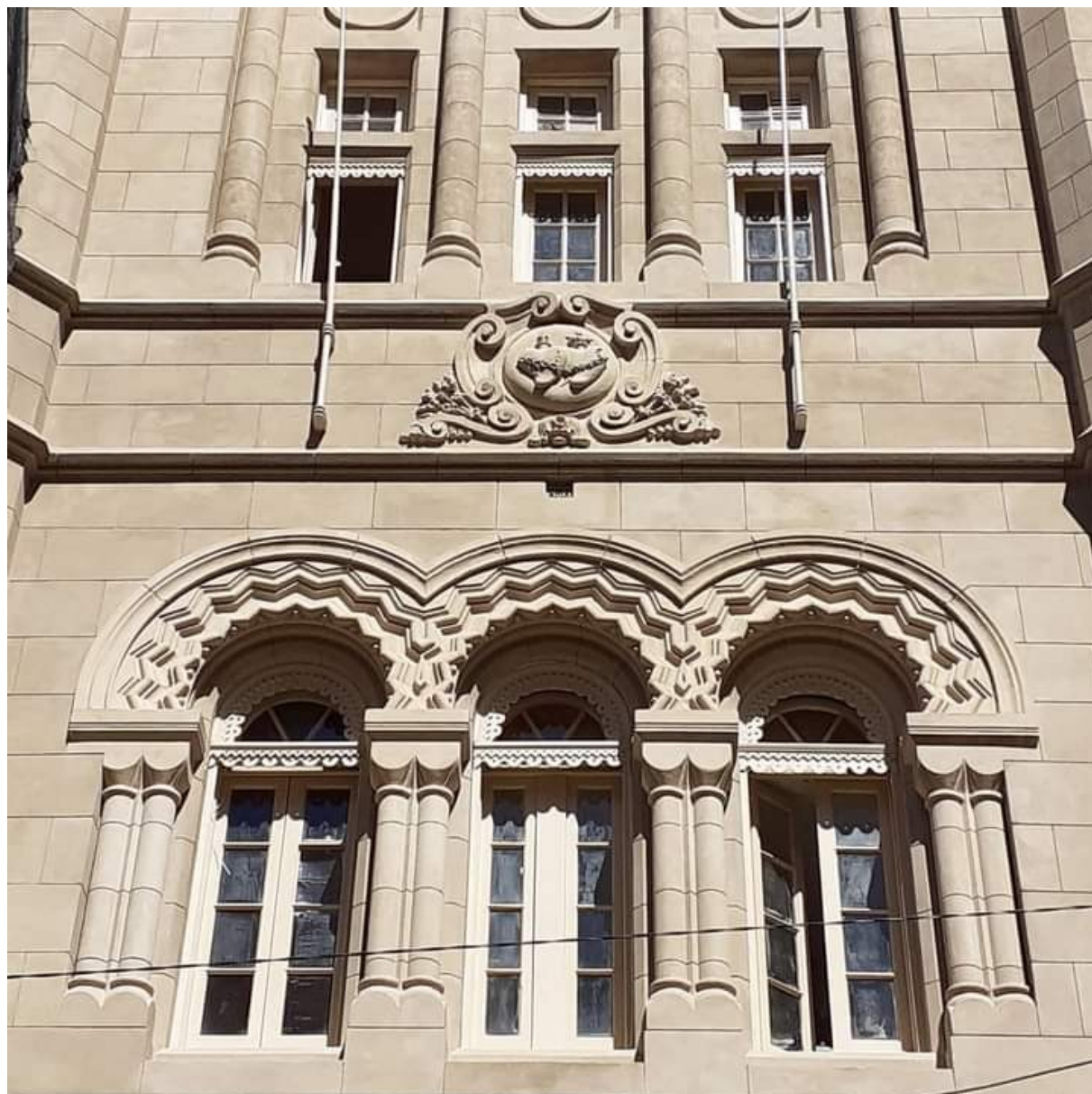


## ***HOMENAJE A ZULEMA SOLANA***

***Estudios sobre lingüística computacional, adquisición y enseñanza  
de lenguas***



**Editado por Carolina Paola Tramallino**

Homenaje a Zulema Solana : estudios de lingüística computacional, adquisición y enseñanza de lenguas / Carolina Paola Tramallino ... [et al.]; contribuciones de Barbara Méndez ... [et al.]; coordinación general de Carolina Tramallino; editado por Carolina Paola Tramallino. - 1a ed . - Roldán: Carolina Paola Tramallino, 2020.

Libro digital, PDF

Archivo Digital: descarga y online

ISBN 978-987-86-7863-4

1. Lingüística. I. Tramallino, Carolina Paola II. Méndez, Barbara, colab. III. Tramallino, Carolina, coord. IV. Tramallino, Carolina Paola, ed.

CDD 410

“Cada vez iré sintiendo menos y recordando más, pero qué es el recuerdo sino el idioma de los sentimientos, un diccionario de caras y días y perfumes que vuelven como los verbos y los adjetivos en el discurso, adelantándose solapados a la cosa en sí, al presente puro, entristeciéndonos o aleccionándonos vicariamente...”

*(Rayuela, Julio Cortázar)*

## Índice

A manera de Prólogo.....	1
<b>Capítulo I. Lingüística computacional.....</b>	<b>11</b>
Comparación de dos técnicas de clasificación supervisada en la categorización de textos y evaluación en datos simulados: Árboles de clasificación y Regresión Logística <i>Celina Beltrán y Ivana Barbona.....</i>	12
Definición de categorías y formulación de propiedades en 5P <i>María José González Díaz y Cristina Pippolo Griego.....</i>	32
Usos de programas informáticos de traducción para la corrección del error en el aula de ELE: debates y propuestas <i>Julia Francés.....</i>	43
El análisis automático de la sílaba escrita del español mediante la herramienta NooJ y su aplicación pedagógica <i>Rodolfo Bonino y Andrea Rodrigo.....</i>	64
Análisis automático de sintagmas preposicionales del español con el sistema NooJ <i>Carolina Caserio y María Paula Garófolo.....</i>	81
<b>Capítulo II. Adquisición y enseñanza de lenguas.....</b>	<b>93</b>
Adquisición del wichí: hacia una metodología para el estudio del lenguaje infantil en una lengua nativa argentina <i>Andrea S. Taverna y Migdalia I. Padilla.....</i>	94
La concepción del error en la adquisición del español lengua segunda o extranjera (ELE/2) - Un análisis de los procesos de derivación sufijal español/portugués y los casos de divergencias morfológicas <i>Natalia Ricciardi.....</i>	116
La gramática del español. Perspectivas teórico-descriptivas y metodológicas en la enseñanza como L1 y L2 <i>Nora Múgica.....</i>	131

Producción escrituraria, argumental y sintáctica, en alumnos universitarios. Un caso representativo: el empleo de “el mismo” y sus variantes con valor anafórico <i>María Cristina Calvi y María Susana Freidenberg</i> .....	147
<b>Capítulo III. Análisis del Discurso</b> .....	161
Lenguaje Cliché y disposición escénica: la perspectiva de Alfred Lorenzer y Víctor Klemperer <i>Diego Alberto Beltrán</i> .....	162
Análisis de la argumentación en el alegato final del juicio a las Juntas Militares en Argentina <i>Carolina Paola Tramallino</i> .....	177

## A manera de Prólogo

"La vida no tiene sentido pero vale la pena vivir, siempre que reconozcas que no tiene sentido".

Albert Camus

Este libro nace de las ganas de brindar homenaje a la Dra. Zulema Solana, más conocida como “Chichina” por todos y todas las que transitamos la querida Facultad de Humanidades y Artes. Quienes como yo, tuvimos la dicha de conocerla profundamente sentimos que nuestra carrera estuvo marcada por su persona. La mía, indudablemente, estuvo signada por su “halo protector”. No creo que haya estudiante, docente o no docente que no recuerde su presencia, sus clases, sus dichos, su voz fuerte, sus pasos en los pasillos y su lucha en las reuniones de Consejo Directivo.

Chichina fue una docente e investigadora que contemplaba el futuro, vanguardista, innovadora. A pesar de haber comenzado como ayudante de lengua griega había virado hacia las nuevas tecnologías y los desarrollos computacionales. Era ella quien nos transmitía el gusto y la curiosidad que debe tener todo investigador, que convertía en sencillo lo complicado, quien podía sintetizar y explicar las teorías lingüísticas más profundas y difíciles.

Era ella, con la palabra justa y el consejo que esperábamos, actuando siempre en consecuencia a sus convicciones, sin ceder en nada, sin claudicar. Quien tornaba “el miedo en prudencia, el dolor en transformación, los errores en iniciación y el deseo en compromiso”<sup>1</sup>. Militante política defensora de la Democracia, había renunciado a lo más preciado: sus cargos docentes universitarios en la época de la Dictadura Militar.

El proceso de llevar a cabo la idea de homenajearla me permitió conocer a otros colegas que querían participar, que sentían la necesidad de rendirle tributo. Fue conmovedor asistir al cariño que habita en cada uno y una de quienes participamos en el proyecto, escuchar sus anécdotas y las huellas que Zulema dejó en sus vidas.

Para mí, Chichina fue muchísimo más que mi directora de tesis, fue una guía, una consejera, una sabia que supo, con la generosidad y humildad de los grandes, otorgarme un lugar, no solo en su cátedra o en su grupo de estudios, sino en su corazón. Me dejó innumerables consejos y enseñanzas que irán tomando sentido cada vez más a medida que transcurra “el tiempo”, o mejor dicho, mi tiempo.

Por lo tanto, este libro es la manera que encontré para decirle GRACIAS.

*Carolina Paola Tramallino*

---

<sup>1</sup> Nassim Nicholas Taleb define así al estoico, es decir, a la persona que vive según los preceptos de la corriente filosófica denominada estoicismo.

La presente obra reúne once artículos de investigación de compañeros y compañeras, estudiantes y colegas que fueron enriquecidos por el recuerdo de clases, proyectos, encuentros y charlas con la Dra. Zulema Solana. Entre ellos, contamos con las valiosas contribuciones de sus hijos: Celina y Diego Beltrán, a quienes agradecemos, especialmente, la participación.

El libro se divide en tres capítulos. El primero integra a los manuscritos encuadrados en la Lingüística computacional. El que inicia esta sección, a cargo de Celina Beltrán e Ivana Barbona, compara el desempeño de dos métodos de clasificación: Árboles de clasificación y Regresión Logística. Dicha comparación se realiza sobre una aplicación en categorización de textos y una evaluación sobre datos simulados bajo distintos escenarios. Luego, el estudio de María José González y Cristina Píppolo muestra el distinto comportamiento de dos clases de artículos en la conformación de los sintagmas determinados a partir de la formalización lingüística realizada desde el paradigma 5P. A continuación, el artículo de Julia Francés analiza, de manera teórica, el rol de los programas informáticos de traducción en el aula de ELE y su implicancia en los materiales didácticos. Tiene como objetivo poner en discusión la relación entre la traducción, el enfoque comunicativo y la corrección en el aula. Luego, el trabajo de Rodolfo Bonino y Andrea Rodrigo propone una perspectiva pedagógica a partir del análisis automático de la sílaba escrita a través del sistema NooJ, para eso se centra en cinco tipos de modelos silábicos del español. Por último, el estudio de Carolina Caserio y María Paula Garófolo exhibe la posibilidad que presenta la herramienta NooJ para el reconocimiento automático de sintagmas preposiciones en textos de español, a través de la confección de gramáticas sintácticas.

El segundo capítulo se refiere a la adquisición y enseñanza de lenguas. Comienza con el trabajo de Andrea Taverna y Migadalia Padilla que expone la metodología de una investigación en curso sobre lenguaje infantil en lengua nativa wichí. El trabajo muestra los resultados que se han obtenido respecto de la adquisición léxica (que incluye la forma de las primeras palabras y su organización semántica) y las prácticas de socialización tempranas. Sigue el artículo de Ricciardi, el cual realiza un relevo del tratamiento del “error” desde diferentes corrientes dentro de las teorías de adquisición de segundas lenguas con el objeto de analizar contrastivamente la derivación sufijal en portugués y español. A continuación, el artículo de Nora Múgica realiza un aporte a la propuesta de enseñanza de gramática del español como segunda lengua, desde distintas perspectivas y enfoques. Para finalizar, cierra el artículo de María Cristina Calvi y María Susana Freidenberg, el cual presenta una

investigación sobre el empleo del “mismo” y sus variantes con carácter anafórico por parte de estudiantes universitarios de la carrera de Letras.

Es necesario hacer mención, con profunda tristeza, del fallecimiento de María Cristina, unos meses después de haber participado en este proyecto. A continuación, la recordamos a través de las palabras de su compañera y amiga, coautora del artículo “Producción escrituraria, argumental y sintáctica, en alumnos universitarios. Un caso representativo: el empleo de “el mismo” y sus variantes con valor anafórico” que forma parte de este volumen.

María Cristina Calvi inició la carrera de Letras con la llegada de la democracia en 1984. Ya egresada pasó a integrar el cuerpo académico de Lengua Española III en la U.N.R, desempeñándose en todas las instancias, desde Ayudante de cátedra hasta Titular. Compañera de tareas e ideales de María Susana Freidenberg, se dedicó con amor y responsabilidad a su verdadera pasión, la docencia. En ese marco, participó junto al equipo de Lengua Española III en numerosos PID, Eventos Científicos Nacionales e Internacionales y publicaciones referidas al ámbito de la Gramática de la Lengua Española.

El capítulo III se compone de dos estudios sobre análisis del discurso. El primero de Diego Beltrán analiza la relación entre el lenguaje cliché y la escena privada e histórica. El objetivo es examinar las expresiones del lenguaje que impiden la reflexión sobre las acciones en el plano ideológico-político y en diferentes contextos históricos. El último, de Carolina Paola Tramallino devela, desde un análisis textual, los argumentos esgrimidos por el fiscal Julio Strassera en su alegato final en el juicio a las Juntas Militares.





Zulema Solana fue la directora del Centro de Estudios de Adquisición del Lenguaje fundado en 1990, perteneciente a la Facultad de Humanidades y Artes. Miembro de la Cátedra UNESCO para la Lectura y Escritura. Profesora Titular de las cátedras Lingüística General I y Lengua Española III de la carrera de Letras de la UNR. Investigadora invitada en GRIL (Groupe de Recherche dans les Industries de la Langue), Universidad Blaise-Pascal Clermont Ferrand, Francia, septiembre-octubre 2003. En 1976 el Gobierno Militar la dejó cesante y retornó recién en 1984 a sus cargos universitarios con el advenimiento de la Democracia. En cuanto a los cargos de gestión, fue vice decana de la Facultad de Humanidades y Artes de la Universidad de Rosario desde mayo de 1994 a noviembre de 1998 y fue Directora de la Escuela de Posgrado en la misma casa de estudios desde 2014 hasta su deceso.

Dirigió numerosos proyectos de investigación referidos a la adquisición del español como lengua materna y extranjera, al análisis automático y a la lectura de textos electrónicos en la educación media. Entre ellos, el más relevante: “INFOSUR: Investigación y desarrollo” proyecto de Lingüística informática con la asesoría del Dr Gabriel G. Bès de la Universidad Blaise-Pascal Clermont-Fd Francia (2004- 2008), proyecto de alto impacto SPU/ Secyt.

A partir de este proyecto surgió la revista digital INFOSUR. Revista de Lingüística Informática, Modelización e Ingeniería Lingüística disponible en [www.infosurrevista.com.ar](http://www.infosurrevista.com.ar) con alcance nacional e internacional y luego, desde 2011, el desarrollo del módulo español (Argentina) para el software NooJ, creado por Max Silberztein de la Universidad de Franche Comté. Dicha tarea fue posible gracias a la implantación en máquina de la información lingüística del español declarada en los archivos Smorph, que fue adaptada según los requerimientos de la nueva herramienta por Celina Beltrán, Rodolfo Bonino, Bárbara Méndez, Andrea Rodrigo y Carolina Tramallino, bajo la dirección de Zulema Solana y la supervisión de Max silberztein.

Además, dirigió los siguientes proyectos de investigación:

- “Incidencia de la argumentación en la comprensión de textos académicos” evaluado por Comisión evaluadora externa SECYT (2001-2003).
- “La producción escrita de los ingresantes a la Universidad” evaluado por Comisión evaluadora externa SECYT (2004-2006).
- “Análisis automático de la interlengua de aprendientes de español” (PID 2009-2011).

- “Análisis automático de la gramática temprana” (PID 2012-2015).
- “La lectura de textos electrónicos utilizados en los procesos de aprendizajes disciplinares en la escuela media” (PID 2013-2016).

A continuación, se detallan las tesis doctorales que dirigió:

- “El orden de palabras en la adquisición del alemán como segunda lengua” de Horacio Dotti, Doctorado de la Facultad de Humanidades y Artes, UNR. 2008.
- “Recuperación de información en bases de datos de texto” de Claudia Susana Deco, Doctorado de la Facultad de Ingeniería y Ciencias Exactas de la UNR. 2009.
- “Análisis automático de las construcciones causativas” de Rodolfo Bonino, Doctorado de la Facultad de Humanidades y Artes, UNR. 2008.
- “El sintagma adverbial núcleo” de Andrea Rodrigo, Doctorado de la Facultad de Humanidades y Artes, UNR. 2009.
- “Análisis automático de la puntuación” de Walter Koza (beca CONICET), Doctorado de la Facultad de Humanidades y Artes, UNR. 2011.
- “Las inferencias en la comprensión de la lectura” de Selva Barros, Doctorado de la Facultad de Humanidades de la Universidad de Catamarca. 2011.
- “El sintagma adjetivo núcleo” de Bárbara Méndez (beca CONICET), Doctorado de la Facultad de Humanidades y Artes, UNR. 2013.
- “El sintagma nominal núcleo: una aproximación al análisis automático de textos” de Andrea Rodrigo Doctorado de la Facultad de Humanidades y Artes, UNR. 2011
- “Análisis morfológico automático de la interlengua de aprendientes de español como L2”, de Carolina Tramallino (beca CONICET), Doctorado de la Facultad de Humanidades y Artes, UNR. 2016.

En cuanto a tesis de "Maestría de Teoría Lingüística y Adquisición del lenguaje" de la Facultad de Humanidades y Artes, dirigió:

- “La no reversibilidad de la adquisición de la lectura y la escritura” de Rodolfo Hachén. 1998.
- “Adquisición de estructuras narrativas” de María del Rosario Fernández. 1998.

- "Las interrogativas en la gramática infantil" de la Prof. Susana Freidenberg. 2000.
- "Información sintáctica en los diccionarios inglés español" de María Andrea Fernández. 2001.
- "La variable sintáctica en la comprensión de la lectura" de Mirta Despósito. 2003.
- "El sintagma nominal núcleo: una aproximación al análisis automático de textos" de Andrea Rodrigo. 2006.

Escribió cuantiosos artículos, capítulos de libros y libros, estos últimos son:

- *La Gramática Modular* en colaboración con Nora Múgica. Hachette, Buenos Aires, 1990.
- *Gramática y Léxico Teoría lingüística y Teoría de Adquisición del Lenguaje* en colaboración con la Prof. Nora Múgica. EDICIAL, Buenos Aires, 1999.
- *Un estudio de la adquisición del lenguaje desde un punto de vista cognitivo*. Centro de Estudios de Adquisición del Lenguaje, Facultad de Humanidades y Artes, UNR, 1999.
- Módulo 1 para el *Curso a distancia Alfabet*, curso de formación para la alfabetización dictado en la Subsección de la Cátedra Unesco para la lectura y la Escritura en la Universidad Nacional de Rosario, 2000.
- *Morfología verbal del español. Estudio de Lingüística Computacional*. Centro de Estudios de Adquisición del Lenguaje, Facultad de Humanidades y Artes, UNR, 2006. ISBN 987-1166-17-6, 166 ps.
- *Lingüística Informática y enseñanza de lenguas*, Centro de Estudios de Adquisición del Lenguaje, Facultad de Humanidades y Artes, UNR, 2006.
- *La interlengua de los aprendientes del español como L2. Aportes de la lingüística informática*. Facultad de Humanidades y Artes, UNR, 2009.

Ganadora del proyecto Latin, iniciativa latinoamericana para la escritura colaborativa de textos. Responsable del equipo que escribió el libro: *Estudios del lenguaje. Niveles de representación lingüística*. Grupo alfa, financiado por la Unión Europea. (Enero de 2014) [http://www.latinproject.org/books/Estudios\\_Lenguaje\\_CC\\_BY-SA\\_3.0.pdf](http://www.latinproject.org/books/Estudios_Lenguaje_CC_BY-SA_3.0.pdf)

Se desempeñó como evaluadora, fue Miembro de la COMISIÓN DE FILOLOGÍA, LINGÜÍSTICA Y LITERATURA de CONICET, de las comisiones categorizadoras de

incentivos de Buenos Aires, La Plata, Mar del Plata y Misiones. Evaluadora externa de proyectos de investigación de la Univ.Nacional del Litoral, de la Universidad Nacional de la Plata, de la Universidad de Buenos Aires y de la Universidad Nacional Sarmiento.

Miembro de Jurados en concursos para profesor titular y adjunto en las Universidades de La Plata, Lomas de Zamora, Misiones, Santa Fe y Rosario. Además de la Universidad de La República (Montevideo). Asimismo, jurado de numerosas tesis de Doctorado en las Universidades de Buenos Aires, Córdoba, Catamarca y Rosario.

**Editora: Dra. Carolina Paola Tramallino (Universidad Nacional de Rosario, CONICET)**

Carolina Paola Tramallino es Profesora de Letras, JTP de la cátedra Lingüística General I de la Facultad de Humanidades y Artes, perteneciente a la Universidad Nacional de Rosario, Argentina. Además es Doctora en Humanidades y Artes con mención en Lingüística y becaria Posdoctoral del Consejo Nacional de Investigaciones Científicas y Tecnológicas de Argentina (CONICET). Su campo de investigación es el de la lingüística computacional aplicada a la adquisición del español como segunda lengua, desempeñándose como directora de proyectos de investigación referidos a dicha temática. Asimismo, dicta seminarios de Posgrado sobre Lingüística computacional, Adquisición del Español y Escritura Académica.

**Comité de Referato**

Dra. Claudia Deco (Universidad Nacional de Rosario).

Dr. Giovanni Parodi (Universidad Católica de Valparaíso, Chile).

Phill M. Silvia Alma Rivero (Universidad Nacional de Rosario).

Dra. Patricia Silvana San Martín (CONICET, Universidad Nacional de Rosario).

Dra. María Luisa Silva (CIIPME, CONICET).

**Revisora**

Dra. Bárbara Méndez (Universidad Nacional de Rosario).

Bárbara Méndez es doctora en Humanidades y Artes con mención en lingüística. Se desempeña como JTP en la cátedra de lingüística General I correspondiente a la carrera de Letras de la Facultad de Humanidades y Artes, UNR. Dicta cursos de escritura académica en el nivel de Posgrado. Es directora del proyecto de Extensión Universitaria *Déjate morder por un libro* y además dirige el Programa formación radicado en la Facultad de Humanidades y Artes Enseñar y aprender con tecnologías. Es miembro representante del SIED ( Sistema Institucional de Educación a Distancia), de la UNR.

### **Claudia Deco**

Licenciada en Matemáticas por la Universidad Nacional de Rosario. Especialista en Gestión de Sistemas de Información por la Facultad Católica de Química e Ingeniería, Pontificia Universidad Católica Argentina. Además es Magister en Informática. Instituto de Computación. Facultad de Ingeniería. Universidad de la República. Uruguay.

Doctora en Ingeniería por la Facultad de Ciencias Exactas, Ingeniería y Agrimensura de la Universidad Nacional de Rosario y es Profesora Asociada e Investigadora de la misma Facultad. También es Profesora Titular en la Facultad de Química e Ingeniería del Rosario de la Universidad Católica Argentina.

### **Giovanni Parodi**

Es Profesor Titular en la Pontificia Universidad Católica de Valparaíso, Chile. Actualmente, ejerce como Editor en Jefe de la Revista Signos. Estudios de Lingüística y Coordinador Científico del Laboratorio Lenguaje & Cognición. Al mismo tiempo, es director para Chile de la Cátedra UNESCO en Lectura y Escritura. Sus principales áreas de interés científico son la lingüística de corpus, la psicolingüística del discurso y la multimodalidad. Su último libro publicado en 2020 es: *Comprensión y Discurso. Del Movimiento Ocular al procesamiento Cognitivo*. Desde el año 2008 es Miembro Correspondiente de la Academia Chilena de la Lengua. Su producción científica se encuentra en [www.giovanniparodi.cl](http://www.giovanniparodi.cl)

### **Silvia Alma Rivero**

Tiene un M.A. y un M.Phil. in Linguistics de la City University of New York, USA. Es Profesora Titular de Lingüística General I y de Lengua Española II de la Escuela de Letras, Fac. Humanidades, U.N.R. Es Directora del Centro de Estudios de Adquisición del Lenguaje en la misma universidad. Su área de investigación es la fonología del español y la adquisición y enseñanza de la gramática de la L1/L2. Trabaja en particular en la formación docente de nivel superior. Dicta seminarios, dirige y evalúa proyectos de investigación y tiene artículos y libros en su especialidad.

### **Patricia Silvana San Martín**

Doctora en Humanidades y Artes por la Universidad Nacional de Rosario (UNR), Argentina. Investigadora Principal del CONICET y Docente-Investigadora categoría I. Vicedirectora del Instituto Rosario de Investigaciones en Ciencias de la Educación (CONICET-UNR). Profesora Titular en la Escuela de Música de la Facultad de Humanidades y Artes (UNR). Coordinadora de la “Red de Cooperación Académica en Patrimonio Cultural Inmaterial de Latinoamérica y el Caribe”. Miembro del Comité Académico de la Especialización a Distancia en Producción de Textos Críticos y de Difusión Mediática de las Artes y, de la Maestría en Crítica y Difusión de las Artes, Universidad Nacional de las Artes, Buenos Aires.

### **María Luisa Silva**

Doctora en Psicología (Univ. Nac. de La Plata), Especialista en Planificación y Gestión de Proyectos I+D (CAEU-OEI), Especialista Principal en Procesos de Lectura y Escritura (UNESCO-FFyL, UBA), Lic. en Letras. Investigadora Adjunta (CIIPME-CONICET) en desarrollo del lenguaje infantil. Profesora de posgrado: Univ. del Salvador; Univ. Nacional de La Plata y de cursos de especialización sobre el desarrollo lingüístico infantil desde una perspectiva cognitivo-funcional. Ganadora del premio Ignacio Chaves Cuevas a la mejor investigación en Gramática de la Lengua española (2009). Participó en numerosos proyectos de investigación de índole nacional e internacional sobre el desarrollo cognitivo y lingüístico de niños de niños con desarrollo típico, atípico y de sectores vulnerabilizados.

## **CAPÍTULO I**

### **Lingüística computacional**



# **Comparación de dos técnicas de clasificación supervisada en la categorización de textos y evaluación en datos simulados: Árboles de clasificación y Regresión Logística**

**Celina Beltrán**

Facultad de Ciencias Agrarias, Universidad Nacional de Rosario, Argentina  
cbeltran2510@gmail.com

**Ivana Barbona**

Facultad de Ciencias Agrarias, Universidad Nacional de Rosario, Argentina

## **Abstract**

This work compares the performance of two classification methods: Classification Trees (CT) and Logistic Regression (LR). This comparison is made 1) on a text categorization application and 2) an evaluation on simulated data under different scenarios.

1) For both methods, the functionality and performance in the texts classification is evaluated, describing how it is possible to use them to categorize and eventually characterize the texts. In this case, the classification criterion is the genre to which the text belongs (Scientific / Non-Scientific). The characterization of the texts is based on the frequency distribution of the morpho-syntactic categories. The texts were classified taking into account simultaneously the measurements made on them. The performance of the techniques was measured with the misclassification rate (MCR) calculated on a sample of texts not included in the estimation of the model and construction of the tree. The classification tree presented a MCR lower than that of the logistic model, managing to classify scientific texts more precisely. For CT the MCR was 4% for scientific texts and 28% for non-scientific texts. For the LR model, the MCR was 14% for scientific texts and 26% for non-scientific texts.

2) In the simulation study, it was observed as a main result, that in conditions where the predictor variables are highly correlated with the response, although the CT showed a significantly lower error rate in the classification, both methodologies work satisfactorily. However, when the conditions for obtaining a satisfactory classification are unfavorable (predictors with little correlation with the response), the CTs achieve a percentage of correct classification that is significantly higher than LR, with the disadvantage of obtaining a tree with numerous terminal nodes using the information from practically all the explanatory

variables. In the unbalanced case, the majority class presents a higher correct classification percentage in logistic regression at the cost of worse performance in the minority class. This behavior is more marked in this technique than in the classification trees.

**Keywords:** Supervised classification techniques; logistic regression; classification trees; simulation; classification of texts

## Resumen

En este trabajo se compara el desempeño de dos métodos de clasificación: Árboles de Clasificación (AC) y Regresión Logística (RL). Dicha comparación se realiza 1) sobre una aplicación en categorización de textos y 2) una evaluación sobre datos simulados bajo distintos escenarios.

1) Para ambos métodos se evalúa la funcionalidad y desempeño en la clasificación de textos describiendo cómo es posible utilizarlos para categorizar y eventualmente caracterizar los textos. En este caso, el criterio de clasificación es el género al que pertenece el texto (Científico / No Científico). La caracterización de los textos está basada en la distribución de frecuencias de las categorías morfo-sintácticas. Los textos se clasificaron teniendo en cuenta simultáneamente las mediciones realizadas sobre ellos. El desempeño de las técnicas fue medido con la tasa de mala clasificación (TMC) calculada sobre una muestra de textos no incluidos en la estimación del modelo y construcción del árbol. El árbol de clasificación presentó una TMC inferior a la del modelo logístico logrando clasificar con mayor precisión los textos científicos. Para el AC la TMC resultó 4% para los textos científicos y 28% para los textos no científicos. Para el modelo de RL la TMC resultó 14% para los textos científicos y 26% para los textos no científicos.

2) En el estudio por simulación, se observó como resultado principal, que en condiciones donde las variables predictoras están altamente correlacionadas con la respuesta, si bien los AC mostraron un porcentaje de error significativamente menor en la clasificación, ambas metodologías funcionan satisfactoriamente. Sin embargo, cuando las condiciones para obtener una clasificación satisfactoria son desfavorables (predictores poco correlacionados con la respuesta) los AC logran un porcentaje de clasificación correcta notablemente superior a la RL, con la desventaja de obtener un árbol con numerosos nodos terminales utilizando la

información de prácticamente todas las variables explicativas. En el caso desbalanceado, la clase mayoritaria presenta un porcentaje de clasificación correcta superior en la regresión logística a costa de un peor desempeño en la clase minoritaria. Este comportamiento es más marcado en RL que en los AC.

**Palabras clave:** Métodos de clasificación supervisada; regresión logística; árboles de clasificación; simulación; clasificación de textos

## 1. INTRODUCCION

El Análisis Multivariado se refiere al tipo de análisis que se realiza sobre  $n$  unidades experimentales sobre las cuales se han medido  $p$  variables y se pretende estudiar a todas las variables (o un gran número) en forma simultánea (Hair, J.F. 1999 [1]). Estas variables pueden ser cuantitativas, continuas o discretas, o cualitativas, nominales u ordinales (Pérez López, C. 2004 [2]). Uno de los objetivos de dichas técnicas es la clasificación de unidades u objetos en grupos (Cuadras, 2014 [3]). En la clasificación supervisada, tarea que concierne a este trabajo, se cuenta con un conocimiento a priori, es decir para la tarea de clasificar un objeto dentro de una categoría o clase se cuenta con la información de  $p$  variables observadas en un conjunto de objetos cuya categoría o clase de pertenencia se conoce. Las técnicas de clasificación pueden diferenciarse en aquellos métodos clásicos estadísticos y los que provienen de la Minería de datos. En las técnicas clásicas se estima un modelo estadístico cuyos coeficientes permitirán caracterizar los grupos y construir la regla de clasificación para nuevas unidades. Las inferencias sobre las estimaciones realizadas permiten detectar aquellas características que aportan en el proceso de clasificación. Esto marca una diferencia con las provenientes de la Minería de datos ya que en estos casos generalmente los análisis son de tipo exploratorios y no se realiza una generalización sobre poblaciones de las cuales se extraen los datos (inferencia).

Otra cuestión a tener en cuenta en esta tarea es la existencia de un desbalanceo de los grupos definidos por la variable respuesta binaria, es decir que existe una clase minoritaria y una mayoritaria, se presenta una dificultad en la clasificación de nuevas unidades. Esta dificultad o inconveniente se refleja en un deterioro del porcentaje de clasificación correcta en los grupos minoritarios, ya que en los grupos de mayor cantidad de observaciones las técnicas seguirán mostrando un buen desempeño. Esta situación es más problemática cuando justamente la clase o grupo de interés es el de menor tamaño.

Entre las técnicas de clasificación, correspondiente al enfoque clásico estadístico y el de minería de datos respectivamente, se pueden citar: Regresión Logística (RL) y Árboles de clasificación (AC) (Beltrán, 2012 [4]).

En este trabajo se propone el estudio de estas dos técnicas estadísticas multivariadas siendo de interés evaluar el desempeño de las mismas. Este desempeño será considerado en una primera

instancia sobre datos reales: clasificación de textos. Para los dos métodos en consideración se presentan su funcionalidad y desempeño describiendo cómo es posible utilizarlos para clasificar y eventualmente caracterizar textos de distintos géneros o disciplinas. El criterio de clasificación es el género al que pertenece el texto (Científico / No Científico). La caracterización de los mismos está basada en la distribución de frecuencias de las categorías morfo-sintácticas. En una segunda instancia de evaluación de las dos técnicas consideradas será sobre datos simulados bajo distintas situaciones que difieren en la estructura de correlaciones entre las variables intervinientes y en el desbalanceo o no de los grupos. En ambas aplicaciones, se considera como medida para la comparación entre métodos la tasa de mala clasificación (TMC), o bien la tasa de clasificación correcta (TCC), calculada sobre una muestra de textos no incluidos en el proceso de construcción de la regla de clasificación.

