el cuerpo docente responsable de la asignatura (cuatro docentes hombres), en las que se les preguntó por los objetivos del tema, así como su consecución, los errores detectados, los rasgos que definían al estudiantado que los había cometido, las razones por las cuales se corrigieron (o no), los planes para su corrección, la valoración del aprendizaje y los motivos que justificaban la detección del error. De esta manera se recabaron opiniones o juicios referentes a la labor docente que se podían confrontar con lo que el observador de la clase había percibido; esto es válido y trascendente, sobre todo, para la detección del error y su incorporación pedagógica en el aprendizaje de la lengua. Como ya se ha establecido, este asunto no se trata en este estudio. Sin embargo, lo cierto es que se tomaron en consideración exclusivamente los errores validados por los profesores. En este sentido, algunos errores quedaron fuera de este análisis.

Por su parte, la detección de los errores de expresión escrita fue tratada en las reuniones con los docentes. Las pruebas fueron tareas, composiciones hechas en clase y exámenes. En estos casos, además de lo ya dicho arriba, se añadía el intento de corrección (fallido o no) hecho por los(as) estudiantes. En este caso, si se percibían tachones, borrones o marcas que tapaban el error, se consideraba solamente una oportunidad de enmendarlo.

Ahora bien, respecto a los errores no considerados (aquellos que solo fueron detectados o registrados por los observadores) se tiene que señalar, en primer lugar, el porqué se tomó la decisión de trabajar de esta manera y, en segundo lugar, las consecuencias de esto. Así pues, respectivamente, si se hubiera encomendado la tarea de registro al docente, en algún momento, este podía distraerse de la clase, la detección o el registro. Además, con esta forma de trabajo se presentaba la alternativa de investigar un tema de sumo interés y de gran valía: la percepción del docente sobre el error. Por otro lado, mediante este proceder se inhibió la posibilidad de combinar valoraciones distintas del error. Por último, toda la información obtenida se encuentra en el archivo del proyecto y será analizada en momentos diferentes.

Al ser este un análisis de acercamiento al tema, en el que se pretende brindar un panorama general y sugerir una sistematización de la información sin hacer ningún tipo de afirmación concluyente, no se ofrece una revisión estadística de la información recopilada.

En este sentido, cuando se haga referencia a algún tipo de error o a un ejemplo de este, se van a presentar lo que se denominan "casos" y no los llamados "tokens".

¿Se corrigió?	¿En otras ocasiones desapareció el error?
¿Se trabajó?	¿Se explicó en qué consistió el error?
¿Quién lo detectó?	Quien da la clase, quien comparte la clase, quien produjo el error
Corrección	Establecer la forma correcta
Tipo de error	Fonológico, morfológico, sintáctico, semántico, pragmático
Tipo de estudiante	Escribir todos los números de la clasificación elaborada, que describan al estudiante
Forma de producción	¿Cómo se produjo el error? ¿Es un error escrito, hablado, escuchado? Mencionar si el error se produjo en la clase en la que se enseñó el tema del cual surgió o fue en una clase diferente
Error	Anotar el error en contexto

Tabla 2. Ficha de registro de datos

Por cuestiones de privacidad (resguardo de información personal), aun cuando se cuenta con la autorización de parte de quienes colaboraron en la investigación para estudiar y presentar en trabajos académicos los errores que produjeron en el aula durante el tiempo que duraron las observaciones, en ningún caso se proporcionan mayores detalles de los grupos.

4. GENERALIDADES

La lengua maya, también conocida como maya yucateco, maya peninsular, yucateco, maya, o, como ya se mencionó, *maayáat'aan*, pertenece, junto con el lacandón (sur y norte), el Itzaj y el Mopán, a la rama yucatecana de la familia lingüística mayance. En la actualidad se habla en algunos cuantos poblados de Belice, una buena parte de los estados mexicanos de Campeche y Quintana Roo, y en la totalidad de los municipios de Yucatán. Desde hace algunos años, debido a un proceso migratorio, también es hablado, mayormente, en la bahía de San Francisco, California.

Según los registros del más reciente de los Censos Generales de Población y Vivienda de 2010 (Instituto Nacional de Estadística y Geografía, 2013) se alcanzaron a contabilizar 786 422 hablantes mayores de 5 años en México; su distribución se presenta en la siguiente tabla. Como se podrá observar, Yucatán es el territorio peninsular en el que hay un mayor número de hablantes del maya.

Entidad federativa	Hablantes mayores de 5 años
Yucatán	547 098
Quintana Roo	163 477
Campeche	75 847
TOTAL	786 422

Tabla 3. Número de hablantes de maya por entidad mexicana (Elaboración propia con datos del INEGI, 2013)

Diferentes organismos han elaborado índices que miden la vitalidad de la lengua. En el caso del maya, *Ethnologue* lo sitúa como un sistema en desarrollo (o sea, se encuentra en las etapas iniciales de desarrollo de escritura, estandarización, modernización), en tanto que el Instituto Nacional de Lenguas Indígenas la ubica en un mediano riesgo de desaparición; esto sucede en dos situaciones generales: la primera, cuando la proporción de infantes hablantes de la lengua es menor a 25% y cuenta con más de una localidad con 30% y más de hablantes, o bien, tiene más de mil hablantes en localidades con 30% y más de estos y, la segunda, cuando la proporción de niños hablantes es mayor a 25% y solo hay una localidad con 30% y más hablantes, o bien, tiene menos de mil hablantes en localidades con 30% y más hablantes.

