

**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*)<sup>1</sup>. 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	
<b>Modo</b>	Escrito/oral	<b>Tipo y modo de recolección</b>	Transversal/longitudinal Manual/digital	<b>Internas: Cognitiva afectiva</b>	Edad/estilo de aprendizaje Motivación/actitud
<b>Género</b>	Diario/ensayos personales	<b>Elicitación</b>	Espontánea/ Por instrucción	<b>L1</b>	Lengua materna del aprendiente
<b>Tipo</b>	Narrativo/Descriptivo/Expositivo Argumentativo	<b>Uso de referencias</b>	Diccionario/ texto fuente	<b>L2 segunda extranjera</b>	L2/ LE
<b>Tópico</b>	General/específico	<b>Limitación de tiempo</b>	Libre/controlado	<b>Nivel de Competencia</b>	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	
<b>Modo</b>	Escrito/oral	<b>Tipo y modo de recolección</b>	Transversal/longitudinal Manual/digital	<b>Internas: Cognitiva afectiva</b>	Edad/estilo de aprendizaje Motivación/actitud
<b>Género</b>	Diario/ensayos personales	<b>Elicitación</b>	Espontánea/ Por instrucción	<b>L1</b>	Lengua materna del aprendiente
<b>Tipo</b>	Narrativo/Descriptivo/Expositivo Argumentativo	<b>Uso de referencias</b>	Diccionario/ texto fuente	<b>L2 segunda extranjera</b>	L2/ LE
<b>Tópico</b>	General/específico	<b>Limitación de tiempo</b>	Libre/controlado	<b>Nivel de Competencia</b>	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 reabrador *off-line* para optimizar la transcripción de grabaciones de corpus orales. La técnica de reabrado *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.