## **2. MATERIAL Y METODOS**

### **2.1. Datos reales**

#### *2.1.1. Diseño muestral*

El marco muestral para la selección de la muestra de los textos científicos está compuesto por textos académicos, resúmenes de trabajos presentados a congresos y revistas científicas, extraídos de internet pertenecientes a distintas disciplinas. Los textos periodísticos fueron seleccionados de un corpus mayor utilizado por el equipo de investigación INFOSUR. Este corpus se construyó con noticias extraídas de las páginas web de periódicos argentinos (noticias de tipo general, no especializadas en español). La unidad de muestreo fue el texto y la selección de la muestra se llevó a cabo empleando un diseño muestral estratificado.

Luego de obtener las muestras de los estratos, fueron evaluadas y comparadas respecto al número medio de palabras por texto. Se requiere esta evaluación para evitar que la comparación entre los géneros se vea afectada por el tamaño de los textos.

La muestra final para este trabajo quedó conformada por 90 textos científicos y 60 No científicos, con 162 y 135 palabras promedio por texto respectivamente.

### *2.1.2. Diseño y desarrollo de la base de datos*

En Beltrán (2009) se presenta la implementación de un etiquetador morfológico de textos basado en la información resultante de los análisis bajo el software Smorph (analizador y generador morfosintáctico desarrollado por Salah Aït-Mokhtar en el Groupe de Recherche dans les Industries de la Langue, Universidad Blaise-Pascal, Clermont II) y el módulo post-smorph que permite resolver ambigüedades que resultan del análisis de Smorph. Esta herramienta se utiliza para el análisis morfológico de los textos. La información resultante de dicho análisis se dispuso en una matriz de dimensión: tantas filas como cantidad de objetos lingüísticos tenga el texto y tantas columnas como ocurrencia+lema+valores. Luego, a partir de esta base de datos por palabra, se confeccionó la base de datos por documento que es analizada estadísticamente.

## **2.2. Datos Simulados**

Se generaron mediante simulación 500 archivos de datos de 150 filas (unidades) y 6 columnas (variables) bajo distintas condiciones o escenarios. La simulación se realizó a partir de distribuciones normales estandarizadas multivariadas con matriz de correlaciones según cuatro estructuras diferentes. Se consideró la primer columna (X1) como la variable respuesta y las restantes variables (X2 a X6) como las variables predictoras o explicativas. Luego de la generación de los ficheros se transformó la variable respuesta (X1) para obtener una variable dicotómica utilizando:

- a La mediana de la distribución teórica (dos grupos balanceados)
- b el cuartil 3 de la distribución teórica (un grupo con el 25% de las observaciones y el otro con el 75% restante)

La variable respuesta siempre se la consideró transformada a categórica ya que el objetivo de este estudio es evaluar las técnicas encargadas de clasificación de unidades. En este estudio, para evaluar el desempeño de las técnicas de clasificación en situaciones de tener dos grupos, se definieron las siguientes modalidades o condiciones sobre las cuales se evalúan los desempeños de las técnicas.

- 1- Escenario 1: Variable respuesta altamente correlacionada con las predictoras ( $0.27 < r < 0.63$ ) y las variables predictoras poco correlacionadas entre sí ( $r < 0.06$ ).

- 2- Escenario 2: Variable respuesta poco correlacionada con las predictoras ( $r < 0.06$ ) y las variables predictoras muy correlacionadas entre sí ( $0.49 < r < 0.84$ ).
- 3- Escenario 3: Variable respuesta muy correlacionada con las predictoras ( $0.36 < r < 0.83$ ) y las variables predictoras también muy correlacionadas entre sí ( $0.43 < r < 0.87$ ).
- 4- Escenario 4: Variable respuesta poco correlacionada con las predictoras y asimismo las variables predictoras poco correlacionadas entre sí ( $r < 0.06$ ).

De esta manera se definen 8 bases de datos, correspondientes a cada una de estas situaciones, que presentan una variable respuesta dicotómica (para el caso a) y b) correspondiente a grupos balanceados y desbalanceados respectivamente) y 5 variables explicativas continuas. Resultan entonces, de la combinación de los escenarios 1, 2, 3 y 4 con las modalidades de la respuesta (a) y (b), las siguientes bases (BASE n° de escenario-modalidad de X1): BASE 1a - BASE 1b – BASE 2a – BASE 2b- BASE 3a – BASE 3b – BASE 4a- BASE 4b

Sobre las bases simuladas detalladas recientemente se comparan las técnicas multivariadas de clasificación anunciadas: ÁRBOLES DE CLASIFICACIÓN Y REGRESIÓN LOGÍSTICA. Por este motivo, para cada muestra, se simularon datos extras o suplementarios para ser considerados en la evaluación de la clasificación sin haberlos utilizados en los procesos de estimación (grupo de prueba).

El proceso de simulación de ficheros de datos como las aplicaciones estadísticas subsiguientes se lleva a cabo en el software R versión 3.4.0 [6] (James, 2013 [7]).

### **2.3. Metodología estadística**

Uno de los problemas de los que se ocupan las técnicas estadísticas multivariadas es la clasificación de objetos o unidades en grupos o poblaciones. Dos enfoques agrupan a las técnicas de clasificación. Uno de ellos es cuando se conocen los grupos o categorías y se pretende ubicar los individuos dentro de estas categorías a partir de los valores de ciertas variables (técnicas supervisadas). El segundo enfoque, que no es el utilizado en este trabajo, ocurre cuando no se conocen los grupos de antemano y se pretende establecerlos a partir de los datos observados (técnicas no supervisadas).

Las técnicas evaluadas en este trabajo tienen por objetivo construir un sistema que permita clasificar unidades en una de las categorías definidas y conocidas previamente en función de las variables relevadas.

Luego de la aplicación de cada una de las técnicas definidas en este apartado se debe evaluar la calidad de los resultados, es decir el desempeño para clasificar mediante la validación del mismo. Esto se realiza particionando el conjunto de unidades en dos grupos. Uno es utilizado para la estimación del mismo (entrenamiento del sistema) y el segundo conforma el grupo de validación para la fase de prueba. Se consideró como medida para la comparación entre métodos el error de mala clasificación calculada sobre una muestra de textos no incluidos en el proceso de construcción de la regla de clasificación.

A continuación se presentan dos de las técnicas multivariadas que tienen por objetivo clasificar unidades en categorías definidas a priori que fueron evaluadas en diferentes aplicaciones por los autores.

### *2.3.1. Análisis de regresión logística*

Esta técnica es un caso particular de los modelos lineales generalizados, modela la probabilidad de que una unidad experimental pertenezca a un grupo en particular considerando información medida o registrada en dicha unidad.

La regresión logística es utilizada en situaciones en las cuales el objetivo es describir la relación entre una variable respuesta categórica, en este caso dicotómica, y un conjunto de variables explicativas que pueden ser tanto categóricas como cuantitativas.

Sea  $\mathbf{x}$  un vector de  $p$  variables independientes, esto es,  $\mathbf{x}' = (x_1, x_2, \dots, x_p)$ . La probabilidad condicional de que la variable  $y$  tome el valor 1 (presencia de la característica estudiada), dado valores de las  $p$  covariables que definen el vector  $\mathbf{x}$  es:

$$P(y=1/X) = \pi(x) = \frac{e^{g(x)}}{1 + e^{g(x)}}$$

$$\text{donde } g(x) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$$

$\beta_0$  es la constante del modelo o término independiente

$p$  el número de covariables

$\beta_i$  los coeficientes de las covariables

$x_i$  las covariables que forman parte del modelo.



Si alguna de las variables independientes es una variable discreta con k niveles se debe incluir en el modelo como un conjunto de k-1 “variables de diseño” o “variables dummy”. El cociente de las probabilidades correspondientes a los dos niveles de la variable respuesta se denomina odds y su expresión es:

$$\frac{P(y=1|X)}{1-P(y=1|X)} = e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}$$

Si se aplica el logaritmo natural, se obtiene el modelo de regresión logística:

$$\log\left(\frac{P(y=1|X)}{1-P(y=1|X)}\right) = \log(e^{\beta_0 + \beta_1 x_1 + \dots + \beta_p x_p}) = \beta_0 + \beta_1 x_1 + \dots + \beta_p x_p$$

En la expresión anterior el primer término se denomina logit, esto es, el logaritmo de la razón de proporciones de los niveles de la variable respuesta.

Los coeficientes del modelo se estiman por el método de Máxima Verosimilitud y la significación estadística de cada uno de los coeficientes de regresión en el modelo se puede verificar utilizando, entre otros, el test de Wald y el test de razón de verosimilitudes.

Una cuestión importante en este tipo de análisis es determinar si todas las variables consideradas en la función de discriminante contienen información útil y si solamente algunas de ellas son suficientes para diferenciar los grupos. Dado que las variables utilizadas para explicar la respuesta es probable que estén correlacionadas, es posible también que compartan información. Por lo tanto, se puede buscar un subgrupo de variables mediante algún criterio de modo tal que las variables excluidas no contengan ninguna información adicional.

En este trabajo se utilizó como evaluación del ajuste del modelo el test de Hosmer-Lemeshow y, dado que el modelo es utilizado para clasificar unidades, se utilizó también la tasa de mala clasificación calculada sobre la muestra independiente excluida en la etapa de estimación.

### 2.3.2. Árboles de clasificación

Los árboles de clasificación son una técnica de análisis discriminante no paramétrica que permite predecir la asignación de unidades u objetos a grupos predefinidos en función de un conjunto de variables predictoras. Esto es, dada una variable respuesta categórica, los árboles crean una serie de reglas basadas en las variables predictoras que permiten asignar una nueva observación a una de las categorías o grupo.

Es un algoritmo que genera un árbol de decisión en el cual las ramas representan las decisiones y cada una de ellas genera reglas sucesivas que permiten continuar la clasificación. Estas particiones recursivas logran formar grupos homogéneos respecto a la variable respuesta. El árbol determinado puede ser utilizado para clasificar nuevas unidades.

Es un método no-paramétrico de segmentación binaria donde el árbol es construido dividiendo repetidamente los datos. En cada división los datos son partidos en dos grupos mutuamente excluyentes. Comienza el algoritmo con un nodo inicial o raíz a partir del cual se divide en dos sub-grupos o sub-nodos, esto es, se elige una variable y se determina el punto de corte de modo que las unidades pertenecientes a cada nuevo grupo definido sean lo más homogéneas posible. Luego se aplica el mismo procedimiento de partición a cada sub-nodo por separado. En cada uno de estos nodos se vuelve a repetir el proceso de seleccionar una variable y un punto de corte para dividir la muestra en dos partes homogéneas.

Las divisiones sucesivas se determinan de modo que la heterogeneidad o impureza de los sub-nodos sea menor que la del nodo de la etapa anterior a partir del cual se forman. El objetivo es particionar la respuesta en grupos homogéneos manteniendo el árbol lo más pequeño posible.

La función de impureza es una medida que permite determinar la calidad de un nodo, esta se denota por  $i(t)$ . Si bien existen varias medidas de impureza utilizadas como criterios de partición, una de las más utilizadas está relacionada con el concepto de entropía

$$i(t) = \sum_{j=1}^K p(j/t) \cdot \ln p(j/t)$$

donde  $j = 1, \dots, k$  es el número de clases de la variable respuesta categórica y  $p(j/t)$  la probabilidad de clasificación correcta para la clase  $j$  en el nodo  $t$ .

El objetivo es buscar la partición que maximiza

$$\Delta i(t) = - \sum_{j=1}^K p(j/t) \cdot \ln p(j/t)$$

De todos los árboles que se pueden definir es necesario elegir el óptimo. El árbol óptimo será aquel árbol de menor complejidad cuya tasa de mala clasificación es mínima. Generalmente esta búsqueda se realiza comparando árboles anidados mediante validación cruzada. La validación cruzada consiste, en líneas generales, en sacar de la muestra de aprendizaje o entrenamiento una muestra de prueba, con los datos de la muestra de aprendizaje se calculan

los estimadores y el subconjunto excluido es usado para verificar el desempeño de los estimadores obtenidos utilizándolos como “datos nuevos”.

### 3. RESULTADOS

#### 3.1. Desempeño en datos reales: Clasificación de textos

En Beltrán 2013 [10] se realizó un análisis exploratorio en el cual se evidencian las características que discriminan los corpus de textos en estudio. En dicho estudio se evidenció que existen diferencias significativas entre los corpus respecto al tamaño de los textos (número de palabras por texto). Esta situación llevó a realizar las sucesivas comparaciones sobre los porcentajes o proporciones de las categorías gramaticales, hallando diferencias significativas ( $p < 0.05$ ) para todas las categorías gramáticas excepto la proporción de clíticos y de verbos en los documentos analizados. Asimismo, en un análisis de componentes principales, se dispusieron los textos en el plano de proyección demostrando que los textos procedentes del corpus No Científico presentan un mayor número de adverbios, respecto a las restantes categorías, que los textos Científicos.

##### 3.1.1. Análisis de Regresión Logística

Se realizó un análisis de regresión logística para obtener una regla de clasificación que permita asignar los textos en estas dos poblaciones, definidas por el género al que pertenecen (Científico / No científico), en base a la frecuencia de cada categoría gramatical en el texto.

Para determinar cuáles categorías gramaticales son las responsables de la discriminación se utilizaron los tres algoritmos de selección de variables. En todos los casos se llega al mismo resultado: las variables que logran diferenciar a los dos corpus de textos analizados son el número de adjetivos, adverbios, conjunciones copulativas, determinantes, nombres y preposiciones.

Tabla 1: Coeficientes del modelo de regresión logística

<b>Estimación máximo verosímil</b>					
<b>Coeficiente</b>	<b>gl</b>	<b>Estimador</b>	<b>Error estándar</b>	<b>Est. Chi-cuadrado</b>	<b>Prob. asociada</b>
<b>Intercepto</b>	1	-0.6868	0.5507	1.5554	0.2123
<b>adjetivos</b>	1	0.1694	0.0562	9.0777	0.0026
<b>adverbios</b>	1	-0.3106	0.0800	15.0862	0.0001

<b>Conj. Cop.</b>	1	0.2769	0.1073	6.6566	0.0099
<b>determinantes</b>	1	0.1216	0.0464	6.8795	0.0087
<b>nombres</b>	1	-0.1995	0.0464	18.5044	<.0001
<b>preposiciones</b>	1	0.1575	0.0544	8.3925	0.0038

Este modelo permite, mediante la utilización de los coeficientes estimados, calcular para cada texto la probabilidad de pertenecer a cada uno de los corpus definidos por el género.

$$P(\in \text{Cien}/X) = \pi(x) = \frac{e^{\beta_0 + \beta_1 x_1 + \dots + \beta_j x_j + \dots + \beta_p x_p}}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_j x_j + \dots + \beta_p x_p}}$$

$$P(\in \text{No Cien}/X) = \pi(x) = \frac{1}{1 + e^{\beta_0 + \beta_1 x_1 + \dots + \beta_j x_j + \dots + \beta_p x_p}}$$

Con este criterio un texto es asignado al corpus cuya probabilidad es máxima.

La bondad del ajuste se evaluó mediante el test de Hosmer-Lemeshow y la tasa de error de clasificación. Con el modelo de regresión logística obtenido durante la selección de variables se obtuvo una tasa de error global, estimada sobre el corpus de prueba, del 20% y la probabilidad asociada en el test de bondad de ajuste es  $p=0.9696$  evidenciando lo adecuado del modelo. La tabla 2 presenta la tasa de error para cada género.

Tabla 2: Tasa de mala clasificación

<b>Medidas de evaluación</b>		
	<b>CIENTIFICO</b>	<b>NO CIENTIFICO</b>
<b>Tasa de error</b>	14%	26%

Tabla 3: Razones de odds estimadas

<b>Razón de odds</b>			
<b>Efecto</b>	<b>Estimación puntual</b>	<b>IC 95%</b>	
<b>adjetivos</b>	1.185	1.061	1.323
<b>adverbios</b>	0.733	0.627	0.857
<b>Conj. Cop.</b>	1.319	1.069	1.628
<b>determinantes</b>	1.129	1.031	1.237
<b>nombres</b>	0.819	0.748	0.897
<b>preposiciones</b>	1.171	1.052	1.302

Los coeficientes del modelo de regresión logística permiten la interpretación de la misma. La razón de odds para el número de adjetivos es 1.19 lo cual indica que la chance de clasificar a

un texto como Científico se incrementa en un 19% al aumentar en número de adjetivos en una unidad. Con respecto al número de adverbios la razón de odds es menor a la unidad por lo tanto si se interpreta el recíproco,  $1/0.73=1.36$ , significa que la chance de clasificar un texto en el corpus No Científico aumenta un 36% al incrementarse en una unidad el número de adverbios. Si analizamos el efecto de las conjunciones copulativas, determinantes y preposiciones, al incrementar en una unidad cada una de estas categorías gramaticales, la chance de clasificar un texto como Científico se incrementa en un 32%, 13% y 17% respectivamente. Al igual que el efecto del número de adverbios, la probabilidad de clasificar un texto como No Científico se incrementa en un 22% ( $1/0.82=0.22$ ) al aumentar en una unidad la cantidad de nombres en el texto.

3.1.2. Árboles de Clasificación

Se aplicó la técnica de Árboles de Clasificación para obtener reglas de clasificación que permitan asignar los textos en dos poblaciones, definidas por el género al que pertenecen: CIENTÍFICO y NO CIENTÍFICO. De la misma manera que en el apartado previo, los predictores utilizados corresponden a la distribución de las distintas categorías morfológicas halladas en el análisis automático de los textos (proporción de cada categoría morfológica).

Las variables que mostraron una buena discriminación de los grupos son la proporción de adverbios, nombres, adjetivos, preposición y verbos. El árbol final presenta 10 nodos terminales.

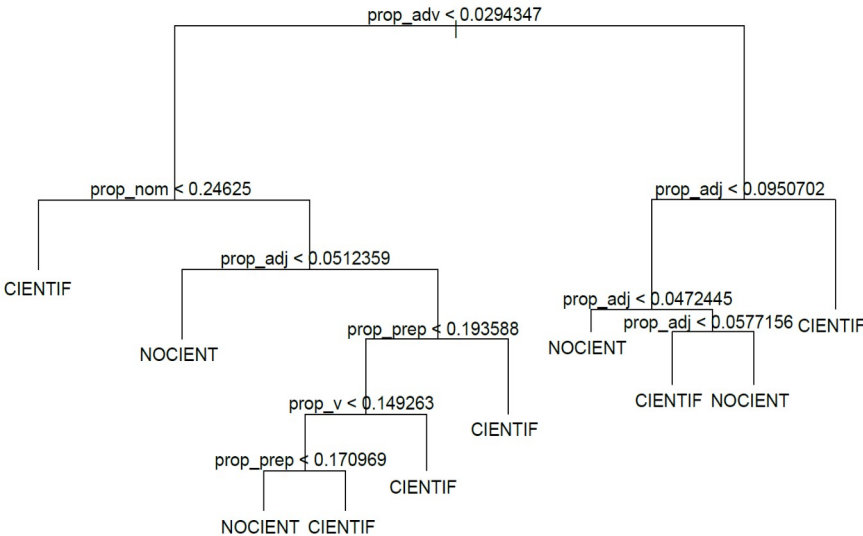


Gráfico 1: Árbol de clasificación

El gráfico 1 muestra el árbol resultante de esta aplicación. La variable regresora más fuertemente asociada con el género es la proporción de adverbios en el texto, categorizada como superiores e inferiores a 0.029 (2.9%). El corpus general queda así dividido en dos grupos, los que presentan más del 2.9% de adverbios y los que no. Luego intervienen en las sucesivas subdivisiones el número de nombres, adjetivos, verbos y preposiciones. Interpretando el árbol resultante, se encuentran 10 perfiles de textos (que corresponden a los 10 nodos terminales) asociados con una de los dos géneros. Estos son:

- Textos con un porcentaje de adverbios inferior al 2.9% y un porcentaje de nombres menor a 25% son clasificados como textos CIENTÍFICOS.
- Textos con un porcentaje de adverbios inferior al 2.9%, un porcentaje de nombres mayor a 25% y de adjetivos inferior al 5% son clasificados como textos NO CIENTÍFICOS.
- Textos con un porcentaje de adverbios inferior al 2.9%, un porcentaje de nombres mayor a 25%, de adjetivos superior al 5%, de preposición menor al 17% y de verbos menor al 15% son clasificados como textos NO CIENTÍFICOS.
- Textos con un porcentaje de adverbios inferior al 2.9%, un porcentaje de nombres mayor a 25%, de adjetivos superior al 5%, de preposición entre 17% y 19%, y de verbos menor al 15% son clasificados como textos CIENTÍFICOS.
- Textos con un porcentaje de adverbios inferior al 2.9%, un porcentaje de nombres mayor a 25%, de adjetivos superior al 5%, de preposición inferior al 19%, y de verbos mayor al 15% son clasificados como textos CIENTÍFICOS.
- Textos con un porcentaje de adverbios inferior al 2.9%, un porcentaje de nombres mayor a 25%, de adjetivos superior al 5%, de preposición mayor al 19% son clasificados como textos CIENTÍFICOS.
- Textos con un porcentaje de adverbios superior al 2.9% y un porcentaje de adjetivos inferior al 4.7% son clasificados como textos NO CIENTÍFICOS.
- Textos con un porcentaje de adverbios superior al 2.9% y un porcentaje de adjetivos entre 4.7% y 5.7% son clasificados como textos CIENTÍFICOS.

- Textos con un porcentaje de adverbios superior al 2.9% y un porcentaje de adjetivos entre 5.7% y 9.5% son clasificados como textos NO CIENTÍFICOS.
- Textos con un porcentaje de adverbios superior al 2.9% y un porcentaje de adjetivos superior al 9.5% son clasificados como textos CIENTÍFICOS.

El árbol final fue evaluado utilizando la muestra de prueba, no fue utilizada en la construcción del mismo, hallando una tasa de mala clasificación del 14%, siendo 4% para los textos científicos y 28% para los no científicos (Tabla 4).

Tabla 4: Tasa de mala clasificación

<b>Medidas de evaluación</b>		
	<b>CIENTIFICO</b>	<b>NO CIENTIFICO</b>
<b>Tasa de error</b>	4%	28%

### **3.2. Desempeño en datos simulados**

#### *3.2.1. Análisis exploratorio de datos simulados*

Para explorar las bases de datos simuladas se aplica el test no paramétrico de Wilcoxon para muestras independientes, en cada conjunto de datos simulado, para llevar a cabo la comparación de los grupos definidos por la variable respuesta binaria respecto a cada variable explicativa. Esto se realiza para cada una de las muestras contenidas en cada escenario. Los resultados de este análisis sugieren que las técnicas de clasificación deberían mostrar sin inconvenientes un buen desempeño en las muestras correspondientes a las bases de los escenarios 1 y 3 ya que los grupos evidencian diferencias marcadas respecto a las variables utilizadas para la discriminación. No se esperaría lo mismo en datos provenientes de los escenarios 2 y 4 en los que los grupos parecen no mostrar discrepancias. Por este motivo en estos casos es interesante evaluar cómo se diferencian los resultados obtenidos en la clasificación con RL y AC, evaluando también el deterioro en la clasificación en el grupo minoritario en el caso desbalanceado.

### *3.2.2. Aplicación de técnicas de clasificación*

#### *3.2.2.1 Regresión Logística*

Se ajustó un modelo de regresión logística para variable respuesta dicotómica y 5 variables explicativas continuas, para cada una de las 500 muestras en cada uno de los 4 escenarios evaluados y considerando los dos casos de variable respuesta: grupos aproximadamente balanceados (a) y no (b). Los resultados hallados en cada caso se presentan a continuación.

Escenario 1a: En promedio, en las 500 muestras se observó en promedio un 16% de mala clasificación, siendo el mínimo observado de 7% y el máximo de 25%.

Escenario 1b: En promedio, en las 500 muestras se observó en promedio un 11% de mala clasificación, siendo el mínimo observado de 3% y el máximo de 19%.

Escenario 2a: En promedio, en las 500 muestras se observó en promedio un 42% de mala clasificación, siendo el mínimo observado de 33% y el máximo de 51%.

Escenario 2b: En promedio, en las 500 muestras se observó en promedio un 25% de mala clasificación, siendo el mínimo observado de 15% y el máximo de 38%.

Escenario 3a: En promedio, en las 500 muestras se observó en promedio un 14% de mala clasificación, siendo el mínimo observado de 4% y el máximo de 23%.

Escenario 3b: En promedio, en las 500 muestras se observó en promedio un 10% de mala clasificación, siendo el mínimo observado de 3% y el máximo de 19%.

Escenario 4a: En promedio, en las 500 muestras se observó en promedio un 42% de mala clasificación, siendo el mínimo observado de 30% y el máximo de 54%.

Escenario 4b: En promedio, en las 500 muestras se observó en promedio un 25% de mala clasificación, siendo el mínimo observado de 15% y el máximo de 35%.

#### *3.2.2.2 Árboles de Clasificación*

Se aplicó la técnica de AC (Árboles de Clasificación) para variable respuesta dicotómica y 5 variables explicativas continuas, para cada una de las 500 muestras en cada uno de los 4 escenarios evaluados y considerando los dos casos de variable respuesta: grupos aproximadamente balanceados (a) y no (b). Los resultados se presentan a continuación para cada caso.



Escenario 1a: En promedio, en las 500 muestras se observó en promedio un 12% de mala clasificación, siendo el mínimo observado de 1% y el máximo de 21%.

Escenario 1b: En promedio, en las 500 muestras se observó en promedio un 10% de mala clasificación, siendo el mínimo observado de 5% y el máximo de 16%.

Escenario 2a: En promedio, en las 500 muestras se observó en promedio un 21% de mala clasificación, siendo el mínimo observado de 13% y el máximo de 29%.

Escenario 2b: En promedio, en las 500 muestras se observó en promedio un 17% de mala clasificación, siendo el mínimo observado de 8% y el máximo de 25%.

Escenario 3a: En promedio, en las 500 muestras se observó en promedio un 12% de mala clasificación, siendo el mínimo observado de 6% y el máximo de 18%.

Escenario 3b: En promedio, en las 500 muestras se observó en promedio un 9% de mala clasificación, siendo el mínimo observado de 4% y el máximo de 15%.

Escenario 4a: En promedio, en las 500 muestras se observó en promedio un 21% de mala clasificación, siendo el mínimo observado de 14% y el máximo de 28%.

Escenario 4b: En promedio, en las 500 muestras se observó en promedio un 17% de mala clasificación, siendo el mínimo observado de 10% y el máximo de 24%.

### 3.2.2.3 Comparación de los resultados hallados en datos simulados

Los porcentajes de clasificación correcta globales son superiores en los árboles de clasificación, para los 4 escenarios. La diferencia en el desempeño de las técnicas es más evidente en los escenarios en los que la respuesta se encuentra poco correlacionada con las variables explicativas. En los casos contrarios la diferencia es leve. Es decir, incluso en el escenario menos deseable los árboles se desempeñan mejor en la tarea de clasificar nuevas unidades.

Tabla 5: Porcentaje promedio de clasificación correcta según escenario, conformación de grupos y técnica estadística.

Escenario	Caso balanceado		Caso desbalanceado	
	Árboles de clasificación	Regresión Logística	Árboles de clasificación	Regresión Logística
1	88.0	84.5	89.5	88.5
2	79.2	58.2	83.1	75.3
3	88.5	86.5	90.6	89.6
4	79.3	58.2	82.9	75.3

Tabla 6: Porcentaje promedio de clasificación correcta según escenario, técnica estadística y clase, en el caso desbalanceado.

Escenario	Caso desbalanceado			
	Árboles de clasificación		Regresión Logística	
	Clase minoritaria	Clase mayoritaria	Clase minoritaria	Clase mayoritaria
1	76.8	93.5	71.2	94.0
2	61.4	90.0	2.9	99.2
3	79.7	94.1	74.6	94.4
4	60.9	89.8	2.6	99.2

Cuando se analiza el caso de grupos desbalanceados, el porcentaje de clasificación correcta dentro de cada clase se muestra en la tabla 6. Como era de esperar, la clase mayoritaria presenta un porcentaje alto mientras que la clase minoritaria muestra un mal desempeño, siendo mayor la diferencia principalmente en los escenarios en los que la respuesta no se correlaciona con las variables explicativas del modelo. Comparando las técnicas, en este caso el modelo de regresión logística presenta una clasificación más acertada sólo en la clase mayoritaria, mientras que los árboles clasifican con un mejor desempeño en la clase minoritaria, respecto a la otra metodología.

#### 4. CONCLUSIONES

Respecto a la aplicación en datos reales (clasificación de textos), el desempeño de las técnicas fue medido con la TMC calculada sobre una muestra de textos no incluidos en la estimación del modelo y construcción del árbol. El árbol de clasificación presentó una TMC inferior a la del modelo logístico logrando clasificar con mayor precisión los textos científicos. Para el AC la TMC resultó 4% para los textos científicos y 28% para los textos no científicos. Para el modelo de RL la TMC resultó 14% para los textos científicos y 26% para los textos no científicos.

La diferencia en la tasa de mala clasificación sólo se diferenció en el corpus de textos científicos para el cual con el árbol se obtuvo un 4% de mala clasificación versus un 14% para el modelo de regresión logística.

En ambos tipos de análisis, las diferencias entre los dos tipos de textos están centradas principalmente en el porcentaje de adverbios, adjetivos, nombres y preposiciones presentes. Sin embargo, en el modelo de regresión logística han intervenido otras variables en la

discriminación como los determinantes y conjunciones copulativas; mientras que el árbol de clasificación utiliza el porcentaje de verbos, categoría morfológica no utilizada en la regresión.

Una ventaja observada en el árbol de clasificación es la adaptación para recoger el comportamiento no aditivo de las variables predictoras, de manera que las interacciones se incluyen de manera automática. Sin embargo, en esta técnica se pierde información al tratar a las variables predictoras continuas como variables dicotómicas.

Dado que en estos datos las categorías de la variable respuesta están desbalanceadas, la clase mayoritaria (textos científicos) presenta un porcentaje de mala clasificación inferior a costa de un peor desempeño en la clase minoritaria (textos no científicos).

En la segunda instancia de evaluación, se ha comparado el desempeño de estas dos técnicas en datos simulados bajo distintas condiciones que diferían en:

- la estructura de correlaciones entre la variable respuesta y las predictoras y entre las predictoras mismas.
- Conformación de la respuesta: grupos balanceados y desbalanceados.

Entre las similitudes y diferencias halladas se puede enunciar:

- En condiciones en que las variables predictoras están altamente correlacionadas con la respuesta, ambas metodologías funcionan satisfactoriamente. Sin embargo, la superioridad de los AC respecto a la TCC resultó significativa.
- Si bien en esta aplicación no se puede evidenciar, por no ser datos correspondientes a una problemática real, los modelos de RL presentan la ventaja de la interpretación de los coeficientes estimados que permiten reflejar información valiosa contenida en los datos.
- En condiciones desfavorables para obtener una clasificación satisfactoria, predictores poco correlacionados con la respuesta, los AC logran una TCC notablemente superior a la RL, con la desventaja de obtener un árbol con numerosos nodos terminales utilizando la información de prácticamente todas las variables explicativas.
- En el caso desbalanceado, la clase mayoritaria presenta una TCC superior en la regresión logística a costa de un peor desempeño en la clase minoritaria. Este comportamiento es más marcado en esta técnica que en los árboles de clasificación.

## Bibliografía

- [1] Hair, J. F., Anderson, R. L., Tatham, R. L. y Black, W. C. (1999). *Análisis Multivariante*. Prentice Hall Iberia, Madrid.
- [2] Pérez López, C. (2004). *Técnicas de Análisis Multivariante de Datos*. Pearson Educación, Madrid.
- [3] Cuadras, C. M. (2014). *Nuevos métodos de análisis multivariante*. CMC Editions, España.
- [5] Beltrán, C. (2012). Árboles de clasificación y su comparación con análisis de regresión logística aplicado a la clasificación de textos académicos. *INFOSUR*. 6.
- [6] Core Team, R. (2017). *A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna. disponible en URL <https://www.R-project.org/>.
- [7] James, G., Witten, D., Hastie, T. y Tibshirani, R. (2013). *An Introduction to Statistical Learning with Applications*. R. Springer, New York.
- [8] Agresti, A. (2002). *Categorical Data Analysis*. Wiley & Sons. New Jersey.
- [9] Hosmer, D. y Lemeshow, S. (1989). *Applied Logistic Regression*. Jhon Wiley & Sons. New York.
- [10] Beltrán C. (2013). Estudio exploratorio para la comparación de distintos tipos de textos: Textos Científicos y Textos No Científicos. *Revista de Epistemología y Ciencias Humanas*, 5.

Restricciones del artículo neutro en la presencia de sintagmas nominales

# Definición de categorías y formulación de propiedades en 5P<sup>1</sup>

**María José González Díaz**

Consejo de Formación en Educación  
Montevideo, Uruguay  
mgonzalez.uy@gmail.com

**Cristina Pippolo Griego**

Consejo de Formación en Educación  
Montevideo, Uruguay  
cristinapippolo@gmail.com

## Abstract

The objective of this proposal is to show, following the procedures of Computational Linguistics, the different behavior of two classes of articles in the conformation of the determined syntagmas. In these structures, the gender feature of the articles conditions the requirements and exclusions. These particular features can be formalized and implanted on a machine. The methodology followed for formalization includes the following stages: definition of the relevant categories for the analysis, according to the 5P paradigm; formulation of the properties of existence, requirement, exclusion and linearity; implementation in the XFST program (Xerox Finite-State Tool, version 8.1.3) and finally, interpretation of the results.