En cuanto al sistema fonológico (importante para las consideraciones que aquí se hacen), el maya cuenta con un sistema vocálico basado en cinco timbres contrastivos (los mismos que el español) con rasgos de duración (breves y largas), estado glótico y tonalidad, así como con un sistema consonántico con nasales, laterales, aproximantes, fricativas, oclusivas y africadas; entre estas dos últimas existe un contraste glotal en las sordas. En las Tablas 4 y 6 se muestra la totalidad del sistema.

Timbres vocálicos			
	Anterior	Central	Posterior
Alta	i		u
Media	e		o
Baja		a	

Tabla 4. Vocales del maya

Duración	Breves	i, e, a, u, o
	Largas	i: (ii), e: (ee), a: (aa), u: (uu), o: (oo)
Estado glótico	Glotales	iʔ (i'), eʔ (e'), aʔ (a'), uʔ (u'), oʔ (o')
	Rearticuladas	iʔi (i'i), eʔe (e'e), aʔa (a'a), uʔu (u'u), oʔo (o'o)
Con tono alto		i:1 (ii), e:1 (ée), a:1 (áa), u:1 (úu), o:1 (óo)

Tabla 5. Otros rasgos de las vocales del maya
(entre paréntesis aparece la grafía empleada en la escritura)

El tono del maya ha sido objeto de discusión recientemente. Uno de los problemas radica en determinar cuáles son los contrastes tonales en la lengua; dicho de otra manera, ¿a cada realización (breve o larga, y glotal simple o rearticulada) se le asigna un tono? Según Sobrino (2013: 160), hay un acuerdo "en que las vocales largas presentan tonos fonológicos". Al respecto, este autor afirma lo siguiente:

[...] la duración parece depender del rasgo suprasegmental de la vocal, es decir, si tiene tono alto, tono bajo o si es rearticulada. El tiempo de realización de las vocales largas con tono bajo es menor que el de las vocales largas con tono alto; por su parte, el tiempo de duración de las vocales rearticuladas es mayor que el de las vocales largas con tono bajo, pero menor al de las vocales largas con tono alto (Sobrino, 2012: 161).

En (1) se observan algunos ejemplos en los que se puede ver y justificar el contraste entre los rasgos de las vocales que se describieron arriba.

- (1) a. Entre vocales breves y largas
kan `cuatro'
kaan `serpiente'
ch'om `piñuela'
ch'oom `zopilote'
- b. Entre vocales largas y rearticuladas
baal `cuñado' (respecto al varón)
ba'al `cosa'
ooch `zarigüeya'
o'och `comida'
- c. Entre vocales largas y tonales
aak `hierba gramínea'
áak `tortuga'
ch'iich' `cascajo'
ch'íich' `ave pequeña'
- d. Entre vocales rearticuladas y tonales
cha'ak `sagú' (palmera con frutos ovoides)
cháak `lluvia'
si'is `entumecerse alguna parte del cuerpo'
síis `frío'

Puntos de articulación	Glotaes	Son							
		Sor	ʔ	h					
	Velares	Son							
		Sor	k	kʔ					
	Palatales	Son							
		Sor			ʃ	tʃ	tʃʔ		
	Dentales	Son							
		Sor	n	t	tʔ	s	ts	tsʔ	l
	Labiales	Son	b		w				
		Sor	m	p	pʔ				
Modos de articulación	Nasales	Simples	Glotaes	Fricativas	Simples	Glotaes	Aproximantes	Laterales	
		Oclusivas			Africadas				

Tabla 6. Inventario de consonantes del maya (Sor es abreviatura de sordo y Son lo es de sonoro)

5. ANTECEDENTES

En México existen pocas investigaciones sobre la enseñanza y adquisición de segundas lenguas que tienen como base una lengua indígena (Gómez, 2015; Hernández, 2020; por ejemplo). Esto, desafortunadamente, no difiere de lo que se puede encontrar en el medio indoamericano. Hay que evitar confundir esta situación con la existencia de cursos, presenciales o en línea, con buen material de soporte o sin él, para el aprendizaje de lenguas indígenas. En este sentido, son notorios los ejemplos del Lakota, el quechua y algunas lenguas mayas, como el Kaqchikel y el mismo maya yucateco en universidades estadounidenses y francesas. Este último caso justifica las breves líneas que se dedican a la ideología de aprendices y enseñantes sobre las lenguas de aprendizaje, enmarcada bajo el término lenguas con fines específicos e, inclusive, las lenguas de instrucción.

Es bien sabido que la concepción del error ha transitado desde la equivocación, pasando por la falta de dominio, hasta la [a]gramaticalidad. En un sentido teórico, el error ya no comporta ningún tipo de connotación estigmatizante; por el contrario, en la actualidad solo tiene connotaciones positivas. Como se ha sostenido previamente, el error es fundamental para el aprendizaje. En este sentido, se ha adoptado una orientación pedagógica en su tratamiento y, en consecuencia, los estudios en la materia sacan a la luz dificultades reales (las cuales también hay que saber distinguir) y son un método eficaz para evaluar el proceso de aprendizaje, las metodologías de enseñanza y los factores "controlables" que influyen en el éxito de estos, así como del dominio de la lengua.