**Keywords:** formalization, Computational Linguistic, natural language.

## Resumen

El objetivo de la presente propuesta es mostrar, siguiendo los procedimientos de la Lingüística Computacional, el distinto comportamiento de dos clases de artículos en la conformación de los sintagmas determinados. En estas estructuras, el rasgo género de los artículos condiciona las exigencias y las exclusiones. Estos rasgos particulares pueden ser formalizados e implantados en máquina. La metodología seguida para la formalización incluye las siguientes etapas: definición de las categorías pertinentes para el análisis, según el

---

<sup>1</sup> Una versión previa de este trabajo, en la que también participó Ivanna Centanino, fue presentada en el seminario Modelización en Lingüística Computacional de la Universidad Nacional de Rosario en 2012.

paradigma 5P; formulación de las propiedades de existencia, de exigencia, de exclusión y de linealidad; implantación en el programa XFST (Xerox Finite-State Tool, versión 8.1.3) y finalmente, interpretación de los resultados.

**Palabras clave:** formalización, lingüística, computacional, lenguaje natural.

## 1. INTRODUCCIÓN

El presente artículo describe el diferente comportamiento de los artículos en la formación de los sintagmas determinados. Aunque el paradigma de los artículos determinantes en español incluye cinco formas: *el, la, lo, los y las*, el rasgo de género condiciona las exigencias y las exclusiones de las estructuras en las que se incorporan. Es consabido que las características del artículo neutro lo diferencian de las demás formas porque no aparece en las mismas construcciones. Estos rasgos particulares pueden ser modelizados en máquina. Para ello, en primer lugar, se definen las categorías con las cuales se trabajó siguiendo el paradigma 5P; en segundo lugar, se presentan las propiedades de existencia, de exigencia, de exclusión y de linealidad, que se establecieron entonces; en tercer lugar, se explica cómo se implementó en el programa XFST (Xerox Finite-State Tool, versión 8.1.3<sup>2</sup>), utilizado en ese momento; y finalmente, se interpretan los resultados.

## 2. DEFINICIÓN DE LAS CATEGORÍAS

A continuación, se definen las categorías que conforman el vocabulario con las que se formuló la hipótesis lingüística, en términos legibles para la computación, que el índice de determinación con artículo neutro es mucho menor que con el resto. Estas categorías son: artículo, nombre común y adjetivo.

P1: amod (saa, [art, nc, adj]).

### ARTÍCULO

La categoría se subdividió en dos subclases, usando como criterio un rasgo que es relevante porque determina posibilidades y restricciones para integrar una estructura. Se trata del rasgo *género*. La subclase artículo 1 (**art1**) se integró por el artículo masculino singular, el femenino singular, el masculino plural y el femenino plural. La subclase artículo 2 (**art2**), por el artículo neutro. La controversia sobre el lugar que ocupa el artículo dentro la clasificación categorial ha mostrado distintas posturas e incluso ha sido discutida la existencia de una forma neutra del artículo. Aunque han sido consideradas en este trabajo como subclases diferentes, no se puede desconocer que la distribución y propiedades gramaticales del neutro no son diferentes a las del resto de las formas del artículo. En este sentido, se puede decir que

2 Actualmente la creación de la plataforma NooJ, diseñada por Marx Silverstein, para el español ha sustituido la utilización del programa XFST usado en el momento de la instrumentación del Seminario.

el sintagma introducido por uno u otros podrá cumplir la misma función sintáctica en una oración.

#### NOMBRE COMÚN

La primera clasificación que plantea la tradición gramatical es la que distingue los nombres propios de los comunes (**nc**). En esta propuesta se descartan los primeros por el tipo de estructura que se somete a análisis. La categoría **nc**, en este trabajo, siempre funciona como núcleo de la estructura. El **nc** es una categoría variable que tiene rasgos de género y número. En la sintaxis, se reconoce como el núcleo del sintagma nominal y tiene la posibilidad de admitir modificadores que, cuando es posible, concuerdan con aquel en género y número.

#### ADJETIVOS

La subclasificación realizada en la categoría adjetivo corresponde a lo que tradicionalmente las gramáticas denominan adjetivos calificativos (**adj1**) y adjetivos relacionales (**adj2**). Estas subcategorías funcionan como núcleo, cuando se cumplen determinadas condiciones. Por ejemplo, un adjetivo es núcleo cuando no aparece un nombre en la estructura.

P2: obligdi (saa, [nc, adj]).

A continuación, se definirá la propiedad que establece cuáles de las tres categorías pueden aparecer por única vez.

P3: unic (saa, [art, nc, adj2]).

Esta formulación significa que el artículo, el nombre común y el adjetivo relacional pueden aparecer una sola vez en la estructura.

### 3. EXIGENCIAS DE LAS SUBCATEGORÍAS ART1 Y ART2

En este trabajo, las propiedades de exigencia representan la necesaria ocurrencia de la categoría **nc** y de las subcategorías **adj1** y **adj2** motivada por la presencia de las subcategorías **art1** o **art2**. La siguiente formulación responde al paradigma 5P:

P+1: exig (saa, [[art1], [[nc],[adj1],[adj2]]]).



Esta propiedad enuncia que el artículo 1 (**art1**) exige la presencia de un nombre común (**nc**), de un adjetivo calificativo (**adj1**) o de un adjetivo relacional (**adj2**).

P+2: exig (saa, [[art2], [[adj1],[adj2]]]).

Esta propiedad enuncia que el artículo 2 (**art2**) exige la presencia de un adjetivo calificativo (**adj1**) o un adjetivo relacional (**adj2**).

#### 4. EXCLUSIONES

Estas propiedades expresan las restricciones que limitan la ocurrencia de categorías, motivada por la presencia de la subcategoría **art2**, **nc** o **adj2**.

P~1: exclu (saa, [[nc], [° adj]]).

La propiedad de exclusión 1 expresa que la ocurrencia del nombre común excluye la presencia del adjetivo como núcleo de la estructura.

P~2: exclu (saa, [[art2], [nc]]).

La propiedad de exclusión 2 expresa que la presencia del artículo neutro excluye la presencia del nombre común.

P~3: exclu (saa, [[art2], [[adj1, adj1]]]).

La propiedad de exclusión 3 expresa que la presencia del artículo neutro excluye la coocurrencia de dos adjetivos calificativos.

P~4: exclu (saa, [[art2], [[adj1, adj2]]]).

La propiedad de exclusión 4 expresa que la presencia del artículo neutro excluye la ocurrencia simultánea del adjetivo calificativo y del adjetivo relacional

#### 5. PRECEDENCIA

La propiedad de linealidad o precedencia expresa el orden de aparición permitido de las categorías en la estructura.

P<1 <saa, { \_ }, precede (art, {nc, adj})>.

La propiedad de linealidad 1 significa que artículo precede al nombre común y al adjetivo.

P<2 <saa, { \_ }, precede (nc, {adj2})>.

La propiedad de linealidad 2 enuncia que el nombre común precede al adjetivo relacional.

P<3 <saa, {\_}, precede (adj1, {adj2}>.

La propiedad de linealidad 3 expresa que el adjetivo calificativo precede al adjetivo relacional.

## 6. IMPLANTACIÓN EN MÁQUINA

Luego de definir las categorías y formular las propiedades según el paradigma 5P, se procedió a la implantación en máquina. Se creó el archivo **catsaa.txt** con los siguientes datos categoriales:

```
define art [art1|art2];
define art1 [% art1];
define art2 [% art2];
define adj [adj1| adj2];
define adj1 [% adj1];
define adj2 [% adj2];
define nc [% nc];
define ems [art|adj|nc];
regex ems;
```

Las categorías resultantes son:

```
xfst[1]: print words
```

```
nc
```

```
adj1
```

```
adj2
```

```
art2
```

```
art1
```

```
xfst[1]:
```

Por otro lado, formulamos las propiedades descritas anteriormente según el paradigma 5P. Para esto, se crea el archivo **prosaal.txt** con las siguientes especificaciones:

```
source catsaal.txt

define amod [ems | % °]+;

define nuc [${% ° nc} | ${% ° adj1}|${% ° adj2}];

define unic1[${?art}];

define unic2[${?% nc}];

define unic3[${?% adj2}];

define unic [unic1 & unic2 & unic3];

define unicnuc [$.% °];

define nucleo [nuc & unicnuc];

define exig1 [[${% art1} & ${nc|% ° adj1|% ° adj2}] ~${% art1}];

define exig2 [[${% art2} & ${° adj}] ~${% art2}];

define exigen [exig1 & exig2];

define exclu1 [${% nc} & ~${° adj} | ~${% nc}];

define exclu2 [${% art2} & ~${% adj1 |% adj1}] ~${% art2}];

define exclu3 [${% art2} & ~${% adj1|% adj2}] ~${% art2}];

define exclu4 [${%art2} & ~${nc}] ~${%art2}]

define exclu [exclu1 & exclu2 & exclu3 & exclu4];

define preced1 [art < [nc|adj]];

define preced2 [% adj1 < % adj2];

define preced3 [% nc < % adj2];

define precednu [ems<[${%art1}][%art2]]];

define preced [preced1 & preced2 & preced3 & precednu];
```

```
define saa [amod & unic & nucleo & exigen & exclu & preced];
```

```
regex saa;
```

Al implantar las propiedades anteriormente formuladas, el programa XFST no despliega resultados. Se indaga sobre los posibles errores de formulación, redundancias, contradicciones y otros aspectos que pudieran detener el proceso. En consecuencia, se decide un planteamiento alternativo que expande la definición de categorías y omite la formulación de propiedades. Para este fin, se crea el archivo **catsaa2.txt**:

```
define art1 [% el| % la |% los|% las];
```

```
define art2 [% lo];
```

```
define adj [adj1|adj2];
```

```
define adj1 [% adj1];
```

```
define adj2 [% adj2];
```

```
define nc [% nc];
```

```
define SAA [art1 nc][[art1 %adj1]][[art1 %adj2]][[art2 %adj2]][[art2 %adj1];
```

```
regex SAA;
```

En este archivo se puede observar la especificación de los artículos que integran la subcategoría art1 y el artículo que integra la subcategoría art2. Se mantuvo la distinción de dos subcategorías para el adjetivo. Sin embargo, no se han establecido ejemplos concretos para evitar la dificultad que traería la concordancia en género y número con el artículo. En este sentido, una combinación del tipo “el + adj1” impide que resulte un caso concreto como «el + dichosa», que evidenciaría la necesidad de considerar la propiedad *concordancia*.

Además, se han definido las posibilidades combinatorias bajo la etiqueta SAA. La definición de estas estructuras se hace en lugar de la formulación de propiedades. De hecho, se está restringiendo la posibilidad de que aparezcan juntas ciertas categorías y subcategorías. Esto es lo mismo que se hace cuando se definen propiedades de exclusión y exigencia.

Los resultados obtenidos se transcriben a continuación:

```
xfst[1]: print words
```

```
lo adj2
```

lo adj1

el nc

el adj1

el adj2

la nc

la adj1

la adj2

los nc

los adj1

los adj2

las nc

las adj1

las adj2

xfst[1]:

## 7. INTERPRETACIÓN DE LOS RESULTADOS

Con el objetivo de visualizar la interpretación de los resultados, se ofrecen ejemplos de cada combinación expresados en lenguaje natural.

lo adj2      Lo regional/ Lo filial/ Lo constitucional

lo adj1      Lo bueno/ Lo malo/ Lo feo

el nc        El abismo/ El padrino/ El cuerpo

el adj1      El turista/ El ilusionista/ El fugitivo

el adj2      El internacional/ El policial

la nc        La fuga/ La tregua/ La trastienda

la adj1      La buena/ La mala/ La fea

la adj2      La regional/ La estatal

los nc	Los cazafantasmas/ Los testigos/
los adj1	Los sospechosos/ Los inadaptados/ Los mercenarios
los adj2	Los regionales/ Los internacionales
las nc	Las horas/ Las bandas
las adj1	Las suplicantes/ Las estafadoras
las adj2	Las fenicias/ Las troyanas

## 8. CONCLUSIONES

La hipótesis que origina la propuesta fue confirmada, ya que no resultaron combinaciones agramaticales del tipo *lo nc*. Al no desplegarse resultados a partir de la implantación en máquina de las propiedades, no se ha podido acceder a resultados con tres elementos, como podría ser la combinación *el nc adj1*. De lograr la modificación en las propiedades que permita hacer correr el programa sería esperable obtener resultados como: *el nc adj1*, *el nc adj2*, *el adj1 nc*, *el adj1 nc adj2*.

## 9. PROYECCIONES

Una posible línea de profundización en este planteo podría realizarse teniendo en consideración la ejemplificación de sustantivos y adjetivos, así como fue presentado en la categoría artículo. Para poder realizar esta ejemplificación sería necesario incluir la especificación de la propiedad *concordancia*. De este modo, se evitaría la aparición de construcciones agramaticales del tipo *\*la hombres sabios*.

Por otro lado, resultaría enriquecedora la incorporación de la categoría *adverbio*. Esto supondría estudiar distinciones dentro de la categoría para determinar las restricciones correspondientes.

## Bibliografía

- Bonino, R. (2009). Presentación de la lingüística computacional, en Solana, Z. (ed.), *La interlengua de los aprendientes del español como L2*. Centro de Estudios de Adquisición del Lenguaje, Facultad de Humanidades y Artes, UNR.
- Gelbukh, A. (2014). *Computational Linguistic and Intelligent Text Processing*. Springer. 15<sup>th</sup> Conference, CICLing. Kathmandu, Nepal. Part. 1.
- RAE y ASALE (2010). *Nueva gramática de la lengua española. Manual*. Espasa, Buenos Aires.
- Solana, Z. y Rodrigo, A. (2005). El sintagma nominal núcleo, en Castel, V. (ed.), *Desarrollo, implementación y uso de modelos para el procesamiento automático de textos*. Facultad de Filosofía y letras, UNCUYO.
- Solana, Z. (2006). *Un aporte al análisis automático de textos: el sintagma nominal núcleo en las producciones escritas infantiles*. RASAL.
- Solana, Z. (2008). Los clíticos como clasificadores de los verbos. *INFOSUR*.
- Trouilleux, F. (2007). Specifying Properties of a Language with Regular Expressions, en Bonino, R. (2009). *Presentación de la lingüística computacional*. Centro de Estudios de Adquisición del Lenguaje, Facultad de Humanidades y Artes, UNR, 8.

# **Usos de programas informáticos de traducción para la corrección del error en el aula de ELE: debates y propuestas**

**Julia Francés**

Universidad Nacional de Rosario. Facultad de Humanidades y Artes  
Rosario, Argentina

## **Abstract**

There is a wide range of computer programs and tools used sometimes systematically, other times randomly, by students and teachers of Spanish as a Foreign Language. Many of these resources directly or indirectly use translation as a strategy to aid in language learning or, at least, in the producing and understanding Spanish texts. This fact can mean that teachers may face new challenges, especially when it comes to a precise understanding of their students' written mistakes, but it can also support new learning goals. The study theoretically and practically analyzes the role translation programs can play in the Spanish classroom.

**Key words:** Spanish as a foreign language, machine translation, automatic morphological analysis, error analysis

## **Resumen**

Existe una amplia gama de herramientas informáticas utilizadas de forma a veces sistemática, a veces espontánea y aleatoria, por parte de los aprendientes de español como lengua extranjera y sus docentes. Muchas de ellas implican directa o indirectamente el uso de la traducción como estrategia para el estudio del idioma o, al menos, para la producción y comprensión de textos en español. Esto puede relacionarse con diversos retos para los docentes, en especial en lo que respecta a un análisis preciso de los errores, así como nuevas metas de aprendizaje. El presente trabajo analiza, de manera teórica y con ejemplos propios y de otros autores, el rol de los programas informáticos de traducción en el aula de ELE. Para ello, se repasan algunas de las herramientas lingüísticas que influyen en la oferta de materiales a disposición del aprendiente, problematiza la relación entre la traducción, el enfoque comunicativo y la corrección en el aula y estudia de manera teóricopráctica el uso de estas herramientas informáticas para trabajar los errores con los aprendientes de ELE.

**Palabras clave:** enseñanza de español como lengua extranjera, traducción automática, análisis de errores



## 1. INTRODUCCIÓN

La proliferación de herramientas informáticas vinculadas al lenguaje, como los programas de traducción automática, los de enseñanza de lenguas por computadora, los diccionarios en línea y los de análisis morfológico y de contenido, por nombrar solo algunas, son cada vez más importantes en el aula de enseñanza de español como lengua extranjera por diversos motivos. En primer lugar, como veremos, pueden influir en los conocimientos previos u obtenidos por fuera del aula por parte de los aprendientes. En segundo lugar, pueden mediar en el trabajo asignado por los docentes a sus alumnos, ya que pueden utilizarse, de forma permitida o no, para ayudar o realizar en su totalidad las tareas asignadas. En tercer lugar, pueden ser de utilidad para los docentes a la hora de afianzar conocimientos, tratar temas específicos o, incluso, como veremos, mejorar la experiencia en el aula para la corrección del error. Es decir que pueden considerarse como obstáculos o como oportunidades.

En el presente trabajo nos concentramos en un tipo de programas que resulta especialmente significativo: aquellos que usan directa o indirectamente la traducción como modelo. Vale la pena repasar los debates que surgen a partir de que algunas herramientas enfocadas al aprendizaje, como por ejemplo las que lo plantean como un juego o los diccionarios en línea, así como los traductores automáticos gratuitos. Todos ellos fueron imponiendo un cambio en la visión sobre la traducción dentro del enfoque comunicativo para ELE. A esto se le suma la importancia que va cobrando la denominada mediación comunicativa, que parecería reforzar la necesidad de desarrollar competencias traductoras en el aula de lengua extranjera. Así, las contradicciones y soluciones superadoras que se revisarán en el presente trabajo pueden ser de interés tanto para la disciplina de la traducción como para la enseñanza de lengua extranjera.

Las dificultades u oportunidades que generan las TIC, así como la compatibilidad de las técnicas de traducción con el enfoque comunicativo en el aula, son cuestiones que han generado diversos debates que nos interesa analizar. Al respecto, en el presente trabajo haremos una revisión bibliográfica donde veremos estudios que buscaron cuantificar los beneficios que puede tener el uso guiado de estas herramientas en el aula. Los pondremos en cuestión frente a aquellos que identifican las dificultades que puede traer su uso no guiado y, a veces, oculto de parte de los alumnos, en particular a la hora de corregir errores y detectar la evolución en el aprendizaje.

En esa misma línea, es nuestra intención profundizar sobre las dificultades y las oportunidades que pueden plantear estas tecnologías específicamente para la corrección del error. Para ello, es importante clasificar tanto los tipos de faltas como los efectos de la corrección en sí misma sobre el proceso de aprendizaje de la lengua. De esta manera, se pueden deslindar diversas cuestiones vinculadas con aspectos claves del enfoque comunicativo y de la didáctica de las lenguas que pueden interactuar de maneras distintas con la presencia de este tipo de tecnología dentro y fuera del aula. Por ello, consideramos interesante estudiar un ejemplo en el que se comparan los fallos de la máquina con los errores de los aprendientes y se utilizan estos análisis para superar algunas dificultades para desfosilizar errores, como la indiferencia, la frustración, la vergüenza, el aburrimiento, etc., y para debatir en el aula misma sobre la utilidad de la traducción.

Todo ello implica que se debatirán, de manera teórica y con ejemplos propios y de otros autores, el rol de los programas informáticos de traducción en el aula de ELE. Para ello, en primer lugar, se repasarán algunas de las herramientas lingüísticas que influyen en la oferta a la que se enfrenta el aprendiente de español, luego se esbozarán algunas maneras en las que la traducción como destreza y Google Translate como herramienta pueden representar un obstáculo para el enfoque comunicativo en el aula de ELE y para el análisis de errores. Más adelante, se indagará sobre las maneras en las que se puede combinar el uso de herramientas como SMORPH y de traducción automática para ayudar al docente a avanzar con los aprendientes en la desfosilización de errores persistentes y faltas morfosintácticas. Para ello, se analizarán ejemplos de diversos estudios previos y, en segundo lugar, una breve aplicación práctica realizada para este trabajo.

En la sección dos del trabajo se discutirá cómo algunas de las herramientas mencionadas en esta introducción alentaron a dar lo que se puede considerar un "giro traductológico" en la enseñanza de segundas lenguas y plantearon algunos desafíos para el enfoque comunicativo y los docentes de ELE. Asimismo, se esbozará el argumento por una mayor presencia de la traducción en el aula en general y el desarrollo de capacidades interlingüísticas. En la sección tres, se introducirá la cuestión del análisis del error y su corrección y algunos desafíos que pueden plantearse con el uso de nuevas tecnologías sin guía por parte del docente. Luego en la sección cuatro se presentará el estudio de caso donde que se utilizó un programa de traducción automática junto con uno de análisis morfológico automático para ayudar en el proceso de corrección de errores en el aula de ELE e indirectamente concientizar a los alumnos sobre los

problemas que puede traer el primer tipo de herramientas. Por último, en las conclusiones, se retomarán los aspectos clave de las tres secciones teóricas y los resultados del estudio de caso.

### **1.1. Las TIC y el aprendizaje del español**

Las herramientas informáticas pueden ofrecer distintas ventajas para aquellos que quieran acercarse al estudio y la comprensión de un texto escrito tanto en la lengua materna del lector como en una lengua extranjera. Para el caso de lenguas de amplia difusión como el español, existe gran número de usuarios de medios digitales con distintos fines lingüísticos. Estos pueden variar según el tipo de estudio a realizar y el nivel de comprensión y conocimiento de la lengua que tenga el interesado.

Una amplia gama de opciones informáticas parece ofrecer un primer acercamiento a un texto en español para un público que no conoce el idioma. En este marco, es cada vez más extensa la oferta y la utilización de las denominadas herramientas de traducción automática, que logran dicho objetivo con distintos niveles de calidad. Esto es particularmente interesante teniendo en cuenta dos aspectos. Primero, que el enfoque didáctico que priorizaba la traducción como método para la enseñanza de lenguas extranjeras fue predominante hasta que se reemplazó por el enfoque comunicativo [1]. Y, en segundo lugar, que la traducción profesional generalizada de textos, así como la interpretación, tiene como objetivo facilitar la comunicación entre hablantes de distintas lenguas sin necesidad de que los emisores y receptores tengan una lengua en común.

Algunas herramientas de traducción automática se publicitan comúnmente como puentes entre lenguas y hasta se insinúa el fin de la necesidad de aprender otro idioma. Estas son las denominadas "traducciones para asimilación", cuyo objetivo es el *gisting*, o la comprensión de lo esencial del mensaje comunicado. En ámbitos empresariales también se promocionan como formas de ahorrar costos en recursos humanos, cuestionando la necesidad de aprender otro idioma, y de facilitar la interconexión sin barreras con objetivos amplios.

Sin embargo, la oferta de "traductores" automáticos es diversa según los ámbitos y varía en cuanto a costos y funcionalidades. Sus objetivos pueden ser el antes mencionado *gisting* o la producción de textos que luego serán posteditados por traductores profesionales. El primero es de especial interés para este trabajo, ya que es el tipo al que tiene acceso el público en general y el que puede realmente tener consecuencias para el aprendizaje del ELE. Esta clase de

programas en general se subdivide según el uso de tres paradigmas principales. Los más antiguos son el basado en reglas y el estadístico, que utilizan diversos corpus lingüísticos y que se pueden combinar en esquemas híbridos [2]. Sin embargo, el salto cualitativo más importante de los últimos tiempos en este sentido es el de la técnica de redes neuronales aplicado desde sus inicios por el traductor DeepL y, en los últimos años, por Google Translate [3].

Existe una gran cantidad de programas de este tipo con diversos niveles de complejidad y de calidad, y pueden variar desde los más conocidos y generalizados como Google Translate, pasando los que van ingresando al mercado de manera lenta pero sostenida como DeepL, hasta los más especializados como el sistema PAHOMTS de la Organización Panamericana de la Salud. Si bien la calidad y el uso mismo del término "traducción" para estos procesos se encuentra en discusión, el rápido crecimiento de las comunicaciones globales lleva de manera cada vez más acelerada a la búsqueda de la concreción de la mítica solución superadora del "caos" y la incomunicación representada por la caída de la torre de Babel. En el mundo de la globalización y las telecomunicaciones en constante desarrollo, el público exige cada vez más opciones "inmediatas" o rápidas para solucionar los límites que encuentra a las interconexiones. Así, con el desarrollo de las tecnologías de la información (TIC), la intermediación humana (la traducción humana y la enseñanza presencial de una lengua) parece considerarse un elemento que "enlentece" u obstaculiza el rápido acceso a la información. Esto lleva a la constante búsqueda y priorización de alternativas automáticas.

El español, al ser la lengua materna de más de 470 millones de personas y la tercera más utilizada en Internet [4], despierta el interés de gran número de personas que no lo comprenden. Para comunicaciones informales, por ejemplo, a través de las redes sociales, se hace progresivamente más común la integración de dichas plataformas con herramientas gratuitas como Google Translate y *crowdsourcing translation* (traducción comunitaria) para ofrecer a los usuarios la posibilidad de relacionarse con personas que no hablen el idioma, sorteando la necesidad de aprender la lengua del interlocutor. La calidad de los productos de dichos procesos varía según la herramienta utilizada y la combinación de idiomas, pero suelen distar mucho de la traducción profesional [3]. Algunas herramientas más recientes, como DeepL, aseguran ofrecer una calidad superior a estas otras más conocidas, aunque para un número limitado de idiomas. Sin embargo, aún no existen estudios científicos independientes lo suficientemente amplios como para asegurarlo con contundencia [18] y los realizados con

textos en español en áreas temáticas puntuales encontraron que pueden obtenerse traducciones de calidad inaceptable [19].

Sin embargo, para algunos alumnos, esta puede ser una primera instancia de aproximación a la lengua extranjera (en el caso bajo estudio, el español). Además, muchas veces el aprendiente de ELE usa la traducción automática como un puntapié inicial para comenzar sus redacciones o comprender mejor las consignas [1]. En relación a ellos, algunos autores plantean que este puede significar un ejercicio útil a la hora de analizar las sutilezas de la lengua meta y las diferencias gramaticales, léxicas y culturales que existen entre esta y su lengua madre [5] incluso para niveles básicos e intermedios de aprendizaje (A2, por ejemplo, según García [1]).

Asimismo, es interesante destacar que algunos estudios indican que, en el caso de interacciones repetidas entre hablantes de distintas lenguas, en particular en redes sociales, el simple uso de herramientas automáticas puede significar cierto aprendizaje léxico por parte del que no es hispanoparlante al comparar el texto de origen con el texto meta [6]. Si bien esta introducción al lenguaje es informal, no estructurada y puede generar aprendizajes errados, no deja de ser un método espontáneo de uso frecuente que debe ser analizado con mayor profundidad para evaluar sus posibles consecuencias.

Otra manera de retomar el viejo enfoque didáctico de la enseñanza a través de la traducción es la de algunas plataformas de aprendizaje "*gamificado*" que fomentan el estudio del español a través de la traducción. Esto se plasma en programas de enseñanza de lengua extranjera como Duolingo, que integra ejercicios de traducción como metodología fundamental para la enseñanza del idioma con un enfoque lúdico [7]. En este caso, sin embargo, la traducción fue previamente realizada y revisada por humanos con fines didácticos y pedagógicos específicos, de manera tal que las frases a traducir (tanto desde la lengua materna como desde el español) se ordenan según el nivel de dominio del idioma alcanzado y el objetivo específico de cada lección. Lo llamativo de estas herramientas es que muchas veces omiten la explicación gramatical o semántica y se basan principalmente en la intuición y la prueba y error a través de ejercitación para la presentación de contenidos en un espacio de "juego" (con puntajes y premios). Para ello también requieren de traducciones simplificadas y, en especial en los primeros niveles, bastante literales y poco creativas (para focalizar la atención en un elemento gramatical o ítem léxico particular) [7].

Paralelamente, para la creación de herramientas de traducción automática con el paradigma basado en corpus, para el desarrollo de programas de enseñanza automatizados y para una amplia gama de estudios lingüísticos automatizados, no solo didácticos (análisis de discurso, tipologías textuales, estudios gramaticales, etc.), es cada vez más útil y necesario el uso de programas estadísticos y de análisis cuantitativo de textos. Los lingüistas dedicados al análisis del discurso en distintas áreas, así como aquellos que investigan las variaciones de un idioma a lo largo de la historia o entre regiones, pueden valerse de diversas herramientas informáticas para facilitar su trabajo. Este es el caso de programas como R, HyperPo, Wordstat y el proyecto NORA, que permiten obtener automáticamente conteos de frecuencias de palabras, contextos en los que aparecen palabras claves, coocurrencias y listas de distribución de diversos elementos léxicos. Estas investigaciones lingüísticas, a su vez, son esenciales para la conformación de libros de texto y estrategias para la enseñanza de ELE a distintos niveles.

Pero además, encontramos otras herramientas que permiten profundizar aun más el estudio de un texto escrito o transcripción de la oralidad. Estas son aquellas que proveen información acerca de la morfología, lo cual es particularmente relevante para el español, dada su amplia variedad de flexiones nominales, adjetivales y verbales (comparado con otras lenguas de uso difundido como el inglés).

Existen entornos de desarrollo lingüístico como NOOJ, que procesan textos y corpus a nivel ortográfico, léxico, morfológico, sintáctico y semántico. Para el caso del inglés, con una morfología flexiva diferente, el investigador dispone de programas gratuitos como Word Hoarder y MorphAdorner. Para el español, las posibilidades son más limitadas y debe tenerse en cuenta la existencia de tres modelos de analizadores morfológicos automáticos [8]: el de análisis (el procedimiento de análisis), el de la gramática que se usa (las clases gramaticales de palabras) y el de la implementación computacional (el formalismo usado). Con un modelo de análisis a través de generación, encontramos el caso de AGME (Analizador y Generador de la Morfología del Español), del Instituto Politécnico Nacional de México.

Otro caso interesante es el de SMORPH, que permite también su utilización para el análisis de producciones de aprendientes de español como L2 de manera tal que se puede analizar la interlengua, con sus características particulares, incluso superando la idea de "error" en el proceso de aprendizaje y focalizando en una visión cuantitativa (que luego permite estudios cualitativos) de este sistema en su totalidad.

## 2. EL GIRO TRADUCTOLÓGICO

Existen múltiples estudios sobre el uso de recursos informáticos en general tanto para la enseñanza de español como segunda lengua como para la investigación sobre este campo. Si bien está claro que el uso de este tipo de herramientas implica un salto cuanti y cualitativo en ambos campos, lo cierto es que es mucho más reducido el número de estudios sobre los cambios paradigmáticos en el ámbito de la didáctica de segundas lenguas.

Una de estas cuestiones se trata en un volumen dedicado al giro traductológico en la enseñanza de L2 reforzado por la proliferación de recursos digitales que facilitan tanto el acceso a traducciones automáticas como a recursos didácticos que usan la traducción como metodología principal. Nos referimos al trabajo de Carreres, Muñoz-Calvo y Noriega-Sánchez [9], que introducen un número del *Journal of Spanish Language Teaching* donde se plantea que se encuentran evidencias del uso de la traducción a lo largo de la historia de la enseñanza de lenguas extranjeras. Sin embargo, aseguran, con la hegemonía del enfoque comunicativo, la actitud hacia ella de parte de investigadores y docentes comenzó a ser negativa y se la consideró contraproducente en el aula. Más específicamente, las autoras señalan que a partir de 2000 los estudios sobre el tema en el ámbito del inglés como lengua extranjera comenzaron a ser cada vez más abundantes, mientras que esta cuestión es mucho menos tratada para el caso de la enseñanza de español. En este marco, plantean el acceso a nuevas tecnologías tanto de traducción automática como aquellas destinadas a la enseñanza de lenguas (como las plataformas de aprendizaje *gamificado*) como una parte del nuevo impulso que tiene el tema. Otra cuestión que las autoras consideran crucial para el renovado interés por el tema es el debate suscitado por la inclusión de la idea de mediación comunicativa en el Marco Común Europeo de Referencia, concepto que relacionan con la destreza de la traducción. Al respecto, identifican una escasez de actividades mediadoras en los manuales de ELE y un trabajo muy limitado sobre la competencia plurilingüe en el aula. En este sentido remarcan que "el fenómeno traductor [está] intrínsecamente vinculado con la comunicación e inserto en un contexto cultural, [lo cual] nos lleva a entenderlo como una actividad mediadora" [9: 101].

La cuestión del fácil acceso a recursos informáticos centrados en la traducción y la de la competencia mediadora ponen en cuestión los intentos de exclusión total de la traducción en el aula de ELE. La pregunta entonces podría ser si traducir es una actividad accesoria e instrumental dentro del proceso de aprendizaje o si es un objetivo en sí mismo. Carreres,

Muñoz-Calvo y Noriega-Sánchez [9] plantean que debe ser ambas cosas a la vez dentro de un enfoque comunicativo adaptado. Otros autores como Clifford, Merschel y Munné [10], Conde Noguero [11] ponen el foco sobre la inevitable realidad del uso persistente que les dan los aprendientes de español a este tipo de herramientas y las actitudes positivas que estos tienen tanto frente a la actividad de traducción (especialmente en el ámbito académico) en sí misma como frente a los recursos que dicen facilitarla. Estos autores, así como Doucar y Houk Schocket [12], abogan por la integración en el aula de las herramientas de traducción automática y las plataformas lúdicas de enseñanza de segundas lenguas a través de la traducción. Si a ello se le añade el concepto de "traducción pedagógica", discutido por Carreres, Muñoz-Calvo y Noriega-Sánchez por ser demasiado reduccionista, pero reforzado por autoras como D'Amore y Bobadilla [13], entendemos que en la actualidad es difícil sostener el rechazo absoluto a cualquier intento de traducción por parte de los aprendientes de ELE, ya sea a través de herramientas informáticas o como ejercicio propio.

Nos preguntamos entonces en primer lugar (de manera general e introductoria) cuáles son algunas de las maneras en las que la traducción automática puede servir de instrumento en el aula. En segundo lugar, indagamos si otras herramientas informáticas como el análisis morfológico automático pueden ayudar a los docentes de ELE a encontrar modos innovadores de tratar la corrección del error.