6. DATOS GLOBALES

Recuérdese que en este artículo se revisa el conjunto de los errores que produjeron distintos grupos de estudiantes universitarios con características diferentes. La intención, como ya fue mencionado, es acercarse al tema mediante una descripción general y una propuesta de sistematización, teniendo en cuenta una lengua que tiene condiciones muy divergentes de aquellas en las que se ha formulado la teoría sobre el error.

6.1. Características globales de los cursos ofertados

Con el propósito de completar la imagen del contexto en el cual se realizó el estudio (ya se sabe que la lengua y sus hablantes experimentan condiciones de uso o desuso muy particulares en el día a día, y que la escuela no ha sido un aliado histórico de la población yucateca), a continuación, se proporciona información referente a los cursos que se suelen programar en Yucatán para aprender maya.

Es común encontrar cursos que justifican su organización de acuerdo con el Marco Común Europeo de Referencia para las Lenguas (MCERL). Sin importar que las características de las lenguas europeas y con fines específicos establecidos difieran de aquellas que poseen las lenguas indoeuropeas, como la maya, lo que sobresale es la falta de estudios gramaticales en estas para sostener tal afirmación y, consecuentemente, poder plantear niveles, contenidos y un sistema de evaluación. Así, por lo general, los cursos existentes (1) no rebasan el segundo semestre, (2) tienen una deserción muy alta, (3) privilegian la gramática y la escritura, (4) no desarrollan suficientemente un enfoque comunicativo, (5) no cuentan con material didáctico idóneo, (6) los imparten personas con entusiasmo, pero sin preparación suficiente o en el área, y (7) se basan en una organización por clases gramaticales (una clase o un curso para revisar sustantivos, por ejemplo) aisladas y fuera de contexto. Para el caso que aquí se analiza, inclusive, los temas propuestos para cada grado (cuatrimestre) exceden las horas para trabajarlas en el aula (10 temas en 26 horas), con lo que no se alcanza a cubrir el programa.

6.2. Errores, tipos y descripción

Los errores recabados para la realización de esta investigación ascendieron a poco más de 150. Entre ellos, el más común fue el hablado (68=45.3%), aun cuando el enfoque comunicativo no se privilegió durante las clases; a este le siguió el escrito (60=40%), por lo que al final quedó el auditivo (22=14.7%). Este conjunto de errores se clasificó según su naturaleza lingüística; a saber:

- A. Fonológica: pares mínimos, pronunciación de fonemas, metátesis, generalizaciones, simplificaciones, sustituciones, elisiones, equivalencias para el verbo del español 'tener'.
- B. Léxica: negación, demostrativos.
- C. Morfológica: plurales, pronombres, tematizador, marcas derivativas.
- D. Sintáctica: orden de palabras.
- E. Semántica: clasificadores.

En el caso de los errores fonológicos se tiene que considerar que estos no solo fueron diversos, sino que también los más frecuentes y numerosos. En el siguiente apartado se ejemplifica y define cada uno de estos errores.

6.3. Ejemplos de los errores

Para cada uno de los ejemplos que se presentan se toma en cuenta (i) el contexto (en negritas se marca el elemento en el cual hay que poner la atención), (ii) la forma correcta (es la primera que aparece en cada numeral), (iii) una traducción y (iv) una definición gramatical.

6.3.1 Negación

La negación en maya tiene cuatro formas: la simple, expresada mediante una partícula (*ma'*); la compuesta, que se conforma por dos partes y empleada, básicamente, en oraciones predicativas, la partícula mencionada y un clítico de cierre (*i'*) que se adjunta al último elemento de la frase; la conjuntiva, equivalente al 'ni' del español (*mix*) y la que atañe a este apartado, la de existencia ('no hay', 'no'+tener), cuya forma es *mina'an*. El error (ver ejemplo 2) en este caso consiste en un procedimiento reflexivo interesante que no se le presentó al grupo: negación simple + verbo tener/haber.

(2) *mina'an in w-alak' peek' → ma' yaan in w-alak' peek'*
'no tengo/hay mi (doméstico) perro' → 'no tener/haber mi (doméstico) perro'

6.3.2 Clasificadores

Como prácticamente todas las lenguas mayances, el maya cuenta con un sistema de clasificadores que dependen de la naturaleza del referente (un tanto reducido si se le compara con otras lenguas mayas como el tsotsil, por ejemplo). Estos son de dos tipos: los nominales (marcan a algunos sustantivos), ejemplo 3, y los numerales (acompañan a los números en conteos), ejemplo 4. En los clasificadores nominales fue frecuente la omisión de este, en tanto que en los numerales se emplearon los dos más generales (aquellos que distinguen la animacidad) y, en ocasiones, inclusive, prefiriendo el de inanimados sobre el de animados.

(3) *yaan óox-túul in w-alak' miis → yaan óox-túul in mis*
'hay tres-animado(s) mi(s) doméstico(s) gato(s)' → 'hay tres-animado(s) mi(s) gato(s)'

(4) *jun-túul kaax → jum-p'éel kax*
'uno-animado gallo' → 'uno-inanimado gallo'

6.3.3 Plurales

Los plurales son marcas opcionales que en algunos contextos pueden producir lecturas diferentes a la de número nominal. Sin embargo, cuando se cuenta, el morfema de plural es agramatical, porque se estaría redundando en la idea.

(5) *óox-túul chi'ik* → *óox-túul chi'ik-o'ob*
'tres-animado(s) tejón' → 'tres-animado(s) tejones'

6.3.4 Pronombres

Las marcas pronominales en los verbos que comienzan por vocal se componen de dos elementos: uno que se pega a la marca de aspecto (*-in*, en el ejemplo) y otro que se integra al verbo (en este caso, *w-*). En este rubro, el error estuvo en que esta segunda marca no se usó.

(6) *t-in w-éemel* → *t-in éemel*
PROGRESIVO-yo subir → PROGRESIVO-yo subir
'estoy subiendo'

6.3.5 Tematizador

Cuando se le quiere atribuir cierta relevancia en el discurso a un elemento oracional se emplea una marca de foco o temática (*-e'*). Debido a que es difícil entender el significado de este morfema, se suele omitir.

(7) *Raúl-e' in suku'un* → *Raúl in suku'un*
'Raúl es mi hermano mayor' o 'Raúl, mi hermano mayor, ...[predicación]'

6.3.6 Equivalencias ('tener')

En maya, las construcciones posesivas del tipo 'tener' numeral + alguien o algo se forman sin verbo; algo similar ocurre en las oraciones existenciales, la diferencia entre estas y las del verbo 'tener' es que en las primeras no hay un verbo 'ser', en tanto que sí hay uno para 'tener'.

(8) *ka'atúul in kiik*
'dos-animada(s) mi(s) hermana(s)'

Lo que se dijo en el ejemplo fue 'tengo dos hermanas', pero lo que se entendió fue 'dos mis hermanas', 'dos de mis hermanas', 'dos son mis hermanas'.

6.3.7 Demostrativos

Las unidades que se usan para indicar distancia respecto de los participantes del acto de habla constan de dos partes (se trata de un morfema discontinuo, como una de las negaciones revisadas en 6.3.1): la primera es una marca que introduce o presenta el nominal y, la segunda, es la marca que establece la distancia en tres términos (cercana, relativamente lejana y no perceptible). Durante la observación, fue común la omisión de la segunda parte de los demostrativos.

(9) *le leek-o' in ti'al* → *le leek in ti'al*
DEMOSTRATIVO recipiente para colocar tortillas de maíz (es) mío
'ese recipiente (para colocar tortillas de maíz) es mío'

6.3.8 Pares mínimos fonológicos

La distinción entre sonidos que contrastan en la lengua en un plano fonológico conlleva muchas dificultades. Este es un error bastante frecuente.

(10) *míis* → *miis* (se confunde una vocal larga con tono alto y una larga)
escoba, barrer → gato

6.3.9 Orden sintáctico

El orden en las frases nominales simples definidas es Determinante, Adjetivo, Nominal. El error fue la inversión de este orden.

(11) *ch'ujuk waaj* → *waaj ch'ujuk*
dulce masa → masa dulce
'pan (masa dulce)'

6.3.10 Pronunciación

Como se refirió en 6.3.8, es difícil distinguir entre los sonidos contrastivos de la lengua. Pero el primer problema que se presenta es la pronunciación del inventario completo de estos. Un error llamativo fue el que se produjo con el fonema fricativo, glotal, sordo (h),⁵ ya que, por alguna razón, varios individuos lo articularon como una fricativa, palatal, sonora (como en 'yunta').

(12) *juntúul* [**h**un.túul → **j**un.túul]

6.3.11 Marcas derivativas

Las raíces verbales del maya requieren, con ciertos aspectos que refieren al estado en el que se encuentra un evento (si comienza, está en desarrollo o concluyó), unas marcas que permiten la flexión. Estos elementos son la transitivización (-t), la causación (-s) o un cero morfológico. Durante las sesiones de los cursos que se observaron fue común la confusión de estas unidades (intercambios entre ellas) y la elisión. Hay que referir que este puede llegar a ser uno de los temas más difíciles de comprender de la gramática de la lengua, por lo que se requiere un uso adecuado.

(13) *táan in jaan-t-ik bu'ul* → *táan in jaanik bu'ul*
PROGRESIVO yo comer-TRANSITIVIZADOR-modo frijol
'estoy comiendo frijol'

Jaan es la raíz del verbo *janal* 'comer' y es intransitiva. De esta manera, para flexionarse requiere transitivizarse cuando el aspecto lo obliga.

6.3.12 Metátesis

Este es un cambio que ocurre en la posición de ciertos sonidos en una cadena de habla. En los casos anotados se ve una tendencia a realizar este proceso fonológico cuando hay más de un mismo sonido en la palabra, como se ve en el ejemplo 14 con la [k].

(14) *kukupkij* → *kupukkij*

6.3.13 Generalizaciones

Este error es exclusivo de los sonidos que tienen contrastes glotales. En el ejemplo que se muestra (independientemente de la elisión de la aspiración representada con una "j" en medio de la palabra) ocurren dos generalizaciones: una con la africada dental sorda simple y su correspondiente glotal, y otra con la vocal rearticulada. El interés del error es que ocurre cuando ya se aprendió el contraste y todos los fonemas simples se articulan como glotales.

(15) *tsajbij* → *ts'a'abij*

6.3.14 Simplificaciones

Es una especie de generalización, pero a la inversa: del rasgo difícil al fácil.

(16) *ko'one'ex* → *konex*

6.3.15 Sustituciones

Aunque tiene forma de generalización, no lo es, ya que en este error no involucra un grado de dificultad entre fonemas. Además, sus posibilidades de producción se reducen a la africada palatal sorda (ch), la fricativa palatal sorda (x) y la fricativa dental sorda (s).