### **3. LA CORRECCIÓN DEL ERROR**

La cuestión del error en el aula de ELE puede trabajarse desde los estudios de adquisición de L2 y en el marco de la didáctica para, por un lado, intentar comprender a qué se deben los fallos cometidos por los aprendientes y cómo clasificarlos, y por el otro, cómo ayudar a los alumnos a superarlos. Desde el campo del análisis del error, autores como Alexopoulou [14] se concentran en su tratamiento explícito en el aula de ELE desde un enfoque comunicativo. Pero también en este marco, el uso de herramientas informáticas puede generar un cambio importante en la disciplina.

Según el estudio realizado por O'Neill [15] en Estados Unidos, el 87,7 % de los aprendientes de español y francés de nivel intermedio usa Google Translate para las tareas asignadas incluso en cursos en los que está prohibido hacerlo. En el mismo trabajo, el autor concluye que las redacciones de alumnos de nivel intermedio en español que usan traductores



automáticos tienen menos errores que aquellas realizadas sin esta herramienta. Si el uso de estos programas es tan generalizado y tiene efectos sobre la detección del error, entonces es esperable tanto que aparezcan errores en las tareas entregadas al docente que pueden haber sido inducidos por la máquina como que se encubran errores del estudiante. El objeto de este trabajo no es analizar cuáles son los errores que surgen por el uso no guiado de este tipo de recursos, pues de ello se ocupan ya otras disciplinas. Pero sí nos proponemos estudiar algunas formas en las que el error de los recursos digitales puede ayudar a los aprendientes de ELE a ser conscientes de los que ellos mismos cometen en sus propias redacciones e incluso en producciones orales.

En el trabajo citado de O'Neill [15], el autor compara el rendimiento en actividades escritas de alumnos que usan esta herramienta con y sin capacitación previa frente a quienes usan solo diccionarios digitales o ningún recurso informático. Su conclusión es que el mejor desempeño se logra con una utilización guiada y entenada de las herramientas digitales. Sin embargo, no hay suficiente información como para asegurar que la producción de escritos mejor calificados implique un mejor aprendizaje de la lengua y un desarrollo adecuado de las distintas competencias. Asimismo, se necesitan más estudios específicos para saber si alentar a que los aprendientes usen Google Translate en todas o la mayoría de sus tareas escritas perjudica o fomenta las destrezas escritas. Fredholm [16], por ejemplo, se ocupó particularmente de sus efectos sobre el desarrollo del vocabulario escrito en español y concluyó que no hay un efecto positivo duradero sobre la variedad léxica del estudiante. Por su parte, O'Neill indagó sobre los efectos motivadores que puede tener el uso de traductores en línea para alentar a los aprendientes a usar con mayor frecuencia el idioma que están estudiando.

Los estudios citados hasta aquí, así como muchos otros trabajos sobre el tema, parecen concentrarse en el entrenamiento de los alumnos para un uso generalizado de Google Translate tanto para las tareas de clase como para su carrera académica por fuera del aula de ELE. En los términos planteados en la sección 2, parece ponerse el foco en la necesidad de desarrollar la destreza traductoril del alumno con la ayuda del programa informático y no tanto en la posible utilidad de este recurso como herramienta para ayudar al aprendizaje mismo de español. Dicho en otros términos, no es igual la producción eficiente de textos mediada por una máquina que el desarrollo de todas las destrezas que implica el conocimiento de una lengua.

La propuesta de aprender a través de los errores del traductor automático aparece en algunos estudios como los de White y Heidrich, [17], Ducar y Schocket [12], aunque suelen focalizarse en la comparación de la L1 y la L2. Otra serie de estudios se concentra focalizarse en debates éticos sobre el uso de este tipo de programas y no tanto sobre sus efectos pedagógicos [15]. No son tantos los trabajos realizados sobre los problemas que puede implicar para el docente el encubrimiento de algunos errores propios de los alumnos al utilizar para sus tareas estas herramientas.

Como indica Alexopoulou [14], en el marco de los estudios de la interlengua, con la aparición del análisis del error y el giro metodológico hacia el enfoque comunicativo en el aula de lengua extranjera, se comenzó a replantear el tratamiento de los fallos de los aprendientes y a remarcar su importancia pedagógica como indicador del proceso de aprendizaje. En este contexto, la corrección se considera esencial para "evitar y prevenir la fosilización de formas erróneas" [14:106], para lo cual debe superar las dificultades planteadas por la indiferencia y el aburrimiento por un lado y la ansiedad y la frustración por otro. Pero, como indicábamos más arriba, ante el amplio uso de recursos informáticos por parte de los estudiantes, en la actualidad la identificación de las dificultades e insuficiencias de los alumnos para mejorar sus producciones autónomas se encuentra con importantes obstáculos. Desde el enfoque comunicativo, se pone al aprendiente en el centro del proceso de adquisición de la L2 y se considera que este se encuentra formulando continuamente hipótesis sobre el sistema lingüístico y estas se ven confirmadas o rechazadas según si la comunicación es exitosa o no. Pero este hecho puede ponerse en duda si el estudiante cuenta con un recurso que le permite prescindir de dichas hipótesis.

Es conveniente entonces mencionar las dimensiones de la corrección para ver cómo pueden interactuar con el uso de traductores en línea. Alexopoulou [14] señala tres. Por un lado, la reparación consiste en la interacción que tiene lugar en el aula de ELE. En este caso, al tratarse de un contexto de producción con mayor facilidad de control sobre el uso informático por parte del docente, es más probable que se pongan en juego hipótesis propias de los aprendientes y que se puedan identificar más claramente las características de su interlengua. En segundo lugar, se trata de desfosilizar, es decir, trabajar sobre aspectos problemáticos de la interlengua de los estudiantes con cierta fluidez pero claras dificultades y escasos indicadores de cambio. En este caso, es más complejo encontrar cuáles son los puntos en los que el aprendiente no logra avanzar, ya que puede haber falsos indicadores de cambio para tareas en las que utiliza

algunas herramientas en línea. Por último, se trata de retroalimentar, es decir, favorecer la fosilización de aspectos positivos de la interlengua para que estos se conviertan en nuevas reglas del sistema. En el caso del *feedback* positivo, este podría ayudar no solo a reforzar lo que hace el alumno correctamente de forma autónoma sino también a hacer consciente y a apropiarse de las características positivas producidas gracias a la ayuda de la máquina.

Estos tres tipos de corrección se pueden cruzar con las tres categorías de fallos o "desviaciones" citadas por Alexopoulou [14]. En primer lugar, el "error persistente" consiste en incorrecciones sistemáticas que no logran ser superadas. En segundo lugar, la "falta" consiste en una equivocación eventual que puede llegar a ser corregida por el aprendiente mismo. Por último, el "lapsus" es una falla cuya explicación se puede encontrar en factores extralingüísticos, como las condiciones psicológicas o físicas de quien lo comete. El error persistente suele requerir de técnicas de desfosilización o reparación, mientras que las faltas y los lapsus pueden requerir retroalimentación para evitar que se conviertan en sistemáticas, pero muchas veces pueden ser corregidas por el mismo hablante.

De estas combinaciones, nos interesa buscar una mirada innovadora sobre las técnicas de desfosilización de errores persistentes y las de reparación de errores y faltas. Para ello, reconocemos el inevitable lugar que tienen las herramientas de traducción automática en el proceso de aprendizaje de español como lengua extranjera y buscamos formas para que los resultados arrojados por su uso no sean un obstáculo al proceso de corrección sino un disparador de debates que permitan cumplir con los mismos objetivos en el aula.

Para ello, un primer paso por parte del docente es lograr identificar los errores (en sentido amplio) tanto producidos en clase como los que aparecen en las tareas sin supervisión. Ello puede implicar un registro de algunas instancias orales, así como de actividades escritas realizadas en el aula. Luego, como indica Alexopoulou [14], estos se pueden agrupar según un criterio lingüístico para saber a qué nivel del sistema corresponden, es decir, si son morfosintácticos, léxico-semánticos, discursivos o pragmáticos. Consideramos que particularmente en el primer caso, es de gran utilidad la aplicación de herramientas de análisis morfológico automático tales como SMORPH sobre corpus constituidos por la totalidad de las producciones de una clase para llegar a una tipología aun más detallada que facilite la comprensión acerca de las hipótesis erradas que llevan a este tipo de desviaciones. Puede ser de

utilidad también hacer comparaciones entre las actividades realizadas en el aula y aquellas que podrían incluir recursos electrónicos.

Luego, la autora [14] plantea ir más allá de la identificación y la clasificación, para llegar a una reflexión sobre las cuestiones problemáticas de la interlengua de parte del aprendiente mismo. En este sentido, considera que es pertinente buscar una sensibilización ante este tipo de fenómenos, un apoyo para el propio reconocimiento de errores y su autocorrección y un incentivo a la investigación para la subsanación de dichos fallos. Para ello, Alexopoulou [14] indica que es fundamental trabajar sobre los elementos afectivos que pueden determinar que la actitud del alumno frente a la corrección sea positiva o negativa. Y en ese sentido ofrece algunos consejos: no exponer a la persona que comente el fallo frente a sus pares y evitar que los aprendientes le tengan miedo a la corrección, alentar a la corresponsabilidad en el aprendizaje y buscar que todas las experiencias sean constructivas.

## **4. OTROS USOS PRÁCTICOS**

En las secciones anteriores se analizaron algunos ejemplos de uso de programas informáticos, especialmente de traducción automática, en el aula de ELE y su vinculación con los errores de los aprendientes. Se mencionaron casos prácticos y estudios estadísticos a corto y mediano plazo en los que se detectaron desafíos para la corrección efectiva del error así como beneficios en ciertos contextos. En este apartado, nos proponemos analizar un ejemplo específicamente diseñado para este trabajo donde se comparan los errores humanos y los producidos por la máquina en un tipo de texto particular y se busca aprovechar didácticamente el error del traductor automático para la corrección del error humano en el aula.

### **4.1. Metodología**

A partir de los consejos indicados en la sección anterior, se estudiaron los errores morfosintácticos más comunes en un grupo de diez alumnos de español como lengua extranjera de entre 25 y 32 años con un nivel intermedio-avanzado de español (entre B1 y C1). En primer lugar, se les pidió que respondieran una consigna: "Piense en la casa de su infancia. Descríbala en un párrafo de tres oraciones en español".

La consigna descriptiva fue diseñada con el objetivo de posibilitar el uso del pasado del indicativo así como alentar a la utilización de múltiples sustantivos y adjetivos que podrían presentar dificultades en cuanto a la morfología desde el punto de vista del género gramatical. Al tratarse de alumnos de nivel medio-avanzado, se entiende que estos temas ya fueron tratados en diversas ocasiones en la clase de ELE y que deben manejarse con bastante destreza. Asimismo, se seleccionó una temática que permitiera poner en juego emociones y experiencias personales, pero que, a la vez, evocara un mismo referente concreto y claro en ambos idiomas para que los textos fuera más fácilmente comparables y el espectro de posibles respuestas en cada uno fuera más reducido.

Por otro lado, los textos descriptivos que se refieren a objetos concretos, sin diálogos ni necesidad de uso de expresiones idiomáticas ni culturemas particulares, presentan un desafío menor para los programas de traducción automática [11]. Si bien la consigna posibilita el uso metafórico del lenguaje, no lo requiere ni lo incentiva específicamente. Además, el hecho de que se pidan tres oraciones permite que el programa tenga bastante contexto para seleccionar opciones léxicas. De esta manera, se presenta una consigna que no es particularmente compleja para la máquina.

Más tarde, se les pidió a los estudiantes que realizaran la misma tarea en su lengua materna sin mirar el texto que habían escrito en español. El ejercicio se realizó en un entorno controlado, donde los alumnos no podían utilizar herramientas informáticas. Luego, se conformaron dos corpus, uno con los textos originalmente en español y otro con los resultados arrojados por la traducción automática de los textos realizados en la L1.

El próximo paso fue analizar del primer y segundo corpus con SMORPH. A continuación, se buscó identificar errores morfosintácticos en los textos generados por Google Translate. Luego, se compararon ambas series de errores y se seleccionaron dos tipos de fragmentos: aquellos en los que el programa en línea cometía errores y los alumnos no y aquellos en los que el traductor digital cometía errores similares a los de los estudiantes. Por último, se trabajaron ambos tipos de errores en clase. Por un lado, se reforzaron con retroalimentación positiva aquellas hipótesis exitosas por parte de los alumnos en los casos en que el programa había fallado. Por otro lado, se debatieron algunos de los errores más frecuentes que los alumnos cometen en sus textos propios y que la máquina comete también.

Cabe destacar que, como indicábamos más arriba, el uso del análisis morfológico automático puede facilitarle al docente la identificación del tipo de errores morfosintácticos más comunes en su grupo de alumnos a nivel global. Ello no significa desdeñar la importancia de los errores léxicos, discursivos y pragmáticos, pero sí permite trabajar en aspectos puntuales que suelen preocupar más a los aprendientes de segunda lengua [14] y que suelen llevarlos al uso de traductores automáticos antes que a diccionarios [15]. Así, se pueden encontrar las áreas en las que es fundamental un trabajo de desfosilización y también de reparación en este nivel en particular.

#### **4.2. Resultados y discusión**

Un primer resultado que consideramos importante destacar es que en nueve de los diez casos estudiados, los textos en L1 y en español tenían diferencias significativas en cuanto a la longitud de las oraciones redactadas, las descripciones realizadas y la variedad de estructuras gramaticales seleccionadas. Es decir que los alumnos eligieron contar diferentes aspectos de lo que en principio parecería ser un mismo referente. Los aprendientes utilizaron distintas estrategias y se valieron de recursos diferentes para comunicarse en una y otra lengua. Este primer resultado puede poner en discusión la necesidad de practicar las habilidades de traducción para la destreza de mediación cultural, ya que podría estar indicándonos que la persona puede expresarse de diferentes maneras respecto a un mismo referente sin necesidad de traducir exactamente el mensaje. Este es un aspecto interesante que vale la pena profundizar en futuros estudios.

A partir del pequeño corpus creado, con la aplicación de SMORPH se pudieron identificar dos problemas morfosintácticos repetidos en los textos en español de los aprendientes: la asignación de género gramatical incorrecto a algunos objetos inanimados, aunque no a personas; la conjugación del pretérito imperfecto del modo indicativo (con el verbo "ser"), aunque no tanto su uso en sí mismo. Como indicábamos en la sección anterior, la consigna se diseñó con el objetivo de ver el uso del pasado y el control gramatical, en particular en lo que respecta al género. Los problemas detectados en las producciones humanas, entonces, responden en parte al diseño de la tarea. Cabe destacar que el análisis comparativo se realizó antes de discutir los errores en la clase, lo cual permitió buscar los temas más productivos para su uso didáctico. Las dos cuestiones identificadas (género gramatical y formas del

pasado del indicativo) son claves para un nivel intermedio de español y, por lo tanto, fue de gran utilidad encontrar que los corpus generados abrían un debate al respecto.

Por un lado, en lo que respecta al pretérito imperfecto, entendemos que Google Translate no produce formas verbales inválidas en español, por lo que, en este aspecto, la máquina representa una ventaja. Sin embargo, sí encontramos que en los textos arrojados por este método había una subutilización de este tiempo verbal mucho más marcada que la identificada en la de las redacciones de los alumnos en español. Esto implica que, si bien las formas morfológicas utilizadas por los alumnos pueden no ser siempre las adecuadas, el uso del tiempo verbal en sí es más pertinente en el caso de los textos de aprendientes que en las traducciones automáticas.

Por otro lado, encontramos que los textos generados por el traductor automático tenían problemas con el género gramatical para personas y pronombres y no con objetos inanimados, a diferencia de las redacciones de los alumnos. Si bien esto requeriría mayor estudio posterior, a priori parecería que, en general, el uso incorrecto de género gramatical para seres humanos puede ser más problemático y generar mayor confusión en los interlocutores que el error en el género gramatical de objetos inanimados.

Estas diferencias y similitudes entre los errores de la máquina y de los estudiantes permitieron que se dedicaran dos clases al debate sobre estos dos temas. Por una parte, al utilizar ejemplos en los que esta herramienta digital había cometido errores y los alumnos mismos no, las discusiones que surgieron a partir de las comparaciones entre unos y otros textos mostraron actitudes positivas sobre el logro de "corregir" o "superar" al traductor automático. En ese sentido, consideramos que reforzaron positivamente las hipótesis exitosas de los alumnos.

Más tarde, se trabajó también con errores que cometieron tanto Google Translate como los alumnos mismos. En primer lugar, se compararon de manera general los textos humanos y los automáticos. Se buscó demostrar que la computadora también arrojó errores y que, por ello, el uso de este tipo de programas debe ser cuidadoso y reflexivo, teniendo en cuenta algunos de los puntos trabajados en la sección 3. Los comentarios al respecto de parte de los aprendientes fueron muy positivos en cuanto a la visibilización de las problemáticas que pueden surgir a partir del uso de estos programas y en cuanto a la capacidad de ellos mismos de actuar como correctores de los textos producidos digitalmente. Al tratarse de una consigna simple, que no parecía diseñada para provocar errores específicos en la máquina, se visualizó que el

programa de traducción puede arrojar errores incluso en este tipo de tareas y no solo en textos más complejos. Además, sin prohibir su uso, el llamado a la precaución y a la reflexión buscó concientizar sobre los desafíos que este tipo de herramientas puede presentar y, de esta manera, generar una discusión más allá de un simple rechazo.

Al mostrar que la máquina comete errores y compararlos con los propios se buscó mitigar la "vergüenza" por equivocarse. En este marco, se alentó al debate y la crítica de frases específicas arrojadas por la computadora, ya que no habían sido redactadas por los estudiantes mismos o sus compañeros. Esto evitó exponer directamente a quienes habían cometido errores similares en sus propias producciones en español y, a la vez, reforzar los aciertos de las producciones humanas frente a la máquina.

La identificación de dos temas fundamentales para el nivel medio de español, como el género gramatical y la conjugación y uso del pasado, permitió que en clase se refrescaran estos conceptos en general y en particular a partir de ejemplos concretos tanto humanos como automáticos.

Entendemos que sería necesario realizar un estudio longitudinal para identificar la utilidad a largo plazo de esta metodología y saber si realmente logra desfosilizar errores persistentes y faltas, así como reforzar la retroalimentación positiva que consolide las hipótesis positivas de los aprendientes. Sin embargo, consideramos que fue un primer paso donde vimos que se facilitó la desestigmatización del error, se alentó al aprendizaje autónomo y la autocorrección a la vez que se reconoció el uso extendido de las herramientas informáticas como Google Translate. Asimismo, se alertó a los alumnos sobre sus falencias, pero rompió con la prohibición absoluta del uso de la L1 en el aula de ELE y se reconoció su utilidad pedagógica, aunque accesoria, incluso en un enfoque comunicativo.

## **5. CONCLUSIONES**

A partir de los artículos repasados en este trabajo y de los ejemplos propios y de otros autores, podemos extraer una serie de conclusiones y ofrecer algunas líneas para continuar investigando.

En primer lugar, hicimos un repaso de la amplia gama de herramientas informáticas que se vinculan directa o indirectamente con la enseñanza de lenguas extranjeras. En este caso, vimos que algunas de ellas, como las plataformas lúdicas en internet y los traductores



automáticos, parecerían forzar un debate sobre el tema de la traducción en la clase de lengua extranjera. Asimismo, identificamos algunos recursos que pueden ser de utilidad para las investigaciones sobre adquisición y enseñanza de español, tales como las herramientas de análisis morfológico automático o las de análisis de contenido.

En segundo lugar, mencionamos estadísticas sobre el uso generalizado de herramientas de traducción automática que demuestran que la prohibición o la negación de su existencia no resuelven el problema. Por un lado, se puede poner en cuestión el enfoque didáctico mismo y, sin guía de parte del docente, se puede generar una reedición del viejo método de enseñanza a través de la traducción. Por otro lado, es posible que le dificulten al docente la identificación de errores persistentes de los aprendientes. Esto puede representar un obstáculo para el análisis correcto del error y para el trabajo en el aula para la desfosilización de errores persistentes y faltas, en especial en la escritura.

Ante esta situación, evaluamos que las opciones no son solo la utilización o prohibición de estas herramientas, sino que existen alternativas para un uso productivo y eficiente que no deje de responder a los objetivos comunicativos en el aula. Al respecto, vimos algunos autores que plantean su utilidad parcial para metas concretas y la necesidad de un entrenamiento para su uso más eficiente. De igual modo, encontramos que la tecnología puede estar al servicio del análisis y la corrección de errores morfosintácticos al facilitar tanto el trabajo de identificación por parte del docente (con análisis automáticos como SMORPH) como ofrecer alternativas en el aula para superar la estigmatización de la equivocación y la exposición de los alumnos frente a sus pares. Sería interesante también analizar la posibilidad de alentar al análisis crítico independiente de los alumnos a través del uso del análisis morfosintáctico automático sobre sus propias redacciones.

Por último, se abren a partir de aquí diversas líneas de estudio. Por un lado, sería necesario un trabajo a largo plazo para analizar en profundidad los efectos psicológicoafectivos del uso de textos arrojados por la máquina en el proceso de corrección en lugar de escritos propios de los alumnos. Además, sería importante evaluar si ello conlleva una mayor facilidad para desfosilizar errores persistentes y reforzar la retroalimentación positiva en el aula.

Asimismo, sería de utilidad estudiar la aplicación de esta estrategia a errores en otros aspectos del lenguaje. Por ejemplo, se podría estudiar si esta estrategia de uso comparativo entre textos producidos por los alumnos y textos generados por Google Translate puede ayudar a

desarrollar un mayor y mejor uso del vocabulario, a diferencia del efecto casi nulo a largo plazo que puede tener el simple uso sin guía de esta tecnología, como encuentra Fredholm [16]. Para ello sería útil la aplicación de una herramienta de análisis morfológico automático como SMORPH (que "limpie" las diferencias morfológicas para un mismo lexema) para luego utilizar sin problemas un programa de frecuencia de palabras o de conformación de mapas mentales comparados.

Por fin, la comparación entre las traducciones de los textos en L1 y las producciones en español podría abrir diversas preguntas acerca de la diferencia en los recursos que utilizan los aprendientes en una y otra lengua y las estrategias que utilizan para sortear las dificultades en la lengua extranjera. Esto, a su vez, posiblemente añadir nuevos matices interesantes a la competencia mediadora, como esbozamos en el punto 4.2.

## Referencias

- [1] García, I. Can machine translation help the language learner? *ICT for language learning*, 3rd Edition, Pixel.
- [2] Ruiz Costa-Jussá, M. (2016). Integración de Paradigmas de Traducción Automática. *Sociedad Española para el Procesamiento del Lenguaje Natural*, Nro. 57, Jaén.
- [3] Bojar, O.; Chatterjee, R.; Federmann, C.; Graham, Y.; et al. (2016). Findings of the 2016 Conference on Machine Translation. First Conference on Machine Translation (WMT16). *The Association for Computational Linguistics*. 131–198.
- [4] Instituto Cervantes (2016). *El Instituto Cervantes presenta el nuevo anuario "El español en el mundo"*. Sala de Prensa. 19 de enero de 2016.
- [5] Lynn, B. (2016). How will machine translators change language learning? *Learning English*. 18 de diciembre de 2016.
- [6] Trusty, A. y Truong, K. (2011). Augmenting the web for second language vocabulary learning. *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 3179–3188.
- [7] Munday, P. Duoling (2017). Gamified learning through translation. *Journal of Spanish Language Teaching*. 4(2): 194-198.
- [8] Gelbukh, A. y Grigori, S. (2006). *Procesamiento automático del español con enfoque en recursos léxicos grandes*. Centro de Investigación en Computación, Instituto Politécnico Nacional, México, 1ra. edición.
- [9] Carreres A, Muñoz-Calvo M y Noriega-Sánchez M. (2017). Translation in Spanish language teaching: the fifth skill/La traducción en la enseñanza del español: la quinta destreza. *Journal of Spanish Language Teaching*. 4 (2): 99-109.
- [10] Clifford J.; Merschel L. y Munné, J. (2013). Tanteando el terreno: ¿cuál es el papel de la traducción automática en el aprendizaje de idiomas? *@tic. Revista D'Innovació Educativa*. 0 (10) 108-121.
- [11] Conde Noguero, M. E. (2018). Los traductores automáticos en línea como recurso metodológico en el aula de español como lengua extranjera. *Transforming education for*

*a Changing World*, Editado por López García y Manso. Adaya Press, 1ra. edición, Eindhoven, 304-312.

- [12] Ducar, C. y Houk Schocket, D. (2018). Machine translation and the L2 classroom: Pedagogical solutions for making peace with Google translate. *Foreign Language Annals*. 51(4): 779-795, invierno de 2018.
- [13] D'Amore, A.M. y Bobadilla, N.R. (2013). La traducción pedagógica en el siglo XXI. *Revista Iberoamericana para la Investigación y el Desarrollo Educativo*. (10), enero-junio 2013.
- [14] Alexopoulou, A. Aproximación al tratamiendo del error en la clase de ELE desde la perspectiva del análisis de errores. *Estudios de Lingüística Aplicada*. (41): 101-125.
- [15] O'Neill, E. (2019). Training students to use online translators and dictionaries: the impact on second language writing scores. *International Journal of Research Studies in Language Learning*. 8(2): 47-65, julio de 2019.
- [16] Fredholm, K. (2019). Effects of Google translate on lexical diversity: vocabulary development among learners of Spanish as a foreign language. *Revista Nebrija de Lingüística Aplicada a la Enseñanza de las Lenguas*. 13(26): 98-117, febrero de 2019.
- [17] White, K. D., y Heidrich, E. (2013). Our policies, their text: German language students' strategies wi th and beliefs about web-based machine translation. *Die Unterrichtspraxis*. 46 (2), 230-250, julio de 2013.
- [18] Benjamin, M. (2019). *Is DeepL Better than Google Translate? A No, A Punt, and A Yes - When & How to Use Google Translate*. Proyecto digital Teach You Backwards, 30 de marzo de 2019.
- [19] Aguilar Canal, E. O. (2019). *Evaluación de la calidad del traductor automático DeepL del inglés al español en textos especializados*. Tesis de Grado, Universidad César Vallejo, Lima.

# **El análisis automático de la sílaba escrita del español mediante la herramienta NooJ y su aplicación pedagógica**

**Rodolfo Bonino**

Centro de Estudios de Tecnología Educativa y Herramientas Informáticas de  
Procesamiento del Lenguaje, UNR  
Rosario, Argentina  
rodolfobonino@yahoo.com.ar

**Andrea Rodrigo**

Centro de Estudios de Tecnología Educativa y Herramientas Informáticas de  
Procesamiento del Lenguaje, UNR  
Rosario, Argentina  
andreafrodrigo@yahoo.com.ar

## **Abstract**

The purpose of this work is to outline and explore our proposal for the automatic analysis of the written syllable in Spanish. After making some general considerations about the syllable, we step into analyzing the syllable in the Spanish language. To do this, we will use NooJ, the software developed by Silberztein (2015). We take a five syllabic model distinction for the Spanish syllable, which are very useful for us to make a thorough analysis of the Spanish lexicon but, which are also very nonspecific as for their combination properties. In order to address these five syllabic models, we created a *productive grammar*. When analyzing a text, in some cases we see that the application of this grammar produces possible analysis but only one of them is correct. These cases serve us as an approach to make a pedagogic application of NooJ, as we understand that the trainee teacher can choose a correct segmentation but always under the guidance of a trained teacher. This process involves the development of metalinguistic skills as well as the use of a software-based tool in order to combine and integrate different areas of knowledge and expertise which will be necessary for their careers.

**Keywords:** automatic analysis, software NooJ, pedagogic application, Spanish, written syllable.

## Resumen

Presentamos aquí una propuesta de análisis automático de la sílaba escrita en español. Luego de hacer algunas consideraciones sobre la sílaba, nos ocupamos de la formalización en NooJ, software desarrollado por Silberztein (2015), para analizar automáticamente la sílaba escrita. Nos dedicamos a cinco tipos de modelos silábicos del español, los más frecuentes dentro del centenar de modelos de que dispone esta lengua. Para abordar estos cinco modelos silábicos, creamos una gramática productiva. Se observa que en algunos casos, la aplicación de la gramática genera dos análisis posibles, siendo solamente uno el correcto. Como se trata aquí de una aplicación pedagógica de NooJ, entendemos que el estudiante, siempre con la guía del profesor, puede elegir la segmentación correcta. Este proceso implica el desarrollo de las habilidades metalingüísticas, tan importantes en su formación, a la vez que la utilización de una herramienta informática implica conjugar diferentes saberes necesarios en la futura práctica docente.

**Palabras claves:** análisis automático, software NooJ, aplicación pedagógica, español, sílaba escrita.

## 1. INTRODUCCION

En las últimas décadas, el estudio de la gramática y de la fonología fue desplazado de los programas escolares por otros contenidos como el análisis del discurso, la pragmática, la sociolingüística y la teoría de la comunicación. La fonología, que tradicionalmente ocupaba un lugar aún más marginal que la morfología y la sintaxis se redujo a la práctica de la separación silábica como base para la enseñanza de la acentuación ortográfica e inclusive en algunos casos, hasta se fue eliminando esta práctica de la clase de lengua del primer ciclo del nivel primario. Esta perspectiva cuenta con el aval de muchos especialistas en educación; pero también ha recibido numerosas críticas, entre las que podemos citar a Di Tullio (2005) y Tapia y Riestra (2014). A partir de nuestra convicción de que la gramática y la fonología son herramientas de gran importancia en el proceso de aprendizaje de la lengua, creemos que estas disciplinas deben adaptarse a los nuevos métodos de enseñanza para reincorporarse actualizadas al sistema educativo. Para nosotros, recuperar el estudio de la fonología en la enseñanza de la lengua en general y del español en particular, constituye una instancia insoslayable, sobre todo en la formación docente.

Por tanto, aunque es previsible que los análisis automáticos que produzcan las gramáticas elaboradas por los estudiantes sean inferiores a los de los programas elaborados por lingüistas; el proceso de elaboración en sí mismo exige el estudio de las gramáticas convencionales la reflexión sobre la lengua y una explicitación del conocimiento que tiene el alumno como hablante nativo.

Este es un punto muy importante a nuestro juicio, puesto que la **reflexión metalingüística** constituye una parte fundamental en el aprendizaje de la lengua y en la formación de los futuros docentes.

En líneas generales, la capacidad de reconocer las sílabas que constituyen una palabra se adquiere sin mayores dificultades en los primeros años de escolaridad; pero la precisión en los resultados puede lograrse de manera intuitiva y no implica necesariamente un conocimiento formal de la sílaba. Además de ser una base para el correcto uso de la tilde (al igual que el resultado intuitivo) es un conocimiento metalingüístico específico. Sin embargo, la dificultad mayor que observamos es que esta práctica de *separar en sílabas* (que como ya anticipamos parece un ejercicio considerado en muchos casos como anticuado y prescindible, aun en los

primeros años de escolaridad primaria) se conforma como algo aislado y sin ninguna relación con otros conocimientos lingüísticos.

En la web, es posible hallar diversos programas de libre acceso que realizan el procedimiento en forma automática con gran eficiencia; pero, como no muestran las gramáticas o los cálculos estadísticos en los que se sustenta el análisis, funcionan como cajas negras que no aportan ningún conocimiento lingüístico más allá del resultado. A tal efecto, y tomando en cuenta las herramientas informáticas para el procesamiento del lenguaje disponibles en línea, el *software* NooJ<sup>1</sup> resulta una herramienta de gran importancia, ya que es posible interactuar creando nuestras propias gramáticas y diccionarios, lo que implica **aprender haciendo**, en este caso en función de los datos del español. Al utilizar NooJ, (herramienta que nos interesa fundamentalmente, porque puede descargarse sin costo alguno) se percibe la sistematicidad del conocimiento lingüístico. Así la elaboración de diccionarios y de gramáticas debe guardar una coherencia interna que de no observarse nos impide el llegar a un resultado satisfactorio.

Actualmente en la página web de NooJ, encontramos recursos lingüísticos de numerosas lenguas: árabe, croata, inglés, francés, alemán, griego, hebreo, italiano, japonés, portugués, ruso y por supuesto español, por citar algunas de ellas. En el módulo español, se encuentra el trabajo del grupo dirigido por Andrea Rodrigo de la Facultad de Humanidades y Artes de la Universidad Nacional de Rosario<sup>2</sup>, cuyos primeros avances fueron a través del Grupo Infosur. Esto nos puede dar una idea de la proyección de esta presentación y del marco en el que se inscribe, ya que la presencia de investigadores de diferentes lenguas en un solo *software* implica por tanto una interacción constante y la idea de una verdadera comunidad de investigación en franco avance.

En este trabajo, en particular, nos proponemos desarrollar un formalismo para el análisis automático de la sílaba mediante NooJ, que, si bien no arroja resultados demasiado precisos, permite reflexionar acerca de la estructura silábica del español y exige elaborar estrategias para su mejoramiento. Primero presentamos algunas consideraciones en torno a la sílaba en general y a la sílaba escrita en particular, luego, la formalización en NooJ (gramática productiva), seguimos con la aplicación de dicha gramática al análisis de un texto y por último presentamos un ejercicio de aplicación pedagógica concreto.

---

1 Disponible en: <http://www.nooj-association.org/downloads.html> [última consulta septiembre 2020].

2 Dicho módulo se hizo como parte del proyecto Infosur entonces dirigido por la Dra. Zulema Solana, hoy está en franco desarrollo por el Centro de Estudios de Tecnología Educativa y Herramientas Informáticas de Procesamiento del Lenguaje (Facultad de Humanidades y Artes, UNR).