(17). *wich* → *wix*

6.3.16 Elisiones

Es la eliminación de un sonido dentro de una palabra.

(18) *bejla'e'* → *belae'* (esta elisión propicia el cambio de intensidad en las sílabas)

6.4 Características del error

Los errores que se presentaron en los apartados anteriores son una muestra heterogénea de lo que se percibió y se detectó como tal. Como podrá observarse, el ámbito de producción se reduce a la palabra, aun cuando estén insertos en una frase (nominal o verbal). Esto remite a la forma en la que se diseñó y se actualiza el curso; esto es, no hay evidencia de errores que pertenezcan a niveles mayores de análisis lingüístico o de una conversación básica, útil y cotidiana.

Llama la atención que los errores que se consignaron se produjeron a lo largo de la sesión en la que se enseñó el tema. Esto, por un lado, se interpreta como el primero de los pasos para dominar una característica o un elemento de la lengua. Por tanto, se asume que es comprensible la producción. Sin embargo y, por el otro lado, también puede entenderse como una falta de atención de los docentes en la detección y el seguimiento de los errores en las clases subsiguientes. Esto representa una gran dificultad para alcanzar la finalidad del aprendizaje y así satisfacer las necesidades de quien aprende.

Lo anterior habla, de igual manera, de la dinámica que se desarrolla en el aula, en donde la práctica no es lo principal y priva la idea de que el docente es el que enseña, es el único agente y, por tanto, el aprendizaje es una actividad, paradójicamente, pasiva. Asimismo, cuando el docente asume este papel debe de hacerlo en su totalidad. Por esta razón, la facultad y autoridad para detectar errores es exclusiva de él. Con esto, quien aprende se autoexcluye de una parte muy importante del aprendizaje de una lengua: la

colaboración. El manejo en el aula de la situación entre iguales es fundamental, como ha sido propuesto en varios textos que explican la interlengua y las emociones.

Dos escenarios prácticos que complementan la idea anterior son la corrección del error y si este fue trabajado o no en las clases. Respecto al primer punto, como se ha insistido, la responsabilidad de la detección recayó en su totalidad en los docentes. De esto derivó que los encargados de la corrección fueron, de igual manera, los docentes.

Ahora bien, en referencia al segundo punto se distinguen tres momentos: el posterior a la producción del error, durante las sesiones en las que se revisa el tema y a lo largo del curso. En el primer contexto, sí hubo una atención del error, mediante su señalización y corrección, además de una pequeña práctica o ejemplos que servían como idea de la estructura correcta. En el segundo y el tercer momento, no hubo mención alguna a los errores del curso, salvo cuando uno de estos volvía a surgir. Aunque se debe de referir que esta situación no fue la que prevaleció en la clase; es decir, no siempre se atendió el error que había surgido con anterioridad. En resumen, se puede afirmar que fue muy poca (en frecuencia, tiempo dedicado y complejidad) la atención que recibió el error como materia de trabajo en el aula.

7. SISTEMATIZACIÓN DE LA INFORMACIÓN

Como se recordará, los errores detectados durante el trabajo de observación fueron catalogados según la modalidad de su producción y el nivel de análisis lingüístico en el que pueden ser analizados. En la Tabla 7 se puede observar la conexión que tienen estas dos variables con otras, como son su frecuencia y la percepción que de ellos tienen los docentes. Así, se verá que esta última variable de análisis cambia de estatus entre más profundo (representa un tipo de significado o contenido gramatical) es el nivel; es decir, cuando se trata de niveles básicos, como el fonológico y el léxico, el error se percibe de la misma manera y es "explicable" o "entendible", en los términos de los docentes, debido a que el maya no tiene los mismos sonidos del español y estos son más complicados que los de esta lengua.

Por lo anterior, el profesorado concuerda, en una catalogación de los datos según su "gravedad". Otros errores, como los que se dan en el manejo de los clasificadores, son tolerables, porque incluso algunos hablantes nativos los cometen.

Plano	Modalidad			Frecuencia	Percepción
	Escritura	Oralidad	Audición		
Fonológico	X	X	X	Alta	Entendible
Léxico	X	X	-	Alta	Entendible
Morfológico	X	X	-	Alta	Tolerable
Sintáctico	X	X	-	Alta	Grave
Semántico	-	X	-	-	Tolerable
Discursivo	-	-	-	-	-
Pragmático	-	-	-	-	-

Tabla 7. Interconexión de variables del estudio

Algo que también es observable es que la frecuencia de producción de cada uno de los errores del *corpus* es alta. En este sentido, lo que llama la atención es que no hubo, a lo largo de los dos periodos de observación, una forma acertada de reducir su producción. Por otra parte, aparentemente, la comprensión auditiva es la modalidad de error de la que se tiene mayor control. Sin hacer alusión a un procedimiento estadístico de análisis, parece

ser que la audición solo genera equivocación en la percepción de los sonidos; algo semejante a lo que se ejemplifica en el ejemplo 17. Este aspecto requiere de mayor estudio porque, como ya se ha explicado, ni los cursos ni las clases se conducen bajo un principio de comunicación, entre otras muchas cuestiones. Esta afirmación se refuerza si se considera la contraparte, la oralidad, que representa la modalidad en la que se encuentran todos los tipos de error consignados. Lo interesante es que en la escritura, donde se cuenta con mayor control y cuidado, tanto por el tiempo de reflexión como por la posibilidad de corrección, se dan con frecuencia buena parte de los errores.