## 2. DESARROLLO

### 2.1. La sílaba

#### 2.1.1. Consideraciones generales

Presentamos aquí una descripción básica de las propiedades la sílaba; en el siguiente apartado, se analizan los problemas que suscitan estas propiedades para la formalización en NooJ. Si bien la sílaba ha sido trabajada por muchos autores, algunos de los cuales se introducen en la bibliografía, no es nuestra intención explorar estas posibilidades, sino solamente detenernos en una de las caracterizaciones más aceptadas, la que proporciona Franch y Blecua (1975)<sup>3</sup>.

Estos autores señalan los problemas teóricos que suscitan la definición y descripción de la sílaba. Ahora bien, dado que nuestro trabajo se centra en la sílaba escrita, podemos soslayar las dificultades que implica la distinción entre sílaba fonética y sílaba fonológica, y definir nuestro objeto de estudio como una cadena de caracteres constituida por grafemas que representan conjuntos de fonemas vocálicos y consonánticos con un alto grado de fusión e influencias recíprocas.

#### 2.1.2. Partes de la sílaba

Se distinguen dos tipos de constituyentes silábicos: el núcleo o cima, que puede ser simple (vocal) o compuesto (diptongo o triptongo) y los márgenes que también pueden ser simples (consonantes) o compuestos (grupos consonánticos). El núcleo es un elemento imprescindible, mientras que los márgenes prenuclear o explosivo (ataque) y postnuclear o implosivo (coda) pueden estar ausentes. Volviendo a Franch y Blecua (1975)<sup>4</sup>: los tres tipos silábicos más frecuentes del español son:

- consonante - vocal (sílaba libre),
- consonante - vocal – consonante,
- vocal,

Siendo el tipo **vocal** el menos frecuente en oposición al tipo **consonante-vocal** como el más frecuente. Estos tres tipos silábicos constituyen más del 80% de las estructuras silábicas que

---

3 Alcina Franch, J. y Blecua, J. M. (1975). *Gramática española*. p. 255 y ss.

4 Alcina Franch y Blecua. *Gramática española*. p. 264

aparecen en español. El tipo silábico **vocal - consonante** y los que tienen núcleo o márgenes compuestos se dan con menor frecuencia, pero no por ello dejan de ser productivos.

Las secuencias de fonemas /b/, /c/, /f/, /g/, /p/ y /t/ seguidos de /r/ o /l/, y /d/ seguido de /l/ son, generalmente, inseparables y constituyen el ataque silábico; las otras secuencias de dos consonantes están formadas por la coda de una sílaba y el ataque de la siguiente (*ab.sur.do*).

Las secuencias /rs/, /ns/, /bs/ forman coda silábica cuando van seguidas de otra consonante (*trans.por.te*, *abs.trac.to*), en caso contrario, integran sílabas diferentes (*an.sioso*, *ab.sur.do*).

Las secuencias de cinco consonantes son excepcionales y se dan solo en palabras procedentes de otras lenguas, en este caso, las tres primeras forman la coda de una sílaba y las dos últimas el ataque de la sílaba siguiente (*tungs.te.no*). En lo que respecta a la delimitación silábica, en las palabras aisladas las consonantes tienden a fusionarse con la vocal siguiente /a.cua.ré.la/, por lo tanto, salvo que sea final de palabra, toda sílaba con coda (sílaba trabada) presupone que la siguiente comienza con consonante (ataque) y toda sílaba con vocal inicial es comienzo de palabra o está precedida por una sílaba abierta.

### 2.1.3. Particularidades de la sílaba escrita

La representación gráfica de los fonemas plantea algunas cuestiones específicas:

- El grafema *h* no representa ningún fonema. En la separación silábica, este grafema se mantiene unido a la vocal que lo sigue (*i.nhi.bir*, *co.he.ren.cia*), excepto en los casos de las interjecciones, donde está a final de palabra (*ah*, *eh*, *oh*, *uh*).
- El grafema *y* constituye una sílaba cuando se encuentra aislado, seguido de vocal es el ataque silábico y precedido de vocal forma parte de un diptongo o triptongo a final de palabra.
- Las secuencias *ll* y *qu* son dígrafos que aparecen solo en el ataque silábico y representan, el primero los fonemas /k/, /dʒ/ o /ʒ/, según la variante dialectal; y el segundo, el fonema /k/.
- La secuencia *rr* habitualmente es un dígrafo que representa el fonema vibrante múltiple /r/, pero, en algunas palabras como *hi.per.rea.lismo*, el primer grafema representa el fonema vibrante simple

/r/ final de una sílaba y el segundo al fonema vibrante múltiple /r/ que inicia la sílaba siguiente.

- La secuencia *gu* seguida de las vocales *e* o *i* es un dígrafo que representa el fonema /g/ (*gui.so, gue.rra*), pero cuando va seguido de consonantes o de vocales distintas de *e* o *i* representa la secuencia de fonemas /gu/ (*a.gua, gu.sa.no*)
- El grafema *x* representa dos fonemas /ks/ que pueden pertenecer a la misma sílaba /éks.tra/ o a sílabas diferentes /ek.sén.to/. En este caso, aplicamos las pautas de separación silábica propuestas por la *Ortografía* de la Real Academia Española (RAE) (2010)<sup>5</sup>: cuando va seguida de vocal se ubica al inicio de la sílaba (*e.xen.to*) y cuando va seguida de consonante es cierre de la sílaba anterior (*ex.tra*)

6) Entre los hispanohablantes no existe uniformidad en la manera de articular las secuencias vocálicas: por diversas razones en algunos casos se pronuncian dentro de la misma sílaba y en otros, en sílabas distintas. Dado que en español existe una relación directa entre la separación silábica y la acentuación ortográfica, la RAE (2010): ...*"ha establecido una serie de convenciones para determinar qué secuencias deben considerarse siempre hiatos o siempre diptongos o triptongos a efectos ortográficos"*<sup>6</sup>...

En el presente trabajo, utilizamos estas convenciones, que se detallan en lo que sigue.

#### 2.1.3.1. Diptongos ortográficos

- 4 Vocal abierta (/a/, /e/, /o/) seguida o precedida de vocal cerrada átona (/i/, /u/),
- 5 Dos vocales cerradas distintas (/i/, /u/)

#### 2.1.3.2. Triptongos ortográficos

Secuencias formadas por una vocal abierta precedida y seguida por una vocal cerrada átona

#### 2.1.3.3. Hiatos ortográficos

- a Vocal cerrada tónica seguida o precedida de una vocal abierta.
- b Dos vocales abiertas distintas.

---

5 RAE, p. 406

6 RAE, p. 225

c Dos vocales iguales.

## 2.2. Formalización en NooJ

Varias de las propiedades de la sílaba descritas en el apartado anterior resultan difíciles de formalizar mediante una herramienta informática. Como se anticipó, no pretendemos lograr un análisis perfecto de la sílaba (tareas automáticas que ya pueden hacer algunos programas disponibles en Internet<sup>7</sup>), sino orientar acerca de los problemas que implica la formalización. Para nuestro análisis, proponemos una gramática productiva (.nom) que agregue a todas las palabras la etiqueta *SBS* más cada una de las sílabas que integra la palabra. De este modo se espera que en el análisis de un texto, por ejemplo, la palabra *cadete* reciba la etiqueta *SBS+ca+de+te*.

Distinguimos los siguientes tipos silábicos:

- 1 SBN: sílaba formada únicamente por el núcleo (simple o compuesto).
- 2 SBAN: sílaba formada por ataque y núcleo (simples o compuestos).
- 3 SBNC: sílaba formada por núcleo simple y coda (simples o compuestos).
- 4 SBANC: sílaba formada por ataque, núcleo y coda (simples o compuestos).

En el gráfico inicial de la gramática, cada tipo silábico se declara como una variable que arroja como salida el valor de la variable:

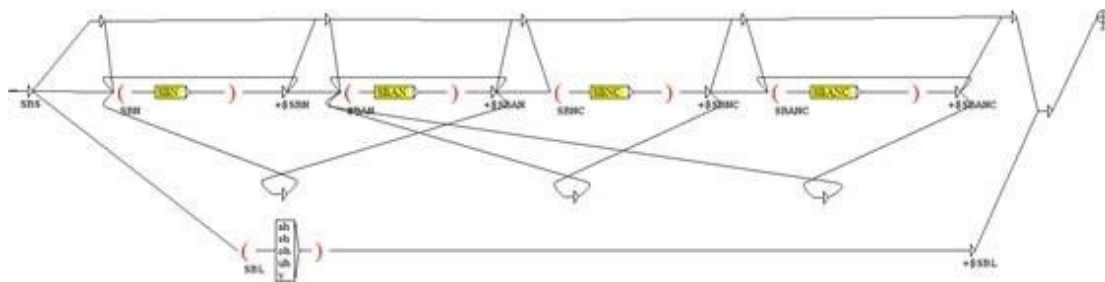


Figura 1. Tipos silábicos y sus posibles combinaciones

Las posibles combinaciones entre estos tipos silábicos se establecen mediante las conexiones: la conexión superior que va desde el nodo inicial hasta el nodo final con entrada y salida en cada tipo de sílaba permite formalizar que todas pueden formar palabras aisladas y que,

<sup>7</sup> Al respecto ver: <https://www.separarensilabas.com/index.php> [Consulta 21 de junio de 2020].

además no tienen restricciones en lo que respecta a la concatenación con los tipos que se encuentran a su derecha (de izquierda a derecha todos los tipos silábicos pueden combinarse con cualquiera de los que le siguen). Las conexiones de cada tipo silábico sobre sí misma indican que un tipo silábico se puede repetir una o más veces. Las conexiones inferiores van de derecha a izquierda y formalizan las restricciones combinatorias: las sílabas cerradas aparecen a final de palabra o seguidas de sílabas que comienzan por consonantes y las sílabas que comienzan con vocal aparecen a principio de palabra o precedidas por otras vocales. La primera de estas restricciones es también la causa de que la variable SBNC no tenga conexión sobre sí misma. La cadena inferior de la gramática permite declarar que el carácter *y* y las cadenas *ah*, *eh*, *oh*, *uh* forman sílabas solo cuando aparecen aisladas.

A continuación, se muestran las imágenes de las gramáticas electrónicas creadas para analizar cada uno de los constituyentes de la sílaba (ataque, núcleo y coda):

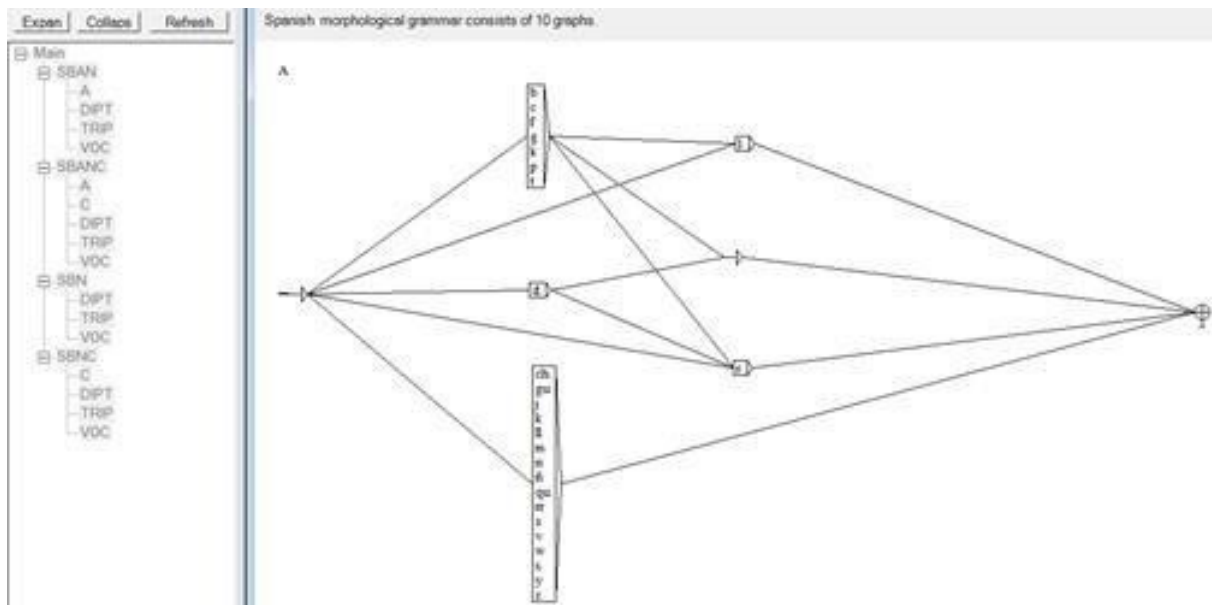


Figura 2. Ataque

El ataque silábico puede estar formado por una consonante o un grupo consonántico cuyo segundo elemento es *r* o *l*. Las consonantes se dividen en tres grupos: las que forman grupo con *r* y *l*, la que forma grupo únicamente con *r* y las que no forman grupo. Los dígrafos se tratan como consonantes simples.

El núcleo puede ser vocal simple, diptongo o triptongo. Cada uno de estos tipos se analiza en un gráfico diferente:

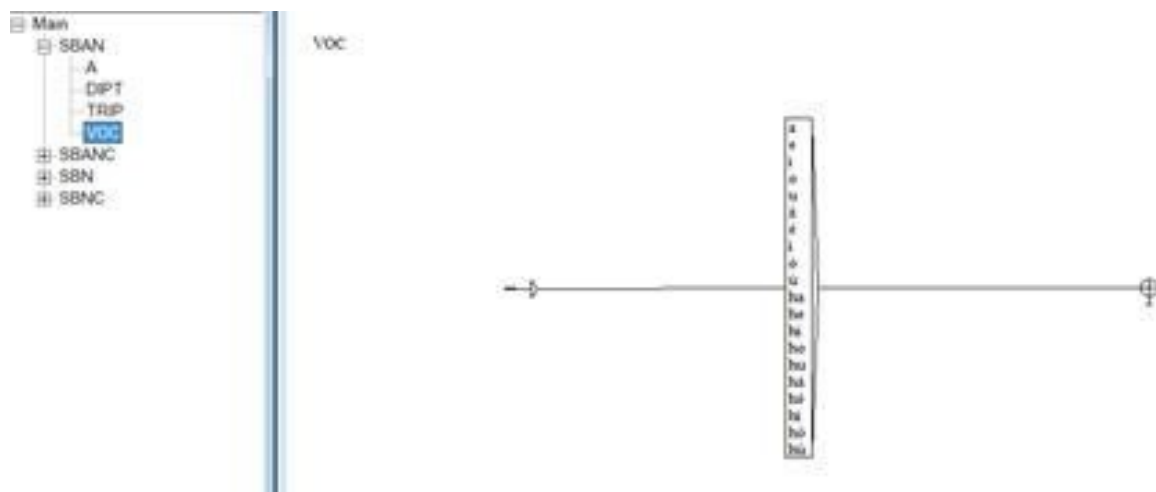


Figura 3. Núcleo vocálico

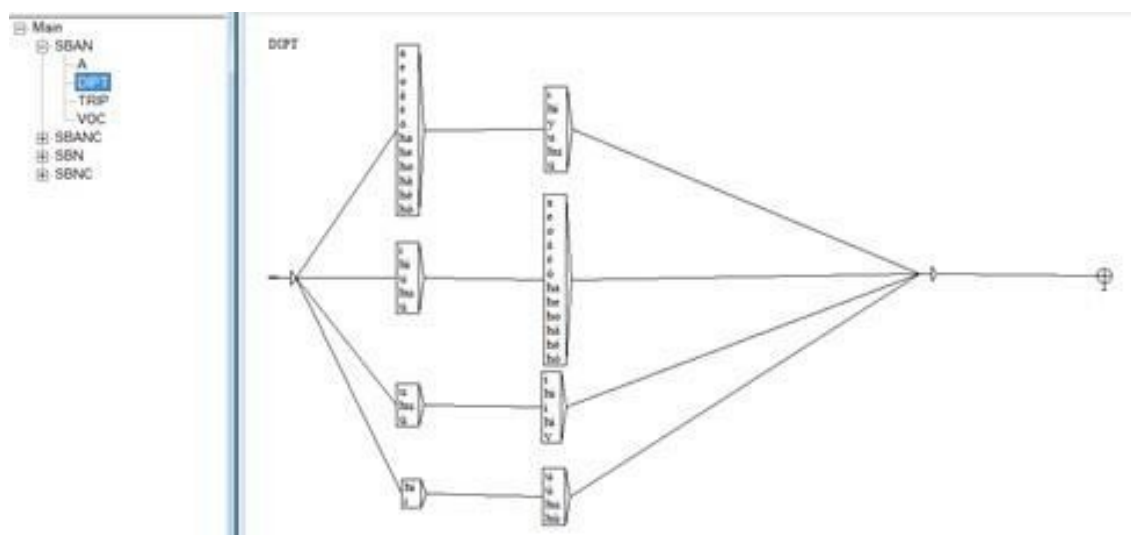


Figura 4. Núcleo diptongo

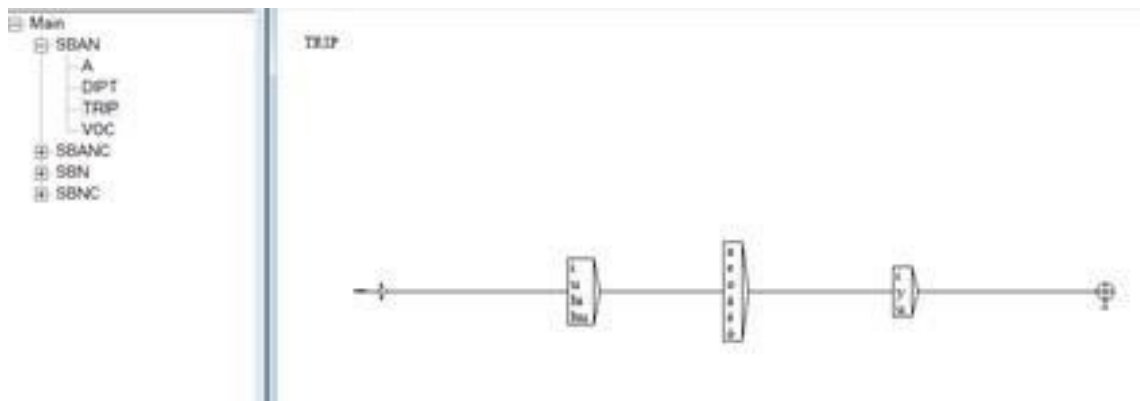


Figura 5. Núcleo triptongo

En la coda, no aparecen todas las consonantes y los grupos consonánticos se reducen a *ns*, *rs*, *bs* y *gns*.

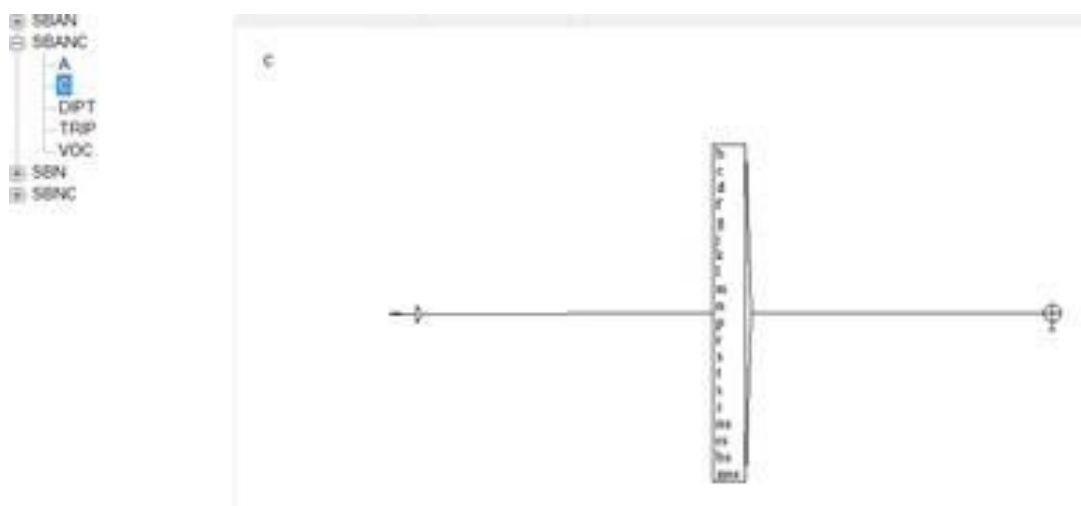


Figura 6. Coda

### 2.3. Aplicación de la gramática al análisis de un texto

De la aplicación de las diferentes gramáticas (Figuras 1 a 6) del punto anterior al análisis de "La muchacha del atado"<sup>88</sup> y completando con los diccionarios y gramáticas flexivas de que ya dispone el Módulo español Argentina, observamos que: 475 *token* (unidades separadas por espacios) solo tres no pueden ser reconocidas y analizadas, 356 reciben un único análisis correcto y 116 además del análisis correcto reciben otros que no lo son.

8 Texto trabajado en el módulo español, Argentina, Disponible en: <http://biblioteca.derechoaleer.org/biblioteca/roberto-arlt/aguafuertes-portenas/la-muchacha-del-atado.html> [Consulta 21 de junio de 2020].

### 2.3.1. Palabras desconocidas (función UNKNOWNNS)

Nuestras gramáticas diseñadas para la sílaba no son capaces de analizar *Arlt*, *Neumann* y *Vladimiro*. Estas palabras tienen combinaciones impropias del español: *rlt* y *nn* no son codas posibles y *vl* no constituye un ataque silábico.

### 2.3.2. Análisis correctos (función UNAMBIGUOUS)

Los 356 *tokens* que reciben un análisis correcto no presentan ninguno de sus constituyentes compuestos.

### 2.3.3. Análisis ambiguos (función AMBIGUITIES)

NooJ efectúa más de un análisis de los tokens que tienen alguno de sus constituyentes compuestos. Extraemos a continuación algunos casos a modo de ejemplo:

- c <apiadando, SBS+a+pi+dan+do> <apiadando, SBS+a+pi+a+dan+do>
- d <hablar, SBS+hab+lar> <hablar,SBS+ha+blar>
- e <suya, SBS+suy+a> <suya,SBS+su+ya>

Como se puede observar, el análisis erróneo de (1) no se puede atribuir al hecho de que las sílabas reconocidas no sean propias del español, dado que la sílaba *pi* resaltada en el ejemplo aparece en otras palabras, como *pi-no* y *a* es una sílaba en *a.mor*. Una gramática que pretenda ser eficiente debería reconocer ambas sílabas; pero debería restringir la combinación de una sílaba terminada en vocal cerrada átona con una que comience con vocal abierta, y viceversa.

Aunque en español existen sílabas con coda *b* (*ab.sur.do*) y sílabas con ataque *l* (*co.la*), la gramática debería evitar la combinación de sílabas terminadas en *b*, *c*, *f*, *g*, *k*, *p*, *t* con sílabas con ataque *r* o *l* y la de sílabas con coda *d* o *r* con sílabas que comienzan con *r*.

Para impedir el análisis erróneo de (3) la gramática debería limitar la presencia de sílabas terminadas en diptongos *ay*, *ey*, *oy*, *uy* o triptongos *uay*, *uey*, *iaiy*, *iey* solo a final de palabra. El texto no tiene palabras que contengan el dígrafo *gu*, pero la función *Debug* de la gramática muestra las dificultades que suscita el análisis de estas secuencias.

	Output
SBAN ("SBAN" ("A" gu ) ("VOC" í ) \$) )\$(SBN ("SBN" ("VOC" ) ) )	SBS+gui+a
SBAN ("SBAN" ("A" g ) ("VOC" u ) \$) )\$(SBN ("SBN" ("VOC" ) )\$(SBN ("SBN" a ("VOC" \$) ) ) )	SBS+gu+i+a
SBAN ("SBAN" ("A" g ) ("DIPT" uí ) \$) )\$(SBN ("SBN" ("VOC" ) ) )	SBS+gui+a

Figura 7. Captura de la pantalla de salida de la función *Debug* para la palabra *guía*



En la captura de pantalla, se puede ver que solo el primer análisis reconoce que la primera sílaba de la palabra está formada por el ataque *gu* seguido de la vocal *í* y la segunda sílaba es la vocal *a*. El tercero, aunque en la columna final produce una segmentación correcta, no es válido, porque analiza la primera sílaba como si estuviera formada por la consonante *g* seguida por el diptongo *uí*. Para producir un análisis correcto, sería necesario especificar que no hay sílabas formadas por consonante *g* seguida de los diptongos *ue*, *ué*, *uí*, *uí*. Aunque en este trabajo, no se pretende elaborar una gramática para obtener el análisis y la generación de todas las sílabas posibles del español, excluyendo las ambigüedades; la observación de los resultados de nuestra propuesta nos permite estimar que para lograrla se deberían distinguir más de un centenar de modelos silábicos con diversas combinaciones entre sí. Esta cantidad de modelos y de composiciones no es formalizable en una gramática como las que mostramos precedentemente que contiene apenas cinco modelos silábicos; por lo que habría que pensar en otra estrategia. No obstante estas gramáticas rudimentarias pueden ser altamente eficaces para la enseñanza de la lengua. Si bien no alcanzan un cien por cien de cobertura, pueden llegar a analizar una gran cantidad de sílabas en español. En el punto 2.4. presentamos algunas de las posibilidades que surgen aplicándolas a la enseñanza del español.

#### **2.4. Aplicación pedagógica**

Antes de pasar a un ejercicio concreto de aplicación didáctica, valen algunas consideraciones sobre nuestra práctica como docentes investigadores. Hemos observado repetidamente que muchos estudiantes no llegan a adquirir elementos básicos propios de su lengua, encontrándose dudas por ejemplo al final de oración, cómo cortar la palabra o dónde cortarla. Esto por qué ocurre. Simplemente, en algún momento de su etapa de alfabetización no se hizo suficiente hincapié en el lazo fonológico de la adquisición de la escritura, a nuestro criterio paso imprescindible. Algunos cambios que se están implementando en la enseñanza de la escolaridad en los primeros años, deben tomar en cuenta que la sílaba debe ser reconocida y que justamente la sonoridad de la misma constituye un elemento que permite reconocer a la palabra en todos sus constituyentes. Este abordaje, si bien se detiene en la sílaba escrita, puede proporcionar elementos para trabajar la sílaba desde el tratamiento automático y simultáneamente integrado a otros conocimientos lingüísticos. Por lo tanto, nos hace reflexionar en el sentido de que no se trata de un contenido caduco, sino de un engranaje de

vital importancia en la enseñanza de la lengua que tranquilamente puede trabajarse desde otro lugar.

Proponemos a modo de ejemplo algunos ejercicios posibles que se pueden trabajar utilizando la gramática presentada en el punto 2.2.:

1. Aplique la gramática al análisis automático de “La muchacha del atado”.
2. Efectúe una búsqueda automática de las palabras desconocidas y explique por qué la gramática no las reconoce.
3. Efectúe una búsqueda automática de las palabras que reciben análisis ambiguo y marque con X la separación silábica correcta.
4. Busque manualmente palabras con diptongos y palabras con hiato, verifique el análisis automático y explíquelo
- 5- Busque manualmente palabras con grupos consonánticos y palabras con secuencias de sílabas con coda seguidas de sílabas con ataque, verifique el análisis automático y explíquelo.
- 6- Busque manualmente palabras con dígrafos *ch*, *ll*, *gu* y *qu*, verifique el análisis automático y explíquelo.

### 3. CONCLUSIÓN

En el presente artículo, nos propusimos presentar algunas cuestiones referidas a la formalización de la sílaba escrita en español valiéndonos del *software* NooJ. En esta propuesta, solamente tomamos cinco modelos silábicos, por lo que constituye apenas una primera aproximación. No ha sido nuestro objetivo abarcar la totalidad de los modelos posibles en español, sino plantear en sí una formalización posible, utilizando gramáticas productivas. Esto nos permite en principio analizar secuencias dejando para un futuro trabajo, la opción de generación, que por una cuestión de espacio no pudimos presentar aquí.

En lo que respecta al análisis, nos remitimos al texto “La muchacha del atado” de Roberto Arlt. Si bien no se logra una perfección total en los resultados logrados, existe un alto margen positivo, ya que de 475 tokens (unidades separadas por espacios), 356 reciben un análisis correcto, o sea, casi el 75% y 116 reciben más de un análisis posible, es decir, que presentan

algún grado de ambigüedad, lo que representa un 24%, finalmente, un 1% restante no puede ser reconocido. Esto nos da un importante margen de cobertura y precisión.

Para culminar, nos detuvimos en proponer algunos ejercicios que podrían constituir una aplicación didáctica de esta gramática. Estos ejercicios no tienen como destinatarios a los estudiantes, sino más bien a los docentes en formación, como un objetivo central de nuestro proyecto. En efecto, reforzar la formación docente, es a nuestro juicio una meta que como investigadores debemos plantearnos para que no haya un abismo entre la investigación y la aplicación de la misma. Esto en el marco del proyecto de aplicación pedagógica de NooJ, “Aplicación de la herramienta informática NooJ para desarrollar estrategias argumentativas en la comprensión y producción de textos académicos” (2018-2021) dirigido por Andrea Rodrigo

## Agradecimientos

A la Doctora Zulema Solana, sin cuya inspiración no hubiera sido posible nuestro trabajo, ya que siempre nos alentó como Directora del Proyecto Infosur a dar impulso a nuestra investigación dentro de la lingüística computacional y creó la Revista Infosur para poder dar visibilidad a la producción científica dentro de este campo.

## Referencias

- [1] Arlt, R. (1942). La muchacha del atado. En: *Aguafuertes porteñas*. Disponible en <http://biblioteca.derechoaleer.org/biblioteca/roberto-arlt/aguafuertes-portenas/la-muchacha-del-atado.html>.
- [2] Alcina Franch, J. y Blecua, J. M. (1980). *Gramática española*. 2da. ed. Ariel, Barcelona.
- [3] Di Tullio, Á. (2005). *Manual de gramática del español*. La isla de la luna, Buenos Aires.
- [4] González Díaz, M. J. et al. (2014). Estudios del lenguaje. Niveles de representación lingüística. *Proyecto LATIn (Iniciativa Latinoamericana de Libros de Texto Abiertos)*. Disponible en [http://www.latinproject.org/books/Estudios\\_Lenguaje\\_CC\\_BY-SA\\_3.0.pdf](http://www.latinproject.org/books/Estudios_Lenguaje_CC_BY-SA_3.0.pdf).
- [5] Hernández, H. (1994). Algunos problemas ortológicos y ortográficos del español actual: sobre los grupos consonánticos *bl*, *br*, *dr*, *tl* y *pl*. En *Actas del IV Congreso de la Asociación para la Enseñanza del Español como Lengua Extranjera (ASELE)*. Disponible en: [http://cvc.cervantes.es/ensenanza/biblioteca\\_ele/asele/pdf/05/05\\_0171.pdf](http://cvc.cervantes.es/ensenanza/biblioteca_ele/asele/pdf/05/05_0171.pdf) [consulta 21 de junio de 2020).
- [6] RAE (2010). *Ortografía*. Espasa Libros, Madrid.
- [7] Real Academia Española y Asociación de Academias de la Lengua Española (2011). *Nueva gramática de la lengua española. Fonética y Fonología*. Espasa.
- [8] Rodrigo, A. y Bonino, R. (2019). *Aprendo con NooJ*. Ciudad Gótica, Rosario.
- [9] Silberztein, M. (2015). *La formalization des langues, L'Approche de NooJ*. Iste, London.
- [10] Tapia, S. M y Riestra, D. (2014). El momento de la reflexión sobre la lengua en el aula: ¿Explicación gramatical azarosa o sistematización de contenidos específicos? En *Saga*

*Revista de Letras*, 1:178–206. Disponible en:  
[https://rehip.unr.edu.ar/bitstream/handle/2133/11901/El%20momento%20de%20la%20reflexi%  
c3%b3n%20sobre%20la%20lengua%20en%20el%20aula.pdf?  
sequence=2&isAllowed=y](https://rehip.unr.edu.ar/bitstream/handle/2133/11901/El%20momento%20de%20la%20reflexi%c3%b3n%20sobre%20la%20lengua%20en%20el%20aula.pdf?sequence=2&isAllowed=y) [Consulta 21 de junio de 2020].

**Software NooJ**, disponible en: <http://www.nooj-association.org/downloads.html> [Consulta septiembre de 2020]

**Módulo Español Argentina**, disponible en: <http://www.nooj-association.org/resources.html> [Consulta septiembre de 2020].

**NooJ Tutorials en español**, disponible en: <http://www.nooj-association.org/tutorials.html> [Consulta septiembre de 2020].

# **Análisis automático de sintagmas preposicionales del español con el sistema NooJ**

**Carolina Caserio**

Facultad de Humanidades y Artes, Universidad Nacional de Rosario Rosario, Argentina  
carolinacaserio@gmail.com

**María Paula Garófolo**

Facultad de Humanidades y Artes, Universidad Nacional de Rosario Rosario, Argentina  
mariapaulagarofolo@hotmail.com

## **Abstract**

The prepositional syntagm is characterized by its heterogeneous composition. The recognition and the analysis require to be able to identify several structures in advance. The syntactic grammars of the NooJ system can identify many types of structures. In this paper, we focus on the syntagm preceded by a preposition. We will make an analysis from the traditional perspective and then move on to the computational one. The objective is to make syntactic grammars that recognize these types of syntagms in any type of text. To achieve the automatic analysis, we will create these grammars using the NooJ tool and we will test them in texts to check that the level of analysis achieved is comprehensive and accurate.

**Keywords:** prepositional syntagm, NooJ tool, syntactic grammar

## **Resumen**

El sintagma preposicional se caracteriza por ser heterogéneo en su composición, por lo que reconocerlo y analizarlo requiere de identificar varias estructuras previamente. Las gramáticas sintácticas del sistema NooJ pueden reconocer muchos tipos de estructuras. En este trabajo, nos focalizamos en el sintagma precedido por una preposición. Haremos un análisis desde la perspectiva tradicional para, luego, avanzar con la computacional. El objetivo es confeccionar gramáticas sintácticas que reconozcan estas clases de sintagmas en cualquier tipo de texto. Para lograr el análisis automático, crearemos estas gramáticas a través de la herramienta NooJ y las probaremos en textos a fin de comprobar que el nivel de análisis que arroja es exhaustivo y preciso.

**Palabras claves:** sintagma preposicional, herramienta NooJ, gramática sintáctica

## 1. INTRODUCCIÓN

En el presente trabajo, nos proponemos analizar el sintagma preposicional desde dos perspectivas: la lingüística tradicional y la computacional. Con este objetivo, realizaremos una descripción sobre las preposiciones y su funcionamiento como sintagmas gramaticales para poder crear gramáticas a través del sistema NooJ que nos permita analizar esta estructura en distintos textos. El propósito es comprobar la efectividad de la herramienta a la hora de llevar un análisis gramatical de alta precisión. Se deja para más adelante la aplicación de estas gramáticas sintácticas en textos de aprendiente de español como segunda lengua.

## 2. DESARROLLO DESDE LA LINGÜÍSTICA TRADICIONAL

Las preposiciones fueron denominadas como “partículas de enlace” por la gramática tradicional. Sin embargo, no solo su función es enlazar, sino que organizan la estructura de los sintagmas como piezas básicas sobre las que se construyen las unidades mayores.

Según Di Tullio (2014), se caracterizan por ser palabras funcionales carentes de todo significado léxico. Algunas determinan semánticamente el valor del sintagma preposicional, como *ante*, *sobre*, *bajo*; nunca son seleccionadas por un núcleo, permiten la alternancia con otras preposiciones o el sintagma preposicional (SP) en su conjunto con adverbios del mismo tipo.

La preposición también se caracteriza por requerir un término. Por lo general, se define la preposición como una palabra relacional que vincula un sintagma nominal con otra construcción. Si bien el término de la preposición es normalmente un sintagma nominal (SN), puede ser también un sintagma adjetival (SA), “aspecto de ochentoso”; un sintagma adverbial (Sadv), “desde muy lejos”, un SP, “por sobre la mesa”, o una cláusula (= SCOMP), “hasta que vuelvas”.

Como explica Di Tullio, el español es una lengua de núcleo inicial: los complementos siguen al verbo, al sustantivo o al adjetivo que los selecciona. Lo mismo ocurre con el término del SP: preposición y término se encuentran en una relación de fraternidad estructural y de adyacencia estricta. El núcleo solo está capacitado para ejercer la reacción cuando precede a su complemento.

Las locuciones prepositivas permiten una expansión mínima:

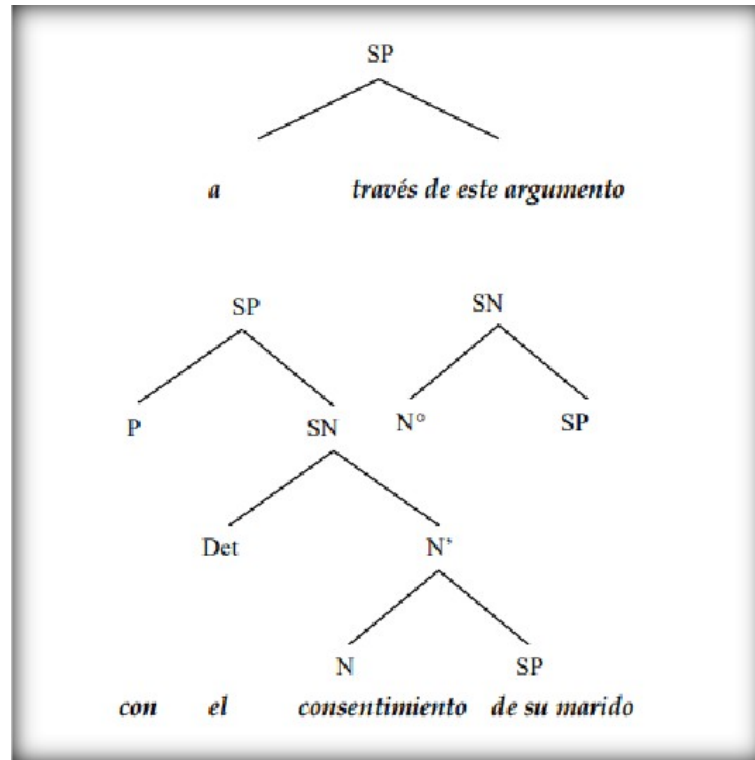


Figura 1: Responde a lo que Di Tullio llama “la estructura interna” cuyo fin principal es determinar cómo está conformada una construcción a partir de sus constituyentes.

Si decimos “el consentimiento”, nos encontramos con una construcción que consta de los mismos constituyentes inmediatos: determinante + sustantivo o construcción nominal. En cambio, “con el consentimiento de su marido”, tiene una estructura interna diferente: es un sintagma preposicional como “desde el jardín, con su permiso”. Los constituyentes de cada una de estas construcciones son la preposición, por un lado, y el sintagma nominal, por otro.

Di Tullio plantea, además, que la noción de estructura interna sirve también para caracterizar los distintos tipos de unidades intermedias que pueden reconocerse. Según la autora, es la preposición la que determina las características de la construcción en su conjunto.

Ignacio Bosque (2008) esquematiza las principales funciones sintácticas que pueden aparecer. En esta oportunidad, haremos hincapié en algunas de las *determinaciones de la preposición* que encabezan un sintagma u oración.

2.a. **Hacia** o **hasta**: dirección. “Estuvimos festejando **hasta** la medianoche”.

2.b. **Desde**: el origen. “**Desde** agosto que no tengo noticias”.



2.c. **Bajo, sobre, ante, entre:** son marcadores impuros porque, además de permitir la relación con el término, tienen un valor lexical.

**Sobre, bajo:** locativo. “El dinero está **bajo** la cama”. “**Ante** la adversidad, sé más fuerte”. “El secreto es **entre** vos y yo”.

2.d. **De:** atributo, cualidad o del sujeto o propiedades físicas. “Ana es **de** Córdoba”. “Ministro **de** Economía”.

Complemento regido. “Elena es guapa **de** verdad”. Completa el significado de “guapa”.

Complemento adjetival. “El bueno **de** tu tío nos llevó al cine”.

2.e. **A:** con referencia a una persona, complemento directo: “Quiero **a** mi madre.”

Complemento indirecto: “Trajo un arco **a** su hijo”.

Complemento regido: “José aspiraba **a** ese puesto”.

2.f. **En:** complemento circunstancial (de lugar). “Viven **en** La Plata”.

2.g. **Por:** agente de la pasiva. “El debate fue rechazado **por** el presidente”.

2.h. **Para:** finalidad, relativa a personas, lugares. “**Para** aprobar, debes estudiar”.

En este sentido, podemos observar que el sintagma preposicional es simple, pero, a su vez, es muy heterogéneo en su constitución ya que puede ir con nombres, adjetivos y adverbios. La preposición es un elemento gramatical que siempre actúa como nexo entre palabras, sintagmas u oraciones, pero no tiene la capacidad de convertirse en núcleo semántico de ninguna unidad gramatical. Asignan el papel temático a sus complementos, ya sea de manera aislada o con otra pieza léxica. Así vemos, al igual que en los casos listados anteriormente, el ejemplo de la preposición *sobre* que indica un complemento con significado de lugar. Este tipo de sintagma puede aparecer dentro del sujeto, como complemento del nombre, en el predicado, como complemento del verbo y, como complemento del nombre, en algún subconjunto sintagmático del predicado.

De esta manera, podemos concluir que, en concreto, el sintagma preposicional es aquel cuyo núcleo es una preposición que asigna caso a un sintagma nominal, siendo este caso el más frecuente. A continuación, lo analizaremos a través de la herramienta NooJ que nos permitirá crear gramáticas que podrán reconocer estas unidades en diferentes textos y representar así su estructura.

### 3. DESARROLLO DESDE LA LINGÜÍSTICA COMPUTACIONAL

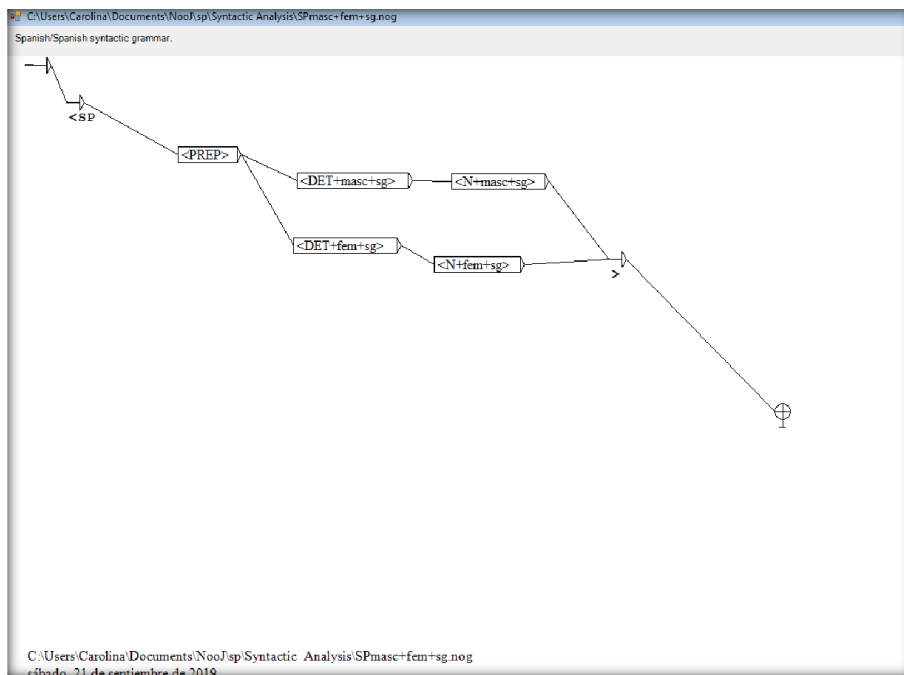


Figura 2: Sintagma preposicional internamente articulado con el sintagma nominal con determinante masculino y femenino singular.

A continuación, comprobamos, a través de la función *Show Debug*, que la gramática creada puede reconocer los siguientes sintagmas:

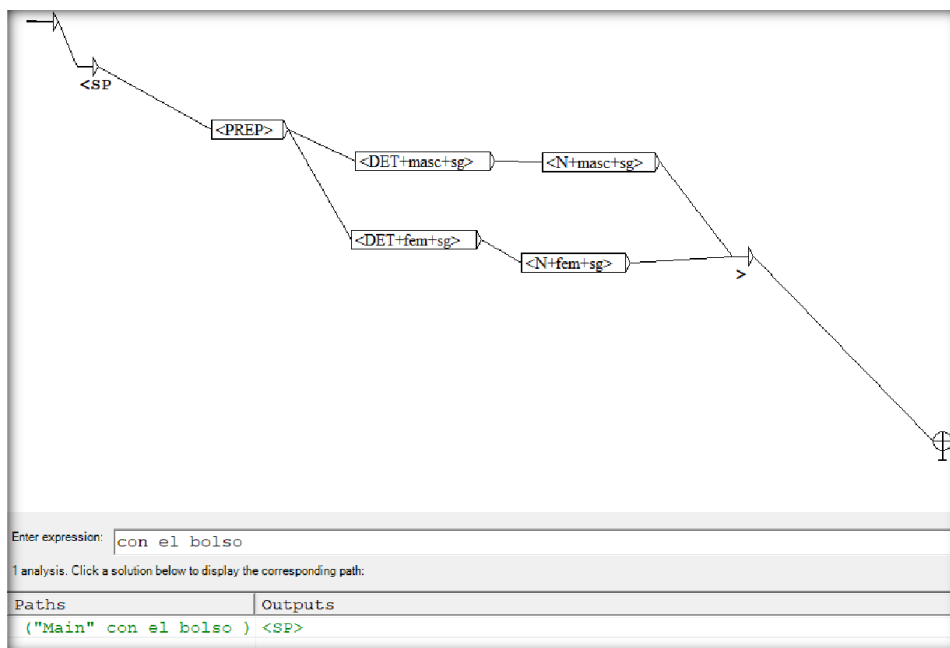


Figura 3:  $SP \rightarrow P + SN \text{ masc (Det + N)}$

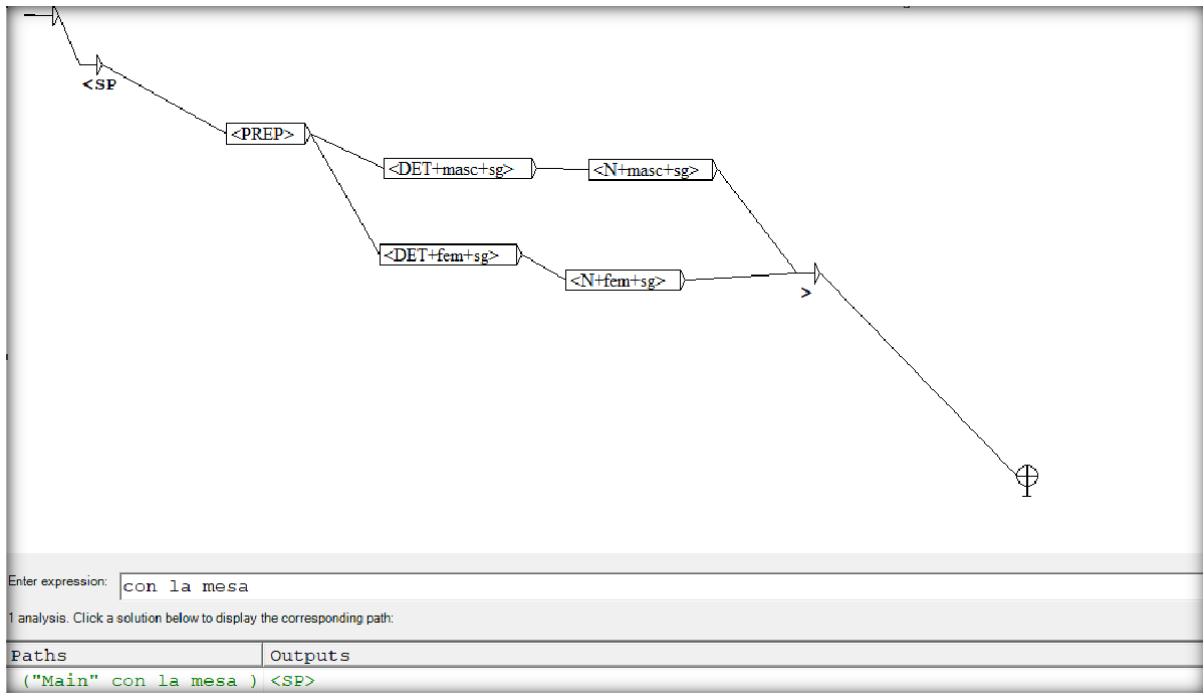


Figura 4: SP → P + SN fem (Det + N)

Luego, analizamos un texto utilizando la gramática creada para verificar que identifique los sintagmas en contexto correctamente. El texto literario analizado es *La muchacha del atado*, de Roberto Arlt.

Flacas, angustiosas, sufridas. El polvo de arroz no alcanza a cubrir las gargantas donde se marcan los tendones; y todas caminan con el cuerpo inclinado a un costado: la costumbre de llevar el atado siempre del brazo opuesto. Y los bultos son macizos, pesados: dan la sensación de contener plomo: de tal manera tensionan la mano. No se trata de hacer sentimentalismo barato. No. Pero más de una vez me he quedado pensando en estas vidas, casi absolutamente dedicadas al trabajo. Y si no, veamos. Cuando estas muchachas cumplieron ocho o nueve años, tuvieron que cargar un hermanito en los brazos. Usted, como yo, debe haber visto en el arrabal estas mocosas que cargan un pebetito en el brazo y que se pasean por la vereda y vigiladas por la madre que salpicaba agua en la batea. Así hasta los catorce años. Luego, el trabajo de ir a buscar costuras; las mañanas y las tardes inclinadas sobre la Neumann o la Singer, haciendo pasar todos los días metros y más metros de tela y terminando a las cuatro de la tarde, pa vestido de percal, preparar el paquete y salir, salir cargadas y volver lo mismo, con otro bulto que hay que "pasarlo a la máquina". La madre siempre lava la ropa; la ropa de los hijos, la ropa del padre. Y ésas son las muchachas que los escuchan la voz del hermano, que grita: -Che, Angelita: apurate a plancharme la camisa, que tengo que salir. Y Angelita, María o Juana, la tarde del sábado trabajan para los hermanos. Y planchan cantando un tango que aprendieron de memoria en El Alma que Canta; que esto, las novelas por entregas y alguna sección de biógrafo, es la única que hablo. Digo que estas muchachas me dan lástima. Un buen día se ponen de novias, y no por eso dejan de trabajar, sino que el novio (también un muchacho que la yuga todo el día) cae a la noche a la casa a hacerle el amor. Y como el amor no sirve para pagar la libreta del almacén, trabajan hasta tres días antes de casarse, y el casamiento no es un cambio de vida para la mujer de nuestro ambiente pobre, no; al contrario, es un aumento de trabajo, y a la vez a estas muchachas sobre la máquina. Han vuelto a la costura. Y al año han un rubro en la cuna. Y esa muchacha ya está armada y escribiendo ahora tiene que trabajar para el hijo, para el marido, para la casa. Cada año un nuevo hijo.

115	121	129	133	136
todo,CUANT+clase=def+género=fem+número=pl	caminar,V+tiempo=pte+modo=ind+persona=3era+número=pl	SP		
		con,PREP	el,DET+clase=artdet+género=masc+número=sg	cuerpo,N+género=masc+número=sg

Figura 5: SP → P + SN masc (Det + N) + sg: “con el cuerpo”

624	632	635	638
existir,V+tiempo=pi+modo=ind+persona=1era+número=sg	SP		
existir,V+tiempo=pi+modo=ind+persona=3era+número=sg	en,PREP	la,DET+clase=artdet+género=fem+número=sg	casar,V+tiempo=pte+modo=ind+persona=3era+número=sg
		la,CL+clase=cl3+persona=3era+número=sg+género=fem+caso=aq	casar,V+modo=imp+persona=2da+número=sg
			casa,N+género=fem+número=sg

Figura 6: SP → P + SN fem (Det + N) + sg: “en la casa”

Text	Before	Seq.	After
Si, aunque marchó aunque marchó por el valle escribe, habré entrado hace mucho a los caracteres aquí grabados los cuales no hay nombre lejos y en todas partes, se cernían las negras alas ya había llegado la alternación me equivoco, el especial espíritu cielo no sólo se manifestaba manifestaba en el globo físico tierra, sino en las almas, imaginación y en las meditaciones el presagio y el recuerdo		por el valle de la Sombra en la región con un estilo sobre la tierra sobre el mar de la peste de aquel año del cielo en el globo de la tierra en la imaginación de la humanidad del Mal	de la Sombra. (Salmo de . (Salmo de David, XXIII) Vosotros de las sombras. Pues en de hierro. El año había . Pues habían ocurrido muchos prodigios y la tierra, se cernían . Para aquellos versados en la 794, en el cual, a la no sólo se manifestaba en físico de la tierra, sino , sino en las almas, en y en las meditaciones de . En una sombría ciudad llamada no podían ser excluidos.

Figura 7: Por medio de la función *Locate*, se reconocen todos los sintagmas preposicionales masculinos y femeninos singulares que aparecen en el texto elegido: *Sombra*, de Edgar Allan Poe.

En función de la gramática creada anteriormente, desarrollamos una nueva que analiza el sintagma preposicional internamente articulado con el sintagma nominal con determinante masculino y femenino plural.

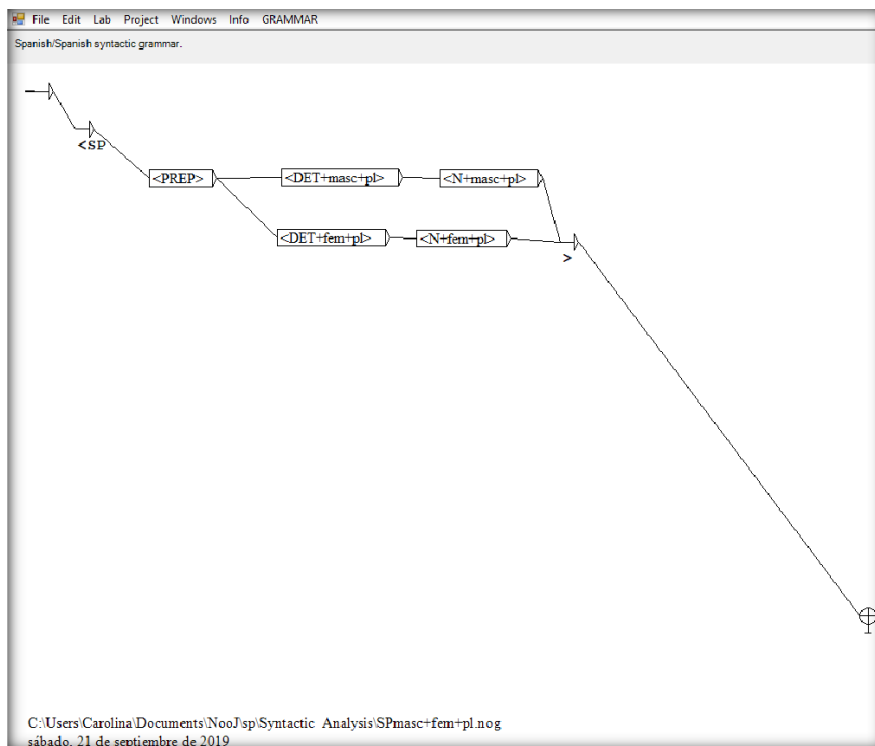


Figura 8: Sintagma preposicional internamente articulado con el sintagma nominal con determinante masculino y femenino plural.

Una vez más, mediante el análisis lingüístico de un texto, observamos que la gramática creada reconoce sintagmas preposicionales articulados con sintagmas nominales femeninos y masculinos plurales.

Show Text Annotation Structure		830 word forms 2 digits	
Si, aunque marchó por el valle de la Sombra. (Salmo de David, XXIII)			
Vosotros los que leéis aún estáis entre los vivos; pero yo, el que escribe, habré entrado hace mucho en la región de las sombras. Pues en verdad ocurrirán muchas cosas, y se escrito. Y, cuando lo hayan visto, habrá quienes no crean en él, y otros dudarán, mas unos pocos habrá que encuentren razones para meditar frente a los caracteres aquí grabados.			
El año había sido un año de terror y de sentimientos más intensos que el terror, para los cuales no hay nombre sobre la tierra. Pues habían ocurrido muchos prodigios y señales de peste. Para aquellos versados en la ciencia de las estrellas, los cielos revelaban una faz siniestra; y para mí, el griego Ómos, entre otros, era evidente que ya había llegado la conjunción con el anillo rojo del terrible Saturno. Si mucho no me equivoco, el especial espíritu del cielo no sólo se manifestaba en el globo físico de la tierra, sino en las almas de los hombres.			
En una sombría ciudad llamada Ptolemáis, en un noble palacio, nos hallábamos una noche siete de nosotros frente a los frascos del rojo vino de Chios. Y no había otra entrada al artesano Corinnos, y, por ser de raro mérito, se la aseguraba desde dentro. En el sombrío aposento, negras colgaduras alejaban de nuestra vista la luna, las cárdenas estrellas.			
105	112	114	118
frente, N+género=fem+número=sg	SP		
frente, ADV	a, PREP	los, DET+clase=artdet+género=masc+número=pl	frasco, N+género=masc+número=pl
		los, CL+clase=cl3+persona=3era+número=pl+género=masc+caso=acc	

Figura 9: SP → P + SN masc (Det + N) + pl: “a los frascos”

A través de la función *Locate*, reconocemos todos los sintagmas preposicionales masculinos y femeninos plurales que figuran en este mismo texto (*Sombra*, de Edgar Allan Poe):

Text	Before	Seq.	After
	<p>los que leéis aún estáis hace mucho en la región aquellos versados en la ciencia físico de la tierra, sino almas, en la imaginación y noche siete de nosotros frente</p>	<p>entre los vivos de las sombras de las estrellas en las almas en las meditaciones a los frascos</p>	<p>; pero yo, el que escribe . Pues en verdad ocurrirán muchas , los cielos revelaban una faz , en la imaginación y en de la humanidad. En una del rojo vino de Chios</p>

Figura 10: SP → P + SN masc (Det + N) + pl: “a los frascos”; “entre los vivos”

SP → P + SN fem (Det + N) + pl: “de las sombras”, “de las estrellas”, “en las almas”, “en las meditaciones”

NooJ permite también crear gramáticas más complejas que combinen distintas variaciones de un mismo sintagma. A continuación, desarrollamos una gramática que analiza sintagmas preposicionales con posesivos tanto masculinos como femeninos en singular y plural.

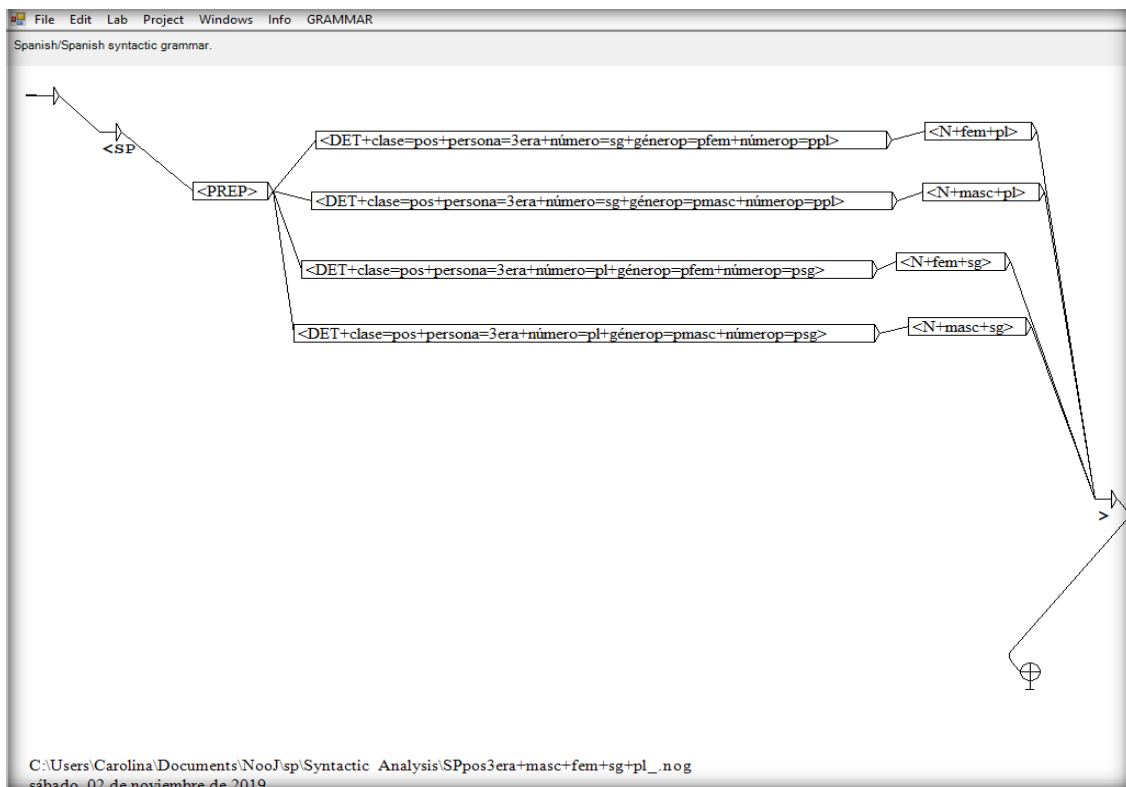


Figura 11: Gramática que analiza sintagmas preposicionales con posesivos tanto masculinos como femeninos en singular y plural.

Del mismo modo, mediante el análisis lingüístico de un texto y la función *Locate*, observamos que la gramática creada reconoce sintagmas preposicionales articulados con sintagmas nominales femeninos y masculinos, singulares y plurales cuyo determinante es un posesivo. El texto elegido es *El corazón delator*, de Edgar Allan Poe.

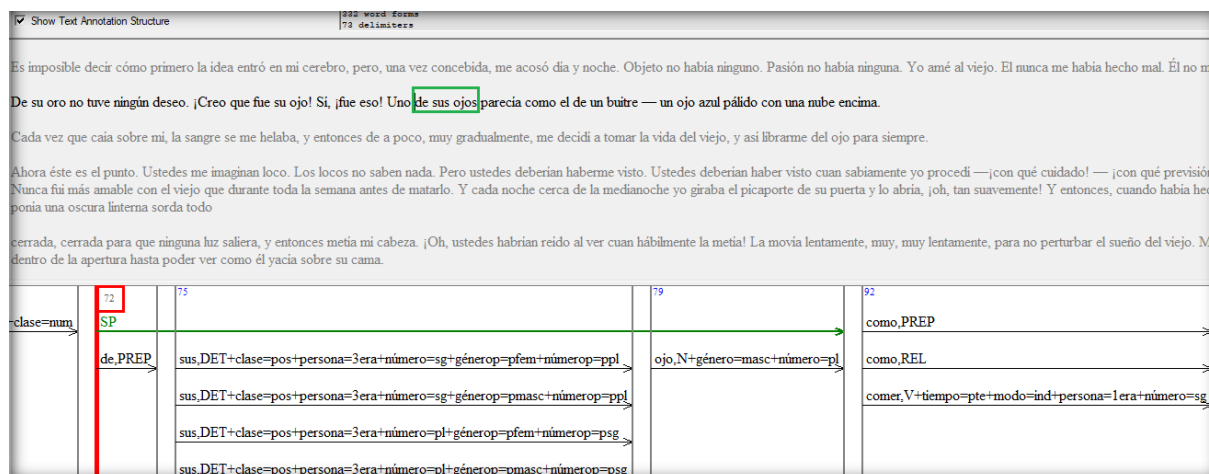


Figura 12: Sintagmas preposicionales articulados con sintagmas nominales femeninos y masculinos, singulares y plurales cuyo determinante es un posesivo.

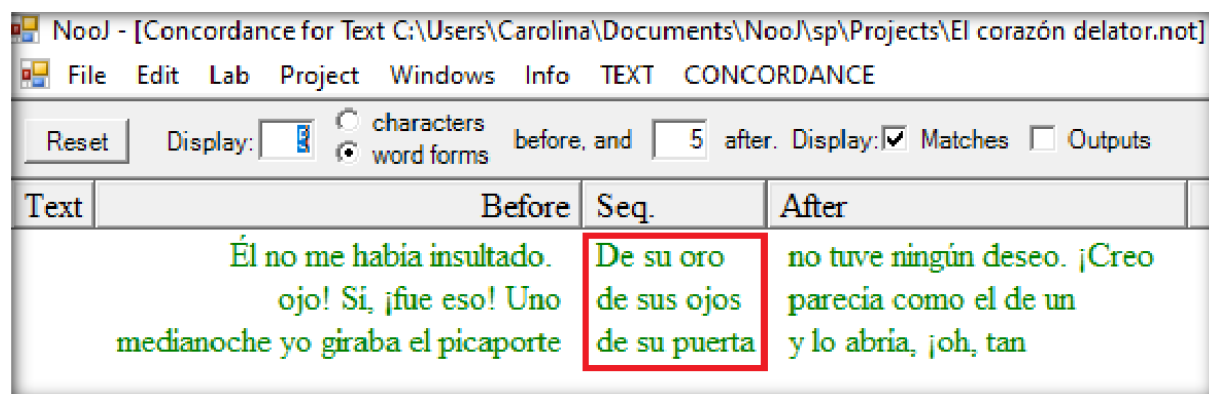


Figura 13: SP → P + SN masc (DETpos + sg): “de su oro” SP → P + SN fem (DETpos + sg): “de su puerta”  
 SP → P + SN masc (DETpos + pl): “de sus ojos”

## 4. CONCLUSIÓN

NooJ es una herramienta informática que crea recursos cuyo objetivo es formalizar fenómenos lingüísticos: tipografía, ortografía, morfología derivacional e inflexional, sintaxis y semántica y que produce analizadores sintácticos capaces de procesar cualquier recurso lingüístico y, a su vez, aplicarlo a un corpus para extraer ejemplos, estructuras correspondientes y presentar análisis estadísticos.

Si bien el presente trabajo se focalizó en la creación de gramáticas que analizan sintagmas preposicionales internamente articulados con otros sintagmas, entendemos que nos encontramos ante una herramienta en proceso de formación que colabora activamente en la tarea de la adquisición y la enseñanza de una L2. Familiarizarnos con esta aplicación nos permitió desarrollar un nivel de análisis gramatical de alta precisión que podemos aplicar tanto en inglés, nuestra segunda lengua, como en español, la segunda lengua que enseñaremos.



## Referencias

- [1] Silberztein, M. (2003). *NooJ Manual*. disponible en <http://www.NooJassociation.org/media/k2/attachments/app/NooJManual.pdf>.
- [2] Di Tullio, Á. (2010). El sintagma adverbial y el sintagma preposicional, en *Manual de gramática del español*. Waldhuter Editores, Buenos Aires.
- [3] Bosque, I. y Gutiérrez-Rexach, J. (2009). Los sintagmas, en *Fundamentos de sintaxis formal*. Akal, Móstoles, Madrid.
- [4] Arlt, R. (1933). La muchacha del atado, en *Aguafuertes porteñas*. Victoria, Buenos Aires.
- [5] Allan Poe. E. (1835). *Sombra*. Southern Literary Messenger, Estados Unidos.  
(1843) *El corazón delator*. James Russell Lowell, Estados Unidos.
- [6] Solana, Z., Beltrán, C. y Tramallino, C. (2013). Análisis automático morfológico con las herramientas SMORPH y NOOJ. *Revista de epistemología y ciencias humanas*, grupo IANUS. ISSN 1852-625X, N.º 5.
- [7] Tramallino, C. (2013). Análisis morfológico con herramientas informáticas. Reconocimiento de nombres en textos de español con el sistema NooJ. *Revista Lingüística y Literatura*. Universidad de Antioquía, Medellín, Colombia, ISSN 0120-5587, 34-64.
- [8] Jiménez López, M.D. (2019). Modelos computacionales de la adquisición y complejidad lingüística. *E-Aesla*. Universidad Rovira i Virigili, 5.