7.1 Otras particularidades de los datos

Los errores con los que se tiene mayor certeza de registro son aquellos de los niveles o planos básicos. Hay estructuras más complejas (como las que son propias de la semántica y la pragmática) en las que no se registran errores, porque simplemente no se revisan en la clase ni son parte del temario.

Merece una consideración especial, no obstante que pueda interpretarse como algo contradictorio el trabajo con los errores que se producen en la oralidad. En efecto, a juicio de los docentes, este tipo de error no se corrige en el momento, como sí pasa con el resto, porque se teme que el grupo pierda confianza y espontaneidad, y porque con estos errores "se aprende".

7.2 Interrelaciones

En la Tabla 8 se presentan los errores de acuerdo con el tipo de estudiante que los produjo. Entre las personas que participaron en el estudio solo hubo hablantes de maya en la categoría de estudiantes de Yucatán con ascendencia maya. Entre quienes se consideraron como estudiantes de Yucatán sin ascendencia maya sin contexto hubo individuos que tenían un conocimiento elemental de la lengua.

De inmediato resalta el hecho de que el error no es exclusivo de un solo tipo de estudiante. Es cierto que los motivos variarán, pero el docente no debe de dar nada por sentado. Si bien en esta situación parte del estudiantado no va a aprender la lengua debido a que ya la habla, es menester que se tenga en cuenta que es posible conocer y aprender gramática. Lo que se notó es que los errores de este grupo de estudiantes son parte de la variación existente en la lengua, derivada o no de la clase de interferencias que se suelen citar en los estudios sobre contacto lingüístico.

En la Tabla se coloca una serie de números. Estos corresponden al orden que en esta investigación se les dio a los errores. Con el propósito de que se recuerden, se repiten en este apartado:

1. Negación
2. Clasificadores
3. Plurales
4. Pronombres
5. Tematizador
6. Equivalencias
7. Demostrativos
8. Pares mínimos fonológicos
9. Orden sintáctico
10. Pronunciación

11. Marcas derivativas
12. Metátesis
13. Generalizaciones
14. Simplificaciones
15. Sustituciones
16. Elisiones

Estudiantes de Yucatán				Estudiantes de fuera de Yucatán			
Con ascendencia maya		Sin ascendencia maya		Con ascendencia maya		Sin ascendencia maya	
CC	SC	CC	SC	CC	SC	CC	SC
2, 3, 6	2, 3, 6, 7, 8, 9, 10	-	1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 13, 14, 16	1, 2, 3, 4, 5, 7, 9, 13, 14, 15, 16	1, 2, 3, 4, 5, 7, 9, 10, 12, 13, 14, 15, 16	1, 2, 3, 4, 5, 7, 9, 10, 11, 12, 13, 14, 15, 16	1, 2, 3, 4, 5, 7, 9, 10, 11, 12, 13, 14, 15, 16

Tabla 8. Tipo de estudiante y error producido (CC=Con contexto, SC=Sin contexto)

Entre los errores analizados, el 2 y el 3 se presentaron con todos(as) los(as) estudiantes. En cambio, el 6 solo apareció con los estudiantes originarios de Yucatán; el 8 solo fue producido por dos de los siete grupos contemplados; el 12 y el 15 únicamente se registraron con estudiantes de fuera de Yucatán; tanto el 11 como el 16 estuvieron presentes en cinco de los siete grupos. Lo cierto, de manera global, es que quienes tuvieron acceso a un contexto maya fueron propensos, con variaciones (la excepción es el grupo de estudiantes de fuera de Yucatán que no tienen ascendencia maya), a cometer menos errores. Como era de esperar, los(as) estudiantes de fuera de Yucatán fueron propensos(as) a hacer errores.

7.3 Sistematización de la información

En la siguiente Figura se interpreta la forma en la que transcurre el error en el aprendizaje de maya. En ella se aprecia que los niveles fonológico y léxico están conectados, por lo que el error en uno repercute en el otro. Los otros niveles, aunque cuentan con reglas particulares que las hacen muy lejanas a las que se tienen en español, no muestran una conexión ni entre ellos ni entre los niveles fonológico y léxico porque hay algo en el curso que lo impide. Una evaluación de las clases señalaría que estas solo examinan los sonidos y las palabras, y no se adentran a otros espacios de la lengua.