## **CAPÍTULO II**

### **Adquisición y enseñanza de lenguas**

# **Adquisición del wichí: hacia una metodología para el estudio del lenguaje infantil en una lengua nativa argentina**

**Andrea S. Taverna**

Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)  
Instituto Rosario de Investigaciones en Ciencias de la Educación (IRICE)  
Rosario, Argentina  
andtaverna@gmail.com

**Migdalia I. Padilla**

Consejo Nacional de Investigaciones Científicas y Técnicas (CONICET)  
Instituto Rosario de Investigaciones en Ciencias de la Educación (IRICE)  
Rosario, Argentina

## **Resumen**

Este artículo presenta la metodología de una investigación en curso sobre lenguaje infantil en una lengua nativa de nuestro país: el wichí. Al tratarse del primer corpus lingüístico del habla infantil en esta lengua originaria, este trabajo describe su diseño metodológico, el que combina técnicas psicolingüísticas y etnográficas de recolección, registro y análisis del habla infantil y adulta wichí y de otras semiosis comunicativas (pauta de mirada, tacto, kinesis) que intervienen en las interacciones materno-infantiles de esta cultura. Bajo una sólida perspectiva del desarrollo, se discuten los recaudos metodológicos que se han implementado en el proceso investigativo y se ofrece un resumen de los resultados que se han obtenido hasta ahora respecto de dos focos de análisis interrelacionados: a) adquisición léxica, incluyendo forma de las primeras palabras y su organización semántica; b) prácticas de socialización tempranas que sirven de contexto a la emergencia del lenguaje infantil wichí.

**Palabras claves:** adquisición del lenguaje, corpus lingüístico, reporte parental, psicolingüística, wichí

## **Abstract**

This article presents the methodology of an ongoing research about infant language, specifically of our country's native language: the Wichi. Since this is the first linguistic corpus of infant speech, our research combines psycholinguistic and ethnographic techniques for collecting, registering and analyzing Wichi children's and adult speech and other communicative bodily semioses (gazing, touching, kinetic movement) relevant in the maternal-infant interactions of this culture. Within a strong developmental perspective, we discuss the methodological precautions that have been implemented in the research process and a summary of the results that have been obtained so far regarding two interrelated analysis issues: a) lexical acquisition, including the form of the first words and their semantic organization; b) early socialization practices that serve as the context for the emergence of the Wichi child language.

**Keywords:** language acquisition, linguistic corpus, parental report, psycholinguistics, Wichi

## 1. INTRODUCCION

Durante décadas, la antropología y la lingüística han sido las disciplinas mayores que han puesto su atención en las comunidades indígenas del monte chaqueño argentino - entre ellas los wichí - produciendo innumerables investigaciones sobre temas culturales, lingüísticos, de política educativa, ecológicos, sociales, jurídicos, etnobotánicos, religiosos, etc. Llama la atención, sin embargo, la escasa presencia de estudios sobre la infancia. Y si bien, existen estudios que involucran aspectos de la educación que reciben estos niños y niñas [3] [4] [19] [28] [31] [41] en general esos estudios se centran en características del sistema educativo formal y, por ende, en niños y niñas que están transitando la escolaridad.

Por otra parte, en la investigación en psicología o psicolingüística del desarrollo hay una ausencia en la documentación sobre la adquisición, en la infancia, de la gran variedad de lenguas indígenas que hay en esta zona de nuestro país, ausencia que podemos notar también en el mapa de la investigación nacional (aunque ver proyecto longitudinal en niños y niñas hispanohablantes y hablantes Qom reciente [1][21][27]). Es probable, entonces, que casi como una circularidad irremediable, la escasez de empeños científicos en psicolingüística- y en particular de las lenguas de los pueblos indígenas (muchas de ellas – como el wichí - en peligro de extinción) - sea lo que oficie de obstáculo para el diseño de nuevos estudios de esta naturaleza, sumado, posiblemente, a la intensidad que se le plantea al investigador que lleva adelante un proyecto longitudinal de estas características.

A contrapelo de esta tendencia, desde hace casi una década llevamos adelante el estudio de la adquisición de una lengua indígena de la Argentina - el wichí-. Específicamente, esta investigación aborda la adquisición del léxico, incluyendo la forma de las primeras palabras, su organización semántica y las prácticas de socialización en las que los infantes wichí se involucran cuando adquieren esta lengua ancestral como lengua materna [37] [34]. Como toda intensa investigación exploratoria que cuenta no con pocos sino nulos antecedentes, este proceso ha labrado un abanico de decisiones metodológicas, las que satisfactoriamente hicieron posible diseñar el estudio de la adquisición del wichí y obtener un corpus inédito del lenguaje e interacciones materno-infantiles en esta lengua. El presente artículo describe dicha metodología, la que al combinar técnicas de la psicolingüística tradicional y la etnografía, se

aleja de las clásicas metodologías de estudios adquisicionistas para estudiar el lenguaje infantil en una lengua nativa inédita.

A continuación, primero, se presenta una breve caracterización de la lengua wichí, la que ha sido descripta recientemente [22] para luego desarrollar la metodología de trabajo de este proyecto psicolingüístico, translingüístico y colaborativo.

## **2. LOS WICHI Y SU LENGUA**

El pueblo wichí, anteriormente conocido como "Mataco", es una población indígena que vive en el monte chaqueño, en las tierras bajas de América del Sur en Argentina. El monte chaqueño se extiende por 1.000 000 kilómetros cuadrados; su terreno se caracteriza principalmente por llanuras cubiertas de hierba, intercaladas con distintas áreas dominadas por matorral, pequeñas plantas leñosas o palmerales. Los wichí dependen económicamente de actividades tradicionales, como la caza, la pesca, la recolección y prácticas hortícolas estacionales de "tala y quema". También son prácticas económicas clave el tejido textil, la cerámica, y, en menor medida, la ganadería y la agricultura (para mayor detalle de las características de la comunidad, ver [2][37] [38] [39][35]).

La lengua wichí pertenece a la familia de lenguas Mataguaya y se habla en las tierras bajas de la región del Gran Chaco en América del Sur. La distribución geográfica de la lengua wichí comprende dos países: Argentina (en las provincias de Formosa, Salta y Chaco a lo largo de los ríos Bermejo y Pilcomayo) y Bolivia. Dentro de las dos amplias variedades del Wichí llamadas pilcomayeño y bermejeño, según si los hablantes viven en los ríos Pilcomayo o Bermejo, la segunda es la variedad que hablan los participantes de esta investigación.

En la región del Gran Chaco donde esta investigación tiene lugar, particularmente en las localidades de Laguna Yema y Chiriguano en Formosa, el wichí se adquiere como primera lengua. Los niños y niñas son introducidos por primera vez al español a los cinco o seis años de edad cuando ingresan a la escuela pública. El español está reservado principalmente para el contexto escolar el que se habla principalmente en interacción con docentes no wichí y otras autoridades escolares. Por lo tanto, dado que el wichí es la lengua de uso exclusivo dentro de la familia y en un contexto más amplio de la comunidad, el ingreso temprano de los infantes al lenguaje refleja aquel que transitan otros infantes que adquieren una sola lengua materna.

Tipológicamente, el wichí es una lengua morfológicamente compleja, aglutinante, con tendencia a la polisíntesis (para una mayor descripción de esta noción ver [9] y para una descripción de la gramática wichí ver [22]) por lo que la misma se sirve de la afijación para expresar la mayoría de las categorías gramaticales tanto verbales como nominales. Esta complejidad morfológica implica que una sola palabra – especialmente si es un verbo – es suficiente para tener lo que en otras lenguas (como el español) se considera una oración o un enunciado gramatical y completo. Los argumentos se indexan en el verbo siguiendo una alineación nominativa-acusativa en la primera y segunda persona y una alineación neutral en la tercera persona (cero para el sujeto y el objeto). Las frases nominales son opcionales, el orden básico de las palabras puede ser VO, SVO. Si bien, las principales clases de palabras son sustantivos y verbos (no hay adjetivos ya que las propiedades, atributos y estados son expresados por un grupo de verbos intransitivos), el verbo es morfológicamente la clase de palabra más prominente. El verbo wichí codifica 15 categorías morfológicas diferentes<sup>1</sup> y hasta 5 de ellas pueden co-ocurrir. El sustantivo, por su parte codifica 8 categorías<sup>2</sup>, de las cuales una de ellas, la posesión, es muy relevante en esta lengua, dividiendo a los sustantivos en dos clases principales: sustantivos alienables vs inalienables. Los sustantivos alienables no tienen un poseedor inherente, incluyendo palabras que típicamente refieren a artefactos, animales y plantas. En cambio, los sustantivos inalienables tienen un poseedor inherente, incluidos los nombres que refieren a las relaciones de parentesco (por ejemplo, madre), las partes del cuerpo (por ejemplo, nariz, brazo), entre otros. Esta distinción alienable / inalienable está marcada en el habla escuchada (*input* lingüístico) que recibe el niño o niña de muchas maneras. Por ejemplo, los sustantivos inalienables (como 'mami' o 'nariz') no se pueden producir solos; en cambio, requieren un prefijo de persona poseedor para indicar la relación posesiva del hablante con el referente. Por ejemplo, *n'ku* [*n'-ku*] 'mi madre'; *aku* [*a-ku*] 'tu madre', la raíz nominal inalienable (ku o madre) aparece de manera consistente, pero toma un prefijo de poseedor diferente según la relación entre el hablante y el objeto de

---

1 Incluyendo persona (sujeto y objeto), número, tiempo, modo, aspecto, reflexivo / recíproco, causativo, aplicativos, instrumento, modalidad / evidencia, direccional, locativo, negación y manera / diminutivo. La cantidad y variedad de categorías morfológicas que admite el verbo (aunque no coinciden todas a la vez) es aparentemente una característica extendida entre las lenguas del Gran [20].

2 Los sustantivos admiten las siguientes categorías: posesión, número, deixis, clasificadores, derivados, aumentativo y diminutivo, tiempo (futuro y no futuro), distributivo (marcador de aspecto verbal) y negación (en sustantivos poseídos).

referencia. Los sustantivos alienables, por su parte, los que a diferencia de los inalienables están constituidos por entidades no inherentemente poseídas, indexicalizan la persona poseedor con los mismos prefijos que los sustantivos inalienables, pero a diferencia de estos últimos, interponen un prefijo o clasificador nominal [40] entre el poseedor y lo poseído. El clasificador adopta dos formas, *ka* combinado con sustantivos que denotan objetos o herramientas útiles y *lo* que se combina con animales domesticables (para una descripción completa ver [22]).

En su conjunto, los wichí son una población nativa que nos ofrece la oportunidad de considerar la adquisición del lenguaje en una comunidad ancestral aún no explorada, con una fuerte lengua vernácula de gran vitalidad y constelación de costumbres que difieren de manera importante de la mayoría de las poblaciones indígenas que se han estudiados a la fecha. Particularmente, la lengua wichí, que se adquiere como lengua materna, representa un tipo de lengua en el que los verbos ocupan una posición privilegiada, caracterizada por una morfología compleja y rica tanto para sustantivos y verbos. Por lo tanto, el estudio de la adquisición del lenguaje en wichí tiene el potencial de arrojar luz sobre cómo las características de la lengua nativa y su contexto cultural de adquisición da forma a este proceso a medida que se desarrolla.

### **3. EL ESTUDIO**

La investigación central fue realizada en las comunidad Wichí Lawet en la localidad de Laguna Yema, comunidad con 3600 habitantes aproximadamente [16] ubicada sobre la ruta nacional 81 en la provincia de Formosa (Fig. 1). En esta comunidad, el wichí es hablado como primera lengua en todos los hogares. El trabajo principal se ha enfocado en 8 familias de esa comunidad quienes se ofrecieron a participar de esta investigación en virtud de los lazos de confianza establecidos con ellas a partir de los estudios previos realizados con niños y niñas de esta comunidad (varios de ellos miembros de estas familias) en años anteriores en temas relacionados con el desarrollo cognitivo y el aprendizaje [2] [25] [35][37] [38] [39]. Puesto que se trata de una comunidad de habla no explorada, para el estudio de la adquisición del wichí consideramos dos ejes fundamentales que se corresponden en un todo con la perspectiva constructivista-emergentista del desarrollo[17]<sup>3</sup> sobre la que descansa el proyecto:

---

3 Esta perspectiva implica que el proceso de adquisición del lenguaje es un proceso complejo y recursivo, en el que interactúan entidades lingüísticas con propiedades emergentes, de tal manera que sus interacciones dan lugar



a) la emergencia y desarrollo de las primeras palabras hasta el inicio de las primeras combinatorias léxicas y b) el contexto interaccional y comunicacional-intencional que se despliega durante dicha etapa psicolingüística en el seno de la relación infante-cuidador.



Figura 1: Localización de Laguna Yema en provincia de Formosa, Argentina

### 3.1. Método psicolingüístico: Reporte parental para la elaboración del primer inventario del léxico infantil wichí

Al tratarse de una lengua no explorada desde el punto de vista de su adquisición, se decidió, en primer lugar, crear un reporte parental en lengua wichí. Los inventarios de desarrollo comunicativo (CDI, *check lists*, etc.) son muy provechosos como herramienta investigativa en comunidades de habla cuyos niños y niñas no están acostumbrados a interactuar con adultos extraños o no pertenecientes a dicha comunidad. Este es precisamente el caso de los niños y niñas de la comunidad Wichí Lawet. En primer lugar, el proyecto se abocó al desarrollo de un inventario que permitiera estudiar el desarrollo del vocabulario de los niños y niñas en estas comunidades, que no sólo proporcionara un punto de comparación con las investigaciones inter-lingüísticas existentes (en español [26][37]), sino que sea pensado específicamente para reflejar las distintas características de la comunidad de estudio y su lengua [37].

---

a nuevas entidades emergentes de nivel superior, con nuevas propiedades emergentes, y así sucesivamente [12].

Dedicamos casi un año completo al desarrollo del vocabulario en lengua wichí. Primero, traducimos la versión en español del Inventario de Desarrollo Comunicativo de MacArthur (MCDI) [17][20] [26] al wichí. En este proceso, consideramos tanto a) las categorías comunicativas del MCDI, y b) las palabras dentro de ellas. En cuanto a las categorías comunicativas, conservamos las del MCDI español, con algunas excepciones. Para adaptarse a la experiencia de los niños y niñas wichí, omitimos tres categorías por completo (‘efectos de sonidos y sonidos de animales’; ‘vehículos’; ‘juguetes’) ya que el wichí no incluye palabras en estas categorías. También omitimos otras palabras en español que no tienen contrapartida en esta lengua ("artículos para el hogar"; "muebles / sala"; "fuera de las cosas"; "lugares para ir"). En el caso de "partes del cuerpo", "acciones y procesos", "propiedades" realizamos adaptaciones gramaticales.

En segundo lugar, convocamos a un grupo focal con tres madres de habla nativa. Nuestro objetivo fue afinar la lista de vocabulario para asegurarnos de haber identificado correctamente la asignación de categorías gramaticales de cada palabra. Todas las sugerencias hechas en estas reuniones fueron revisadas antes de finalizar el instrumento. El reporte maternal en lengua wichí en su versión final incluyó 340 palabras wichí (168 sustantivos, 119 verbos). Véase la Tabla 1 en donde se muestra un fragmento de este inventario. Para su administración, debido a que la mayoría de las madres en nuestra muestra (y en esta comunidad) no leen, una hablante nativa entrenada leyó cada palabra de la lista en voz alta, pidiéndole a la madre que indicara si su hijo: (a) comprende la palabra, (b) produce la palabra, y si fuera así, (c) cómo la pronunciaba. El inventario de vocabulario en lengua wichí fue administrado a 22 madres de niños y niñas de 12 a 47 meses (Véase [37] para consultar el inventario de palabras wichí completo).

Tabla 1: Fragmento del primer inventario del léxico infantil wichí con sus correspondientes equivalencias léxicas en español

WICHÍ	ESPAÑOL
<b>Tshotoy</b>	<b>Animales</b>
afwenche	pájaro
alhetaj	yacaré
Asinoj	perro
Asnu	burro
Awetsaj	jabalí
chasn'u	quirquincho
chilay'i	palomita

chitan'i	tortuga
Chukuk	mariposa
fwotsajh	víbora
hu'u	gallina
hu'ulhos	pollito
Hukwinataj	paloma
Istiwín	águila
Kayla	chiva
kolunp'un	abeja
Kutsi	cerdo
Matsitaj	abeja
mawu	zorro
Mitsi	Gato
Nakwu	abeja
Nakwutaj	abeja

#### **4. DESDE LA PSICOLINGÜÍSTICA HACIA LA ETNOGRAFÍA: LA VISITA PERIÓDICA**

De las 22 familias participantes que proveyeron reportes parentales, trabajamos con ocho que aceptaron participar en un estudio longitudinal para obtener un corpus del lenguaje infantil y del habla adulta circundante, como así también de las interacciones del infante con su entorno. Como todo proyecto longitudinal, el criterio para la selección de esas familias fue la edad de los niños y niñas. Es reconocida la dificultad e intensidad que se le plantean al investigador que propone llevar adelante un proyecto longitudinal a largo plazo de esta naturaleza, el que involucra a niños y familias durante varios años. Más aún si el mismo involucra a familias indígenas por su poca familiaridad con las prácticas de investigación. Teniendo en cuenta estas dificultades, se optó, como en el caso de estudios previos [8][37], por muestras dinámicas en las que si bien todos los niños y niñas participantes debían ser observados dentro del período promedio de adquisición del lenguaje (1-4 años de edad) al momento de iniciar las observaciones, algunos se encontraban al inicio del proceso y otros hacia el final. Por ejemplo, para el estudio de la adquisición del lenguaje verbal, nuestra muestra estuvo conformada por cuatro participantes observados periódicamente (mensual, bi o trimensual), dos de ellos durante el año y los dos años (Niña 1, Niño 3), uno durante el tercer año (Niño 5) y otro participante durante el tercer y cuarto año (Niño 4). Esta selección casi como ‘en cascada’ es posible dado que la edad no es el criterio evolutivo que consideramos en el

análisis de las muestras de habla; por el contrario, puesto que existen amplios acuerdos acerca de la falta de confiabilidad de las medidas cronológicas desde donde comparar el progreso lingüístico, en nuestros estudios consideramos la Media Longitud del Enunciado (MLE), una medida de desarrollo morfosintáctico que se obtiene como resultado de dividir por cien la suma de los morfemas de cien emisiones de un corpus espontáneo cualquiera [7].

La MLE es una medida que ha demostrado mayor confiabilidad que las medidas cronológicas para el estudio del desarrollo lingüístico en especial en etapas tempranas [6].

A dicha muestra, se sumaron 5 niñas más (Niñas 2, 6, 7, y 8) a fin de estudiar los episodios interactivos entre los infantes y sus madres. Las interacciones cotidianas y la comunicación intencional que utiliza otras semiosis más allá del lenguaje verbal, fueron observadas cada 6 meses entre el año y los dos años y medio, conformando un corpus total de 8 niños y niñas para el estudio de las prácticas de socialización tempranas en la cultura y lengua wichí.

Este seguimiento tuvo lugar entre los años 2012-2017 en una serie de trabajos de campo realizados por la primera autora.

Tabla 2: Lista de participantes en función de las edades de observación y estudio

<b>Participantes</b>	<b>1 año</b>	<b>2 años</b>	<b>3 años</b>
<b>Niña 1</b>	1, 1;2, 1;6, 1;8, 1;11	2, 2;1, 2;3, 2;4, 2;5, 2;6, 2;7, 2;8, 2;9, 2;10, 2;11, 2;12	3, 3;2, 3;3
<b>Niña 2</b>	1;3, 1;6, 1;9	2, 2;2, 2;3, 2;4, 2;6, 2;7, 2;8, 2;10	
<b>Niño 3</b>	1, 1;4, 1;6, 1;9, 1;10, 1;11	2, 2;1, 2;2, 2;3, 2;4, 2;5, 2;7, 2;8, 2;9, 2;10, 2;11	3, 3;1
<b>Niño 4</b>		2;2, 2;3, 2;5, 2;7, 2;11	3, 3;1, 3;2, 3;4, 3;5, 3;6, 3;7, 3;8, 3;9, 3;10, 3;11, 3;11; 4
<b>Niño 5</b>			3;2, 3;3, 3;5, 3;7, 3;11
<b>Niña 6</b>		2, 2;1, 2;2, 2;3, 2;5	
<b>Niña 7</b>	1;2, 1;5		3, 3;11
<b>Niña 8</b>	1;6	2;1, 2;2	

La técnica utilizada es la visita periódica del estilo del psicolingüista [5] pero adaptada al contexto de estudio [8]. Se trata de visitas mensuales en las familias focales. En un esfuerzo por capturar una muestra representativa del habla y de las interacciones en la vida cotidiana de estas familias, las visitas tuvieron lugar en diferentes momentos del día. En esta comunidad, la vida se lleva a cabo fuera de la casa, en el patio o en las áreas circundantes, las personas a menudo comen afuera, allí también las mujeres tejen y lavan la ropa (mujeres), o simplemente están sentados bebiendo mate; los niños y niñas a menudo juegan y las interacciones que tienen lugar en esos contextos incluyen más que el niño o niña y la madre: otros niños y niñas y adultos generalmente están presentes, interactuando ya sea con el niño o niña o entre ellos. Típicamente, muchas personas estuvieron presentes durante las grabaciones, algunas hablaron mucho y otras raramente.

#### **4.1. Registro etnográfico y soporte electrónico**

Esta investigación usa formas de registro que se diferencian de la psicolingüística tradicional, centrada, en general, en el lenguaje verbal. Por lo tanto, como se mencionó, uno de los focos de análisis también son las interacciones cotidianas y la comunicación intencional que utiliza otras semiosis más allá del lenguaje verbal, y que sirven de contexto al uso del lenguaje. Para ello, el registro incorpora la interacción espontánea, en forma de gestos, pautas de mirada, movimientos corporales como manifestaciones comunicativas niño-entorno. El trabajo de registro combina observación, registro de notas sincrónicas y asincrónicas (o a-posteriori) así como también el registro electrónico o grabaciones de video y audio. El uso del video es una herramienta esencial en este tipo de investigaciones en tanto herramienta etnografía potenciada [8] brinda la posibilidad de observar repetidamente y con diferentes perspectivas en las mismas interacciones permitiendo un proceso más minucioso de interpretación y comprensión del fenómeno en estudio. Cada visita tiene una duración de una hora y media de grabación. Para el registro empleamos un grabador de audio Zoom H4N así como una cámara Countour + 2 colocada en un trípode y equipadas con un micrófono externo unidireccional. A modo de 'control', se combinan dos tipos de videograbaciones, las directas realizadas por el investigador y las indirectas realizadas por los propios padres, con previo entrenamiento en el uso de medios electrónicos. De ese modo, al comparar las videograbaciones con presencia y ausencia del investigador se hace posible considerar los usos en los que participa

explícitamente un participante externo. El uso del soporte electrónico resultó en el registro y captura de un corpus denso que tiene un total de 129 horas de grabación.

#### **4.2. Transcripción y anotación**

Todos los enunciados se transcriben en wichí y se traducen en español usando Elan 4.2, por dos hablantes nativos diferentes en colaboración con las autoras. Luego, se trabaja con consultores nativos (todos ellos padres o madres) en grupos focales en su función de jueces calificados para hacer preguntas o consultas y generar acuerdos inter-jueces en especial sobre las transcripciones y/o traducciones y también sobre la lista de palabras de los reportes parentales. Esta etapa del análisis del corpus con los hablantes nativos constituye un proceso muy interesante para la producción de conocimiento y también para el despliegue de las habilidades metacognitivas y metalingüísticas de los propios hablantes sobre su lengua materna. En este punto, muy a menudo observamos que para los propios hablantes es un momento de disfrute hablar sobre su lengua vernácula y el conocimiento que sobre ella poseen [13].

#### **4.3. Procesamiento del lenguaje verbal y de otras semiosis en la interacción infante-cuidador**

Como se anticipó, los temas de interés de la investigación van desde el lenguaje infantil, particularmente estudio de la adquisición y desarrollo de la estructura de la palabra y su organización semántica, hasta aspectos que hacen a los marcos comunicativos y de socialización que los infantes desarrollan en relación a su entorno, para los que se tiene en cuenta no solo el habla escuchada (*input* lingüístico) a la que los infantes son expuestos sino también otras semiosis comunicativas propias de la interacción materno-infantil.

Para el estudio del léxico y semántica temprana, en primer lugar para cada niño o niña calculamos la MLE, y en base a esta medida de desarrollo morfosintáctico agrupamos a los infantes en dos fases: fase mono-morfémica o de producción de un morfema o palabra a la vez (niños y niñas con un MLE =1); fase de combinación de morfemas (niños y niñas con un MLE= mayor a 1). Luego, aislamos las unidades de análisis, examinamos ocurrencias de palabras y de morfemas en función de las dos etapas del desarrollo morfosintáctico; discutimos nuestros resultados evaluando hipótesis y predicciones teóricas respecto de la adquisición léxica y morfosintáctica.

Así, en [37] discutimos la universalidad de la ventaja del sustantivo sobre las demás clases de palabras en el vocabulario temprano [20] y mostramos que para el caso de los niños y niñas wichí-hablantes quienes están expuestos a un predominio del verbo en el habla escuchada (*input* lingüístico), los verbos aparecen bastante temprano en su vocabulario, cuando los infantes comienzan a combinar los primeros morfemas (M=1.5). Observamos también que la emergencia temprana de verbos en el lenguaje infantil wichí está relacionada con el dominio productivo creciente de las categorías gramaticales que caracterizan la morfología de los verbos y sustantivo en esta lengua.

En cuanto al dominio de la morfología tanto verbal como nominal, se revisa si y cuál es la morfología utilizada por el niño o niña para cada categoría gramatical verbal y nominal y se analizan sus producciones en función de las dos etapas morfosintácticas referidas. Por ejemplo, para el caso de la morfología nominal, actualmente estamos analizando la trayectoria evolutiva del uso de la posesión en wichí. Como se observa en (1) resultados preliminares muestran que en la etapa mono-morfémica los infantes tienden a producir tanto sustantivos alienables como inalienables, éstos últimos sin la marca prefijal del poseedor, es decir que solo producen formas no reconocibles en el habla escuchada, las que coinciden desde el punto de vista semántico con el significado léxico correspondiente al objeto poseído.

(1) *tela	(la- <b>katela</b> )	(Niño 4, 2;2; 1MLE)
	3pos-abuela	

Más tarde, durante la etapa de las primeras combinaciones entre morfemas, los infantes comienzan a producir el significado proposicional posesivo tanto inalienable como alienable, aunque el mecanismo morfosintáctico de éste último es cognitivamente más sofisticado por incorporar clasificadores además de la marca del poseedor. Así, vemos en (2) un sustantivo alienable en una construcción posesiva con errores. Este niño (Niño 5, Tabla 2) quien transita la etapa de las primeras combinaciones de morfemas (MLE de 1,5) produce el pronombre posesivo pero no aún el clasificador nominal correspondiente ‘*lo*’, que es el que en el habla escuchada habilita a los sustantivos alienables (animales domésticos) por no tener marcas de posesión inherente, a ser poseídos. En cambio, en (3) otro niño (Niño 4) con un nivel morfosintáctico similar

(MLE de 1.5) ya parece dominar el uso del pronombre posesivo y el clasificador correspondiente (‘*ka*’) de la estructura posesiva en cuestión, que es para este caso, un juguete.

(2) (Niño 5, 3;3; 1,5 MLE)

\*n'-chuku (n'-lo=chuku)

1pos-\*-perro 1pos-cl=perro

‘mi perro’

(3) (Niño 4, 2;11; 1,5 MLE)

tim-n'-ka- toto

[3suj] tragar-1pos-cl-juguete

Para el estudio del habla escuchada (*input* lingüístico) y de los aspectos interactivos-comunicativos no verbales, nuestros indicadores son el habla y las manifestaciones corporales, movimientos y pauta de mirada dirigidas al infante focal. Para su análisis, emprendemos en primer lugar una fase cualitativa tendiente a crear para cada semiosis, un sistema de categorías observacional que refleje de manera adecuada y concienzuda el fenómeno observado y lo hacemos utilizando el método comparativo constante [33]. Dependiendo el foco de interés, primero, analizamos de forma interactiva las interacciones y/o enunciados buscando detectar aspectos relevantes, teniendo en cuenta los antecedentes teóricos y evidencia empírica en los temas de estudio, pero muy especialmente la matriz de datos potenciando el proceso ‘desde abajo hacia arriba’ [15], es decir desde los datos hacia la teoría. Luego, tanto para el procesamiento de las interacciones verbales como corporales expresadas a través de la mirada, tacto, movimiento, en base a las categorías descritas se codifican las interacciones, proceso que se realiza utilizando, como para el caso de las transcripciones, el programa ELAN y desde el cual luego exportamos a una matriz de datos que servirá para su posterior procesamiento cuantitativo.

Como ejemplo, para el estudio de las características del habla escuchada (*input* lingüístico), este análisis nos permitió identificar un repertorio de categorías que conforman un registro particular que denominamos habla materna infantil wichí (HMIW) con características específicas y distintas del registro utilizado en las interacciones wichí adulto-adulto en esta comunidad de habla, interacciones que son hasta ahora las más estudiadas en la literatura [22]. Asimismo, descubrimos en el HMIW un léxico específico de las interacciones infante-adulto wichí, no registrado hasta el momento en la bibliografía sobre esta lengua. Específicamente



revelamos el uso de nueve ítems léxicos, los que formalmente poseen una estructura simple del tipo CVCV combinando una de las formas canónicas de las sílabas en wichí, CV [22]. Semánticamente, se trata de términos equivalentes a los del habla adulta referidos a aspectos de interés para el niño, tales como animales, relaciones de parentesco, juguetes, emociones, alimentos (Tabla 3).

Tabla 3. Lista de palabras infantiles del HMIW organizadas semánticamente con su equivalente término en wichí y español

<b>Término adulto</b>		
	Wichí	Español
<b>Entidades y Personas</b>		
<i>Chuku</i>	<i>asinoj</i>	perro
<i>Lulu</i>	<i>tulu</i>	vaca; carne
<i>Tata</i>	<i>lafwcha</i>	papá
<i>Titit</i>		autito
<i>Toto</i>		juguete
<b>Estados emocionales</b>		
<i>Kuku</i>	<i>nuway</i>	sentir miedo
<i>Nana</i>	<i>oytaj</i>	sentir dolor
<b>Necesidades básicas</b>		
<i>Chichi</i>		pecho materno
<i>Mimi</i>	<i>inot</i>	agua

El método comparativo constante del habla escuchada wichí también reveló aspectos pragmático-discursivos propios del HMIW, a saber: el uso de prescripciones o emisiones que tienden a regular las acciones del niño o niña (*Khel, taypo* [vení, sentate]) a menudo asociadas a advertencias (*tenlhok atauflhy ichojhila'am* [si llorás te va a llevar]); referencias o denotaciones, que dirigen la atención del niño a aspectos del entorno (*hin siwele* [mirá el criollo]) y enunciados como interrogativas, respuestas, afirmaciones, negaciones, comentarios (*¿tenlo inot?* [¿quierés agua?]; *techoj* [hace frío]), *temey tanhi* [Se escucha desde el monte (la chuña)]) que tienden a constituir rutinas conversacionales [25]. Curiosamente, sin embargo, en el HMI wichí no observamos las características prosódicas tradicionalmente identificadas en los estudios en lenguas mayoritarias tales como tono de voz alto y exagerado [11][30]. Con esta evidencia discutimos la universalidad ‘del registro denominado maternés (*‘baby talk’* o

'*motherese*'), registro observado mayoritariamente en interacciones típicamente diádicas en culturas mayoritarias [10] demostrando que algunos de los rasgos y características que lo definen pueden no ser generalizables a todas las comunidades de habla del mundo.

Para el caso del estudio de los datos interactivos no estrictamente verbales, codificamos formas semióticas corporales dirigidas hacia y desde el infante, codificamos así la pauta de mirada e identificamos diferentes categorías que describen un gradiente atencional de los participantes de la interacción, permitiéndonos conocer cómo se organiza la atención en estas interacciones, si la atención del infante se concentra en los cuidadores (persona), en los objetos de su entorno (objeto); si el adulto involucra al niño en interacciones con objetos (atención conjunta), etc. [36]. A partir de este análisis hipotetizamos que incluso la atención, proceso cognitivo básico donde se apuntala tempranamente la comunicación intencional y la socialización temprana, posee una organización cultural que sirve de base para otros procesos intersubjetivos tempranos, los que también son perfilados culturalmente. Precisamente, teniendo como base estos marcos comunicativos e intersubjetivos tempranos, estudiamos la Mentalización Corporeizada Materna (MCM) o capacidad materna para interpretar las intenciones del infante. Para ello se codifica el habla dirigida, si la madre dice 'querés', 'deseas' o utiliza cualquier otro vocabulario referido a estados mentales se considera que tiene en cuenta los estados mentales del infante. Al considerar la mentalización como corporeizada, además del habla dirigida, estudiamos la pauta de mirada y el movimiento, tacto o kinesis de la madre hacia el niño, por ejemplo, si al infante se le cae un objeto y lo busca y el cuidador o cuidadora "lee" la intención, lo busca y se lo entrega, también se considera que mentaliza y tiene en cuenta el mundo interno del infante. Entonces, si la madre, utiliza estos vocabularios y lleva a cabo acciones que van en sincronía con lo que el infante desea se considera que la madre no solo lo mentaliza, sino que además lo hace en sintonía con los deseos del infante [23]. En una investigación transcultural en curso estudiamos estos procesos intersubjetivos entre infantes y cuidadores wichí comparativamente con sus pares de cultura mayoritaria buscando identificar variaciones y puntos de encuentro en los procesos de mentalización materna que sientan las bases e impulsan los procesos de representación y simbolización del niño o niña apuntalando su teoría de la mente [24].