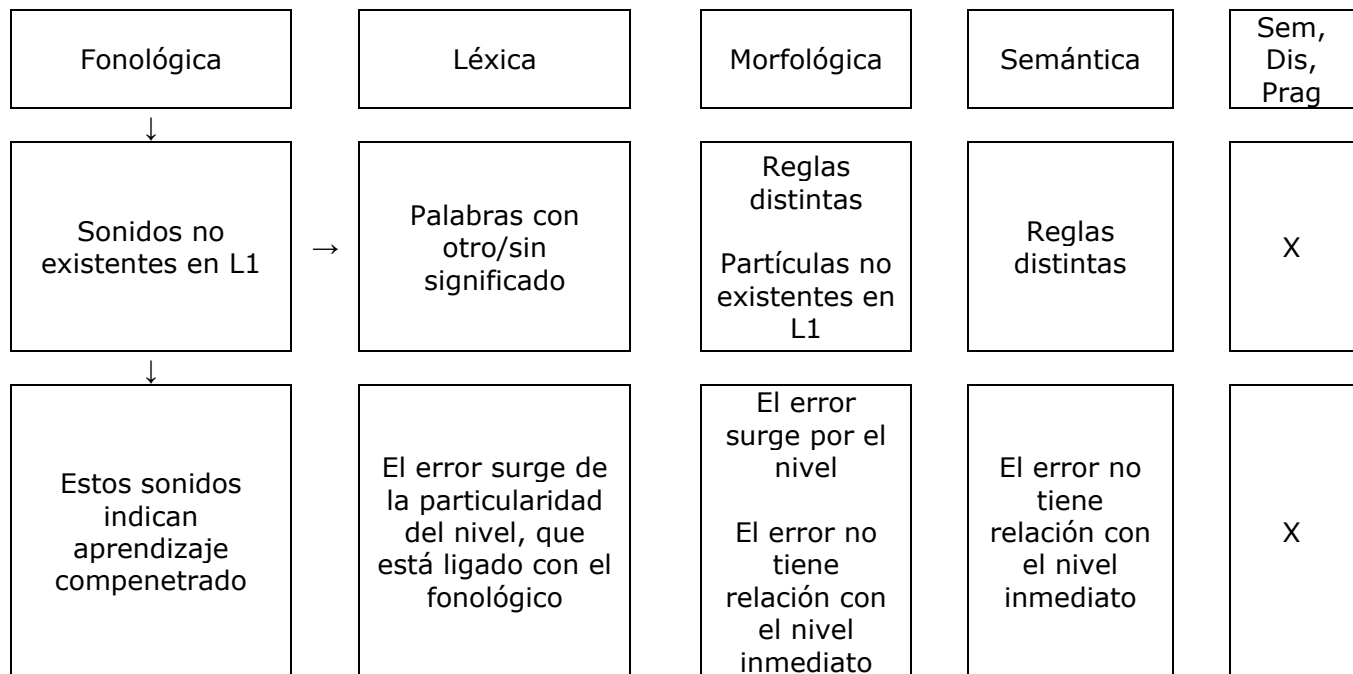


Figura 1. Vía en la que se vinculan los errores detectados (Sem=semántica, Dis=discurso, Prag=pragmática)

8. ALGUNAS CONSIDERACIONES FINALES

De este tipo de trabajos surgen tres tipos de conclusiones: unas vinculadas con los datos, otras pensadas como el paso siguiente o la aplicación del análisis y unas más que se formulan como pendientes de investigación.

Según la información que se manejó, se pudo observar que los errores (i) se encuentran en los niveles más elementales de la lengua e (ii) inhiben, dificultan o retrasan el aprendizaje de cuestiones más complejas (por esto no se avanza en la elaboración de oraciones con las que se pretenda comunicar una idea), y (iii) varían en complejidad dependiendo del nivel de análisis lingüístico. Estas conclusiones obligan a poner más atención en la fonología. Esto, junto con otras cuestiones, se tiene que atender antes de la división y asignación de niveles para los cursos.

También se vio que hace falta material didáctico especializado y que los docentes acostumbran a impartir aspectos generales de la lengua, en un sentido gramatical, además de que se revisan pocas cuestiones con un enfoque comunicativo y el estudiantado, comúnmente, deja los cursos debido a la falta de interés (motivación por acceder a un siguiente nivel que no se vislumbra) y visión para emplear la lengua, así como la tendencia a contar con cursos "desorganizados" o poco planificados. Sería deseable que no se dependiera de la escritura y que se ponderara más la práctica, así como la oralidad en el aula.

En otro sentido, es visible que surge la idea de para qué sirve estudiar la lengua cuando se ve su complejidad. De igual manera, se plantea que no es conveniente organizar los cursos con base en categorías gramaticales (además de que se debe de procurar entrelazar los niveles de análisis): esto hace que el estudio de la lengua sea independiente de su uso. Con el propósito de que los errores sean verdaderos instrumentos de aprendizaje, es vital que quienes participan se involucren en la detección y asimilación de los errores. Por esto,

se debe propiciar una mayor comunión en el grupo para hacer más activo el aprendizaje y que sean los iguales en papel y condiciones (el estudiantado) quienes corrijan el error y tomen conciencia de la reflexión sobre la aceptabilidad de los errores.

En un futuro el tema podría revisarse desde una perspectiva particular: examinar por separado cada uno de los aspectos que aquí se desarrollaron para alcanzar mayor detalle. Así, se nota que muchos son los temas que quedan pendientes en esta investigación. Independientemente de la planificación que se podría hacer para dar seguimiento a cada uno de los individuos del alumnado, la búsqueda de más ejemplos (al igual que la incorporación de estructuras más complejas), de su manejo estadístico, de su correlación con factores como su utilidad y manejo en el aula, así como la percepción que tanto docentes como estudiantes tienen de ellos, en algún momento se requerirá hablar sobre los fines específicos de la lengua, que son los que hacen que se muestre interés por aprenderla.

Como se sabe, la enseñanza de la lengua para fines específicos se centra en los procesos de enseñanza y aprendizaje que facilitan el dominio de una comunicación especializada. Para el caso del maya, se deberán considerar los contextos más comunes que viven quienes se interesan por aprender la lengua: el laboral o el académico, por ejemplo, además de privilegiar el enfoque comunicativo (las condiciones están presentes: poblaciones en las que el uso de la lengua es cotidiano, hablantes bilingües, cercanía entre las localidades, comunidades mayormente rurales o semiurbanas, así como familiares, amistades y vecinos(as) que dominan la lengua).

Otro elemento nuclear para la mejora de los cursos, que se relaciona con los fines específicos y vale, por tanto, la pena investigar es la serie de necesidades que tiene el estudiantado para estudiar la lengua. Con esto, en un primer momento, se podría garantizar una respuesta adecuada a las expectativas de cada sujeto. Dicho de otra manera, es muy raro que se revisen temas de interés concreto para el alumnado, de acuerdo con su campo de especialidad. Por último, no hay duda alguna de que las ideologías y las actitudes lingüísticas negativas impiden desarrollar fines específicos para lenguas nacionales habladas en México, entre las que se encuentra la maya. Pero si no se conocen, describen y lideran la investigación en el campo, la tendencia será repetir la situación que aquí se remarca. Por esta razón, se optó por presentar un panorama general sobre los errores en el aprendizaje en una de estas lenguas. Este es el principio indispensable para comenzar a plantear proyectos más amplios y específicos, a la vez.

AGRADECIMIENTOS

Se agradece y reconoce la participación de Alejandra Moo Batún, Gregorio Hau Caamal y José Cipriano Dzib Uitzil en una primera etapa (2015) del proyecto de identificación, recolección y análisis de errores durante el aprendizaje del maya en diferentes medios. Igualmente, gracias a los cuatro docentes y al estudiantado de los grupos en los que se recolectó la información, por comprender la trascendencia y posible aplicación del estudio, así como por autorizar el empleo de los datos que generaron durante sus clases.

NOTAS

1 Los municipios que, según la administración pública del estado, pertenecen a la región oriente de Yucatán son: Cantamayec, Cuncunul, Chacsinkin, Chankom, Chemax, Chichimilá, Chikindzonot, Dzitás, Kaua, Peto, Tekom, Temozón, Tinum, Tixcacalcupul, Uayma,

Valladolid y Yaxcabá. En 2005 había 527 279 hablantes de maya en Yucatán, por lo que los 132 180 de la región oriente corresponden a poco más del 25% del total. Por otro lado, la tasa de monolingüismo de la población hablante de maya mayor de 5 años a nivel estatal en el año 2000 fue de 8.8, la decimotercera más alta del país.

2 Comisión para la Educación de la Unión Europea.

3 Individuos que se reconocen como hablantes de maya.

4 En otro tipo de cursos, no siempre de naturaleza informal (en los términos en los que se describe en este artículo), suelen asistir profesionistas que requieren aprender maya porque les ayuda considerablemente en el desenvolvimiento de su trabajo: profesionales en antropología, arqueología, epigrafía o turismo, por ejemplo.

5 En realidad este fonema se articula un poco más adelante del punto velar (es una aspiración y no una velar).

REFERENCIAS BIBLIOGRÁFICAS

- Alba, V. de (2009). El análisis de errores en el campo del español como lengua extranjera: algunas cuestiones metodológicas. En *Revista Nebrija de Lingüística Aplicada*, 5 (3) (pp. 1-16).
- Corder, S. P. (1967). The significance of learner's errors. En *International Review of Applied Linguistics in Language Teaching*, 5 (4) (pp. 161-170).
- Dietz, G. (2017). Interculturalidad: Una aproximación antropológica. En *Perfiles Educativos*, 39 (156) (pp. 192-207).
- Embriz, A. & Zamora, Ó. (Coords.) (2012). *Lenguas indígenas nacionales en riesgo de desaparición: Variantes lingüísticas por grado de riesgo*. México: Instituto Nacional de Lenguas Indígenas.
- Hernández, M. (2020). La enseñanza del tsotsil en las escuelas de enfermería del estado de Chiapas. En *Revista Latinoamericana de Educación y Estudios Interculturales*, 4 (3) (pp. 59-70)
- Larsen-Freeman, D. & Long, M. H. (1995). *Introducción al estudio de la adquisición de segundas lenguas*. Madrid: Gredos.
- Polinsky, M. & Kagan, O. (2007). Heritage languages: In the "wild" and in the classroom. En *Language and Linguistics Compass*, 1 (5) (pp. 368-395).
- Richards, J. (1974). *Error Analysis: perspectives on second language acquisition*. Londres, Inglaterra: Longman.
- Shiló, F. (2011). *Análisis de procesos de enseñanza y aprendizaje del tsotsil como segunda lengua en la Universidad Intercultural de Chiapas, México*. Tesis de maestría. Universidad Mayor de San Simón. Cochabamba, Bolivia.
- Sobrino, M (2013). Descripción fenética de los tonos del maya yucateco. En *Estudios de Cultura Maya*, 41 (pp. 157-173).
- Torijano, J. Agustín. (2004). *Errores de aprendizaje y aprendizaje de los errores*. Madrid: Arco Libros.
- Vázquez, G. (2009). El concepto de error: estado de la cuestión y posibles investigaciones. En *Actas del Congreso de la Lengua Española* (pp. 497-504). Sevilla: Instituto Cervantes.

Revista
Nebrija

de
Lingüística
Aplicada a la
Enseñanza
de las
Lenguas