#### **4.4. Perspectiva de trabajo colaborativo**

Ambas metodologías, los reportes parentales y la visita periódica, se realizan desde una perspectiva totalmente colaborativa. Desde este enfoque, investigadores y hablantes nativos construyen el trabajo de campo y, por lo tanto, los miembros de la comunidad de habla participan activamente en todo el proceso, el que incluye la recopilación de los datos, transcripción y análisis. Para ello, fue necesario capacitarlos en el uso de herramientas tecnológicas, por ejemplo grabadores de audio y video digitales, *software* para bases de datos, etc. Los colaboradores y participantes del proyecto son: Aurelia Pérez, quien es además Maestra Wichí de la Escuela Wichí Lako en Laguna Yema, Formosa, María Elena Segundo y Margarita Pérez, ambas de la Comunidad Wichí-Lawet, y participan también otros hablantes nativos así como padres y madres de los niños y niñas participantes. Los padres que aceptan participar son entrenados por las autoras y los colaboradores del proyecto para usar la videocámara para que podamos reunir los datos que necesitamos. Esta alternativa ha resultado muy provechosa, primero, porque los padres están involucrados en el proceso de recolección, y segundo, porque el contexto de las grabaciones es más natural que cuando un investigador está en la escena.

### **5. CONCLUSIONES**

Como puede apreciarse, la construcción de la aproximación metodológica al estudio del lenguaje y socialización infantil en una comunidad de habla indígena representa un trabajo intenso, laborioso, muy artesanal y creativo. Es preciso adecuar métodos sin perder de vista los objetivos de la investigación y las condiciones para su concreción. Se advierten las limitaciones de los métodos de la psicolingüística centrada en general en el lenguaje verbal y la necesidad de abrir al estudio de las otras semiosis que forman parte del entramado intersubjetivo cuidador-infante y que hacen de contexto a la emergencia del lenguaje infantil en las etapas tempranas. Se propone la combinación de varios métodos de obtención de muestras con el fin de obtener una diversidad de perspectivas. Indicamos que el reporte parental es una buena puerta de entrada para ese tipo de poblaciones permitiendo obtener un número mayor y diverso de datos, mientras que las muestras naturalísticas permiten profundizarlos analizando su uso en las interacciones cotidianas pero también ampliando a su interacción con otras formas semióticas como el gesto, el tacto, el movimiento y la pauta de

mirada. Se indica la necesidad de generar categorías de análisis que emerjan de los datos y que sean fieles a las interacciones que se observan y estudian. La combinación de notas etnográficas, diario, registro audio y video es necesaria en todo momento de la muestra. Proponemos que el involucramiento de los hablantes nativos que forman parte del proyecto es esencial, desde su interacción espontánea cotidiana, sus comentarios *in situ*, sobre temas que emergen en la vida cotidiana y en muestras de registros. Finalmente, esperamos que los conocimientos que aquí discutimos puedan servir de referencia y alienten a otros investigadores a considerar más centralmente el estudio del lenguaje y la socialización en diferentes culturas infantiles.

## Referencias

- [1] Audisio, C., Rosemberg, C., Ojea, G. y Cuneo, P. (2019). *Bilingual input to children from a QOM community in Chaco* (Argentina). Presentation at The Fifth Intergenerational Transmission of Minority Languages Symposium: Language, Culture and Ethnicity in Indigenous and Migrant Languages.
- [2] Baiocchi, M., Waxman, S., Pérez, E., Pérez, A. y Taverna, A. (2019). Social-ecological relations among animals serve as a conceptual framework among the Wichi. *Cognitive Development*. 52. 10.1016/j.cogdev.2019.100807.
- [3] Ballena, C. (2013). *Tojha itetshan*. Universidad de Buenos Aires.
- [4] Ballena, C. (2013). *Tochefwenjajn-wu*. Universidad de Buenos Aires.
- [5] Brown, R. (1973). *A first language: the early stages*. Harvard University Press.
- [6] Clemente Esteban, A. (1989). Medida de desarrollo morfosintáctico. Los problemas de la medición y el uso de la MLE (Medida de longitud de emisión). *Anuario de Psicología*, 42(3), 105-113.
- [7] Courtney, E. y Saville-Troike, M. (2002). Learning to construct verbs in Navajo and Quechua. *Journal of Child Language*, 29, 623–54.
- [8] De León Pasquel, L. (2005). *La llegada del alma: lenguaje, infancia y socialización entre los mayas de Zinacantán*. CIESAS. México.
- [9] Evans, N. y H.-J. Sasse. (2002). Introduction: problems of polysynthesis en Evans, N. y Sasse, H.-J. (eds), *Problems of Polysynthesis*, 1-13. *Studia typologica* 4. Akademie-Verlag, Berlin.

- [10] Ferguson, C. A. (1964). Baby Talk in six languages. *American Anthropologist*, 66(6), 103–14.
- [11] Garnica, O. (1977). Some prosodic and paralinguistic features of speech to young children, en Snow, C.E. y Ferguson, Ch.A. (Eds.), *Talking to children: language input and acquisition* (pp. 63-88). Cambridge University Press.
- [12] Gentner, D. (2006). Why verbs are hard to learn, en Hirsh-Pasek, K. y Golinkoff, R. (Eds.), *Action meets word: how children learn verbs* (pp. 544–64). Oxford University Press.
- [13] Goffman, E. (1981). *Forms of Talk*. Blackwell, Oxford.
- [14] Golluscio, L. (1993). Clases de sustantivos y sistema cultural: la posesión en wichí. *Signo y Seña*, 3, 221-239.
- [15] Hill, C. E. (1990). Exploratory in-session process research in individual psychotherapy: A review. *Journal of Consulting and Clinical Psychology*, 58, 288-294.
- [16] Instituto Nacional de Estadísticas y Censos-INDEC. (2010). Censo nacional de población, hogares y viviendas 2010. Censo del bicentenario.
- [17] López Ornat, S. (2011). La adquisición del lenguaje, un resumen en 2011. *Revista de investigación en Logopedia*, 1,1, 1-11.
- [18] López Ornat, S., Gallego, C., Gallo, P., Karousou, A., Mariscal, S., y Martínez, M. (2005). Inventarios de Desarrollo Comunicativo MacArthur. *Manual Técnico*. TEA Ediciones.
- [19] López, A., Silvestre, A., Rosemberg, C.R, Gutiérrez, F., Ojea, G., Díaz, H., Codutti, J.R., Zalazar, M.A., Vallejos, M.L., Suárez, M., García, O., Zenón, S. y Fernández, D. (2010). *Las aventuras de Huaqajñe* [Huaqajñe's adventures]. CIIPME CONICET.
- [20] Jackson-Maldonado, D., Al, D., Marchman, V. A., Bates, E., et al. (1993). Early lexical development in Spanish-speaking infants and toddlers. *Journal of Child Language*, 20, 523–49.
- [21] Migdalek, M. J. y Rosemberg, C. R. (2020). SES Differences in Children's Argumentative Production. *Europe's Journal of Psychology*, 16(2), 193-209.

- [22] Nercesian, V. (2014). *Wichi lhomtes. Estudio de la gramática y la interacción fonología-morfología-sintaxis-semántica*. Lincom.
- [23] Padilla M.I. (2019). *La Mente Mentalizante en la diada adulto-infante: un estudio observacional, longitudinal de caso único* (tesis de grado). Universidad Abierta Interamericana. Rosario, Argentina.
- [24] Padilla, M.I. y Taverna, A. (en evaluación). *Categorías observacionales para el estudio transcultural de la Mentalización Corporeizada Materna (MCM): Evidencia desde interacciones madre-infante wichi e hispanohablante*.
- [25] Pérez, A., Pérez, E. M., Taverna, A. y Baiocchi, M. C. (2017a). *Hal'o*. EDUNaF.
- (2017b). *Laloy*. EDUNaF.
- (2017c). *Tshotoy*. EDUNaF.
- (2017d). *Tshotoy fwiyo'ohen*. EDUNaF.
- (2017e). *Tshotoy inhot lheley*. EDUNaF.
- [26] Resches, M., Kohan Cortada, A. y Picon Janeiro, J. (2018). Trayectorias del desarrollo léxico y Gramatical inicial en niños hablantes de español rioplatense. Experiencias tempranas y contexto social, en Rosemberg, C. (org.), *Conferencia llevada a cabo en el Encuentro Internacional sobre lenguaje, cognición e interacción social en la primera infancia*. Buenos Aires.
- [27] Rosemberg, C. R., Alam, F., Audisio, C. P., Ramirez, M. L., Garber, L. y Migdalek, M. J. (2020). Nouns and verbs in the linguistic environment of Argentinian toddlers: Socioeconomic and context-related differences. *First Language*, 40(2), 192-217.
- [28] Rosemberg, C. R., Borzone, A. M. y Flores, E. (2002). *Las aventuras de Ernestina. [Ernestina's adventures]*. Ministerio de Educación, Ciencia y Tecnología. Programma de Acciones Compensatorias.
- [29] Rosemberg, C. y Stein, A. (2019). Child language and development contexts / Lenguaje infantil y contextos de desarrollo. *Infancia y Aprendizaje*.

- [30] Sachs, J., Brown, R. y Salerno, R. (1972). Adults' speech to children en Raffler-Engel, W. von y Lebrun, Y. (Eds.), *Baby talk and infant speech*. Swets & Zeitlinger.
- [31] Silvestre, A., Gutiérrez, F., García, O., Romero, N., Inderecio, B., Sánchez, A., López, C., Maidana, M., Rosemberg, C. R., Ojea, G. y Fernández, D. (en prensa). Aso Nogotole Le'enaxat Cintia. [A Girl Named Cintia]. Universidad Carlos III de Madrid.
- [32] Stoll, S., Bickel, B., Lieven, E., Banjade, G., Bhatta, T. N., Gaenszle, M., Paudyal, N. P., Pettigrew, J., Rai, I. P., Rai, M. y Rai, N. K. (2012). Nouns and verbs in Chintang: children's usage and surrounding adult speech. *Journal of Child Language*, 39, 284–321.
- [33] Strauss, A., y Corbin, J. (1990). *Basics of qualitative research*. Sage publications.
- [34] Taverna, A. (en evaluación). El habla materna infantil en la lengua wichí.
- [35] Taverna, A. S., Medin, D. L. y Waxman, S. R. (2018). 'Inhabitants of the earth': reasoning about folkbiological concepts in Wichi children and adults, en Marshal, P. J. y Brennehan, K. (Eds.), *Young children's developing understanding of the biological world* (pp. 7–27). Routledge.
- [36] Taverna, A. S., Padilla M. I. (en preparación). The cultural organization of attention: evidence from Wichí and Spanish-speaking mother-infant interactions.
- [37] Taverna, A. y Waxman, S. R. (2020). Early lexical acquisition in the Wichi language. *Journal of Child Language*, 1-21.
- [38] Taverna, A. S., Waxman, S. R., Medin, D. L., Moscoloni, N. y Peralta, O. A. (2014). Naming the living things: Linguistic, experiential and cultural factors in Wichi and Spanish speaking children. *Journal of Culture and Cognition*, 14, 213–233.
- [39] Taverna, A., Waxman, S., Medin, D. y Peralta, O. (2012). Core folkbiological concepts: New evidence from Wichi children and adults. *Journal of Cognition and Culture*, 12(3–4), 339–358.
- [40] Vidal, A. y Nercesian, V. (2005). Sustantivo y Verbo en wichí. Hacia una taxonomía de clases de palabras. *Liames* 5, 7-24
- [41] Zidarich, M. y colaboradores (2006a). *Tsalanawu. Libro de Lectura para la Alfabetización Inicial*. Universidad de Buenos Aires.

(2006b). *Tsalanawu. Fichas de Actividades para la Alfabetización Inicial*. Universidad de Buenos Aires.

(2006c). *Chalanero. Libro de Lectura para la Alfabetización Inicial*. Universidad de Buenos Aires.

(2006d). *Chalanero. Fichas de Actividades para la Alfabetización Inicial*. Universidad de Buenos Aires.



# **La concepción del error en la adquisición del español lengua segunda o extranjera (ELE/2)**

## **Un análisis de los procesos de derivación sufijal español/portugués y los casos de divergencias morfológicas**

**Natalia Ricciardi**

Escuela de Lenguas, Facultad de Humanidades y Artes,  
Universidad Nacional de Rosario

Rosario, Argentina

[natalia.ricciardi@gmail.com](mailto:natalia.ricciardi@gmail.com)

### **Abstract**

The present work aims to consider some theories of language acquisition (cognitive, behavioral, mentalistic, sociological-interactionist), along with theories regarding language errors produced in speakers in their second language (L2), in order to contrast and analyze the suffix derivation in Portuguese and Spanish. We specifically propose to identify and analyze words that present morphological divergences in their suffixes, ones that have maintained their same or highly similar lexemes but are manifested as divergent suffixes, thus leading students to interlanguage production. We think that these idiosyncratic productions in those students who acquire a second language in a language-immersion situation are related to cognitive, psychological and social factors. We also think that a better comprehension of the subjective aspects involved in these processes will contribute to promoting greater respect in intercultural contexts.

**Key-Words:** Second Language Acquisition, Error, Suffix Derivation, Morphological Divergences, Portuguese-Spanish

### **Resumen**

El presente trabajo tiene por objetivo relevar algunas consideraciones de las teorías de adquisición de lenguas (cognitivistas, conductistas, mentalistas y sociológicas) y las teorías sobre el error en las producciones de hablantes de una segunda lengua (L2), con el fin de analizar contrastivamente la derivación sufijal en portugués y español. Específicamente nos

proponemos identificar y analizar aquellas palabras que presentan divergencias morfológicas en los sufijos, es decir, palabras que, en una y otra lengua, han mantenido los mismos lexemas (o muy semejantes) pero manifiestan sufijos divergentes, llevando muchas veces a los alumnos a producciones de interlengua. Entendemos que estas manifestaciones idiosincráticas en alumnos que adquieren la L2 en contexto de inmersión se deben a factores cognitivos, psicológicos y sociales, y que una mejor comprensión de los aspectos involucrados en estos procesos contribuirá a promover un mayor respeto en contextos interculturales.

**Palabras claves:** Adquisición de segundas lenguas, Error, Derivación sufijal, Divergencias morfológicas, portugués-español.

## 1. INTRODUCCION

Los movimientos migratorios de estudiantes se presentan como un fenómeno global y Argentina se destaca como un epicentro en Latinoamérica que atrae estudiantes de diversos países de la región debido a la oferta de enseñanza libre y gratuita de las carreras universitarias en el país (Angelucci, Pozzo, 2020). En esta coyuntura histórico política la enseñanza del español como lengua extranjera se presenta hoy, más que nunca, como un campo que requiere atención y justifica la diversificación de estudios que aborden las varias aristas involucradas. Desde la necesidad de un análisis cuidadoso de las políticas lingüísticas que promueven (o ignoran) el impulso a la enseñanza del español a extranjeros en nuestro país, hasta las diversas propuestas de aplicación concreta a través del análisis de material didáctico para la enseñanza de español lengua extranjera y / o segunda (ELE/2) en Argentina, pasando por los estudios comparativos con fines didácticos y metodológicos que promueven la identificación de semejanzas y diferencias entre dos lenguas tan cercanas, el estudio y problematización de las manifestaciones de interlengua de los luso hablantes que aprenden español se presenta como un campo de investigación complejo y rico. Estudios comparativos y contrastivos han dado lugar a innumerables trabajos que abordaron aspectos vinculados al léxico, la fonética y la fonología, aspectos morfosintácticos como el género de los sustantivos (los heterogénicos); el uso de artículos; los pronombres posesivos; los paradigmas pronominales de objeto directo e indirecto; divergencias y convergencias en las formas y tiempos verbales; el uso de las preposiciones y la regencia verbal y pronominal, entre otros aspectos (Cf. Benedetti, 2002; Bagnó, 2015). La morfología derivativa representa un área productiva de observación y análisis de diferentes ocurrencias de interlengua que manifiestan procesos cognitivos activados por los hablantes en ciertas etapas de su de aprendizaje. Nos interesa explorar, especialmente, las formas idiosincrásicas en la derivación sufijal para la formación de sustantivos en luso hablantes que aprenden español. Nos proponemos entonces reflexionar sobre el papel del error desde las diversas teorías de la adquisición de las lenguas considerando el impacto de los factores sociales y subjetivos en los procesos de aprehensión, para detenernos en la observación de uno de los fenómenos del sistema idiosincrático de los aprendientes de ELE/2 luso hablantes: la morfología derivativa y los casos de divergencias morfológicas. En primer lugar, iniciando con una breve discusión terminológica, presentaremos las principales consideraciones sobre las teorías de adquisición de lenguas

(cognitivistas, conductistas, mentalistas y sociológicas) y las teorías sobre el error (Análisis contrastivo, Análisis de Errores y Teoría de la Interlengua). A continuación, presentaremos resumidamente los procesos de derivación sufijal relevando los principales sufijos formadores de sustantivos en portugués y en español que dan lugar a fenómenos de divergencia morfológica. Estudios de esta índole son necesarios tanto para profesores como para estudiantes, a fin de tomar conciencia de los procesos de fosilización que representan un nivel de interlengua que puede asumirse como suficiente para comunicarse de forma eficiente provocando, en muchos casos, situaciones de incompreensión.

## **2. LAS TEORÍAS DE ADQUISICIÓN/APRENDIZAJE DE LENGUAS**

La preocupación por entender los procesos de adquisición y aprendizaje de las lenguas y los factores que intervienen y los condicionan ha dado lugar a numerosos trabajos que han intentado explicar, desde las aportaciones de diversas corrientes teóricas, la relación entre los sujetos, las lenguas primeras (LM), segundas (L2) y extranjeras (LE), los procesos cognitivos y los factores sociales / contextuales involucrados. A los fines de establecer un vocabulario común que nos permita delimitar las diferentes circunstancias en las que pueden darse los procesos de aprehensión de una lengua y discutir la especificidad de los factores involucrados en estos procesos, proponemos revisar brevemente algunos conceptos.

En primer lugar, retomamos las nociones de adquisición y aprendizaje entendiendo que pueden configurarse como procesos diferentes dependiendo de las circunstancias. Pese a que en muchos casos estos términos suelen utilizarse como equivalentes, Krashen (1985) establece una diferencia proponiendo que la *adquisición* se da por contacto directo con una L2 en contextos de inmersión, mientras que el *aprendizaje* supone un proceso formal en contextos institucionalizados. Mientras que en el proceso de adquisición el individuo concentra su atención en la transmisión de significados por necesidad de comunicación, en el proceso de aprendizaje el alumno, consciente de las reglas del sistema, focaliza su atención en la corrección gramatical. Este modelo dicotómico de competencia dual propuesto por Krashen entiende que entre estos procesos diferenciados no hay posibilidad de interrelación. Si bien este modelo ha recibido críticas debido a su falta de rigor científico y la laxitud de definiciones como "adquisición", "aprendizaje", "consciente" y "subconsciente", que impiden determinar exactamente en qué consiste la "lengua adquirida" y la "lengua aprendida" (McLaughlin, 1987, pp.21-24), no nos detendremos aquí en esta discusión terminológica. Nos

referiremos a la adquisición como el proceso espontáneo e inconsciente que se da en contextos naturales de uso de la lengua objeto, y al aprendizaje como el proceso mediado por la aprehensión consciente de las reglas del sistema y por procesos formales (generalmente institucionalizados), entendiendo que puede existir un solapamiento entre ambos en diversas situaciones de aprehensión de la lengua<sup>1</sup>.

Por otro lado, si partimos de la idea de que el aprendizaje de una lengua puede darse en variados contextos, este factor contextual o social determina la forma y usos que asumen las nociones de Lengua Extranjera (LE), y Lengua 2 (L2). Si bien algunos autores engloban en la sigla L2 tanto los conceptos de Lengua Segunda como Lengua Extranjera (Martín, 2004), en ciertos contextos las circunstancias de aprendizaje /adquisición determinan usos explícitos diferenciados. De esta forma, en el caso de aprendizaje de una lengua en contexto natural en el que el aprendiz está en exposición constante y contacto permanente con la lengua objeto de aprendizaje estaremos hablando de una *segunda lengua* (L2) entendiendo *segunda* en oposición a la noción de *lengua materna* (LM), mientras que la noción de lengua extranjera (LE) refiere al aprendizaje que normalmente se produce en una situación formal, institucionalizada, en contexto de aula. A su vez, existe un contexto mixto, esto es, aquella situación en la cual el aprendiz desarrolla su vida social (por trabajo o estudio) entre hablantes nativos del país a la vez acompaña el proceso de aprehensión a través de la instrucción formal. Para un inmigrante brasileño que estudia una carrera universitaria en Argentina, por ejemplo, la adquisición del español se producirá en un contexto natural, en contacto con profesores, estudiantes y amigos, a la vez que puede aprender la lengua en un instituto de idiomas. Hablaremos entonces de ELE/2 (Español como Lengua Segunda y Extranjera) para referirnos a los procesos de adquisición de español como lengua extranjera en contexto natural que pueden apoyarse también en situaciones de enseñanza formal.

Hechas estas distinciones, de forma resumida y a riesgo de caer en simplificaciones, presentaremos ahora cuatro tendencias teóricas que, a nuestro entender, han marcado las formas de concebir la adquisición tanto de LM como de L2 y, consecuentemente, la concepción del error en el proceso de adquisición de la L2.

<sup>1</sup> Nos inclinamos a entender que la adquisición y el aprendizaje no son dos términos discretos y dicotómicos que no permiten el trasvase de conocimiento consciente a conocimiento inconsciente. La hipótesis de la *interfaz* propone que existe un continuum de las distintas formas de conocimiento que se produce por diferentes mecanismos y procesos vinculados con el uso y la práctica y la concienciación de las reglas del sistema (Baralo, 1995, p. 65).

Las teorías *cognitivistas* equiparan el proceso de aprendizaje de lenguas a otros aprendizajes, esto es, en tanto proceso mental que implica el desarrollo y adquisición gradual de estructuras lingüísticas a través de estrategias de percepción, almacenamiento y producción. Piaget (1926) afirma que las intervenciones formativas del entorno social no tienen un impacto en el proceso de adquisición del lenguaje, si no que existe un juego de asimilaciones y acomodaciones que dan lugar al pensamiento consciente. Si bien Vygotsky ([1934], 1995) y su escuela suelen encuadrarse en esta misma corriente, este último considera los factores sociales (los preestructos históricos) como siendo decisivos, junto con la evolución biológica, para la constitución del pensamiento formal. De esta forma, el lenguaje es un proceso de interiorización producto del aprendizaje en sociedad desde los primeros estadios del niño. Por su parte, las teorías *conductistas* afirman que el ser humano es capaz aprender a partir de un proceso de formación de hábitos (inclusive lingüísticos) a través de la secuencia *estímulo-respuesta-refuerzo*. El aprendizaje de una lengua sería así el resultado de la formación de hábitos que se adquieren por imitación y repetición (Skinner, 1957).

Surgida como respuesta crítica al conductismo, la teoría *mentalista* afirma que no se aprende por condicionamiento externo (formación de hábitos a partir de un estímulo) sino que los humanos nacen genéticamente dotados de un dispositivo de adquisición del lenguaje (DAL) que permite el desarrollo de la capacidad lingüística (Chomsky, 1959). Este mecanismo dispone de una gramática universal (GU) que se rige por principios comunes a todas las lenguas naturales a la vez que, al procesar el lenguaje, el hablante aplica los parámetros concretos de una lengua en particular.

Dentro de las corrientes *sociológicas*, las teorías *interaccionistas* intentan explicar los procesos de aprendizaje de lenguas considerando tanto los factores innatos como los contextuales. Estos estudios consideran que todos los estudiantes tienen características diferentes y están sometidos a diferentes condiciones contextuales que influyen en el grado de procesamiento del input en la lengua extranjera. Dentro de estas últimas corrientes la *sociolingüística* enfatiza en el factor social sin negar los factores cognitivos aportando un enfoque más integral. El modelo de Spolsky (1989, citado en Moreno Fernández, 2016) considera que el contexto social en el que se lleva a cabo el proceso determina una serie de actitudes diversas que dan lugar a una motivación para aprender. En esta amplia interpretación de *contexto social* Spolsky incorpora tanto aspectos personales (actitudes, conocimientos previos) del propio aprendiz y de otros sujetos que acompañan el proceso

(familiares y afectos cercanos en el proceso de adquisición de la LM; compañeros y profesor en el caso del aprendizaje de una LE o colegas de trabajo o estudio en el caso de una L2), así como los rasgos y características de los grupos sociales en los que el sujeto está adquiriendo o aprendiendo la lengua.

En contextos de inmersión en los que el aprendiz convive con una cultura y un contexto lingüístico diferente del propio, la *comunicación intercultural* determina en gran medida la interacción comunicativa, favoreciendo o perjudicando el proceso de adquisición. En espacios geográficos marcados por procesos inmigratorios, suelen darse situaciones discriminatorias que conllevan actitudes de hostilidad y valoración negativa del “acento extranjero”. Estas situaciones determinan en gran medida la producción lingüística y llevan frecuentemente al aprendiz a cometer errores en su comunicación (Moreno Fernández, 2016).

### **3. EL TRATAMIENTO DEL ERROR EN LAS DIFERENTES CORRIENTES TEÓRICAS**

Los errores, aquellas manifestaciones singulares o rasgos de la producción oral o escrita de los aprendices que se desvían de los que son propios de la lengua meta, han ido adquiriendo diferentes valoraciones en estrecha vinculación con las corrientes psicolingüísticas vigentes en cada momento histórico.

Al considerar que el aprendizaje se basa en la adquisición de conductas (inclusive lingüísticas) a través de la repetición, para el *conductismo* el error es un aspecto negativo, puesto que representa la adquisición de una estructura lingüística equivocada. El objetivo principal en el proceso de aprendizaje es evitar la formación de hábitos incorrectos, es decir, de errores. Con el surgimiento de la teoría mentalista (Chomsky, 1965), como respuesta al conductismo, el error es considerado como un proceso cognitivo natural que consiste en la formulación de hipótesis sobre el sistema de la lengua que el sujeto está aprendiendo. Con el *análisis contrastivo*, teoría que intenta establecer las diferencias y similitudes entre la lengua materna y la lengua meta, el error comienza a ser visto como el resultado de las *interferencias* de la lengua materna del alumno en el proceso de aprendizaje de la lengua meta, esto es, la tendencia del alumno a reemplazar rasgos (morfológicos, fonológicos, sintácticos) de la lengua extranjera por otros de su lengua materna. A partir de la evolución del modelo de Corder (1981/1992), se afianza el *análisis de errores* desde una perspectiva más empírica al

analizar los errores efectivamente producidos por los estudiantes de lenguas extranjeras, dejando de lado la predicción, que ocupaba un lugar central en el análisis contrastivo. En esta línea de pensamiento los errores son considerados (...) “exponentes de procesos psicológicos subyacentes capaces de ilustrar, aunque fuera parcialmente, el proceso de aprendizaje (...)” (Santos Gargallo, 1993, p.391). Se establece entonces que existen paralelismos observables entre los procesos de adquisición de la LM y la L2, y en consecuencia el proceso de aprendizaje de la segunda se ve facilitado por la experiencia previa de adquisición de la primera. Surge el concepto de *Interlengua* para designar aquellas manifestaciones lingüísticas singulares del aprendiz en determinada etapa del proceso de aprendizaje que pertenecen a un sistema independiente de la LM y la L2 (Selinker, 1972). Este sistema lingüístico del hablante no nativo en una determinada etapa del proceso de aprendizaje, presenta elementos de la lengua materna, otros de la lengua meta y algunos exclusivamente idiosincrásicos (Santos Gargallo, 1993, p.393). Con estos aportes teóricos se pasa a considerar al error como la demostración de que el aprendiz elabora, comprueba o refuta hipótesis sobre el sistema de la lengua meta.

A pesar de las diferentes consideraciones teóricas que han resignificado el papel del error en el proceso de aprendizaje de una L2, aún hoy los errores en la comunicación intercultural (en contextos de inmersión) pueden tener impactos sociales negativos en diferentes situaciones y ámbitos. Moreno Fernández (2016) destaca los contextos de acceso controlado, en los que un hablante nativo ejerce control sobre el discurso del no nativo, como en las situaciones de evaluación oral de los exámenes de certificación en los que el examinador nativo interactúa y evalúa al candidato con parámetros de la conversación propios de la cultura meta. A su vez, el factor subjetivo también impacta significativamente en las emisiones de los hablantes de una L2 y en las posibles manifestaciones de interlengua. La *identidad social* fuertemente vinculada a procesos de categorización social, formación de una conciencia de identidad social, comparación social y búsqueda de una distintividad psicológica (McNamara, 1997b), tiene un impacto decisivo en procesos de conformación subjetiva mediados por el lenguaje. Muchas veces la presencia de rasgos de la LM en el discurso del hablante no nativo puede responder a procesos de reafirmación de la propia identidad.



#### 4. LA DERIVACIÓN SUFIJAL Y LOS CASOS DE DIVERGENCIA

En análisis previos hemos observado que los errores o manifestaciones idiosincrásicas en producciones de sujetos luso hablantes que aprenden ELE/2 se manifiestan en todos los niveles del sistema de la lengua (Cf. Ricciardi, 2015). En el nivel morfológico, la derivación supone una serie de operaciones de adición que requieren del conocimiento de los sufijos formadores de nuevas palabras y de las diversas formas posibles de combinación entre raíces y afijos derivativos (Varela Ortega, 2003).

La derivación puede realizarse por adición de un afijo o por otro medio no afijal, como la formación regresiva, que consiste en un proceso de sustracción y no de añadidura (*sostener*>*sostén*). Con mayor frecuencia, tanto en español como en portugués, la derivación se presenta a través de la adición de un sufijo antes o después de la raíz, o ambos casos en simultáneo. Llamamos prefijos a aquellos morfemas que se colocan antes de la raíz para formar una nueva palabra (*suponer* >*presuponer*), sufijos a los morfemas que se posponen a la raíz y forman una nueva palabra (*príncipe*> *principado*) y parasíntesis a la ocurrencia simultánea de prefijo y sufijo (*anocheecer*). Las palabras derivadas pueden tener distintos grados de complejidad, ya que se pueden formar palabras con varios sufijos en secuencia (*rentabilización: renta(r) > renta-ble>rentabil-iza(r) >rentabiliza-ción*); o inclusive agregar un prefijo a una palabra formada por secuencia de sufijos (*abastece(r)>abastec-i-miento>auto-abastecimiento*). (Varela Ortega, 2005)

Tanto el portugués como el español presentan formación de nuevas palabras a través de los procesos de derivación prefijal, sufijal, parasintética y regresiva (Cf. Varela Ortega, 2005; Infante, [1995] 2001). Con respecto a la derivación sufijal, a pesar de existir una extensa lista de sufijos compartidos por ambas lenguas (con algunas diferencias ortográficas destacadas, como en el caso de *esperanza / esperança; diferencia / diferença; maldad / maldade*) se identifican determinados casos en los que la junción de lexema y sufijo en el proceso de derivación se dio de forma divergente en la historia de ambas lenguas. En otras palabras, si bien el lexema permaneció igual o muy semejante, cada lengua optó por sufijos diferentes en la formación de nuevas palabras a partir de una palabra primitiva. (*gratitud / gratidão; campesino / camponês*). Este fenómeno da lugar a ciertas manifestaciones idiosincrásicas de la interlengua de aprendices de español luso hablantes que consisten en combinar de forma

inexacta raíz y sufijo, dando lugar a ocurrencias de *divergencia morfológica en la combinación raíz + sufijo*).<sup>2</sup>

Algunos de los sufijos del portugués formadores de nombres divergentes en relación al español son:

**-dão:** este sufijo en portugués es formador de una serie de sustantivos que indican cualidad, propiedad o estado. Observemos, en primer lugar, los sustantivos *gratidão, lentidão, retidão*, que se manifiestan en español con el mismo lexema (*-grati,-lenti*.) o con una variación ortográfica (*-recti*) pero que, sin embargo, optaron por el sufijo derivacional *-tud* (*gratitud, lentitud, rectitud*). Destacamos que en portugués la forma del sufijo *-tud* se manifiesta como *-tude* y forma sustantivos como *juventude, similitude, amplitude, magnitude, pulcritude*.

Observamos también la ocurrencia de la palabra *podridão* y *mansidão*, que en español, manteniendo casi el mismo lexema, pasó a formarse con el sufijo *-dumbre* (*podredumbre* y *mansedumbre*) y *escuridão* que en español se manifiesta como *oscuridad*.

**-ista:** tanto en español como en portugués este sufijo formador de sustantivos y adjetivos a partir de otros sustantivos y adjetivos puede indicar:

-adeptos a ciertas doctrinas: calvinista, budista, materialista, capitalista.

-oficios o agentes: flautista, maquinista, florista, tenista

-adeptos a determinadas formas de actuar o pensar: oportunista, revanchista

Sin embargo, observamos un grupo de nombres que indican ciertas especialidades médicas que en portugués se forman con el sufijo *-logista* en lugar del sufijo por el cual ha optado el español para indicar las mismas especialidades (*-logo*) dando lugar a *cardiologista, odontologista, ginecologista, endocrinologista, oftalmologista, neurologista*, etc.

Otros ejemplos de manifestaciones divergentes en la morfología flexiva del español y el portugués con palabras que mantienen la misma raíz son las que listamos a continuación:

---

2 Estas manifestaciones idiosincráticas de divergencia morfológica en la elección del sufijo formador de sustantivos y adjetivos no es exclusiva de luso hablantes aprendientes de español, sino que se da con frecuencia en hablantes de español que aprenden portugués (*\*solidade, \*oscuridade, \*mantenimento, \*financiação*, entre otras).

<b>Español</b>	<b>Portugués</b>
<i>Arañazo</i>	<i>Arranhão</i>
<i>Campesino</i>	<i>Camponês</i>
<i>Financiación</i>	<i>Financiamento</i>
<i>Herrumbre</i>	<i>Ferrugem</i>
<i>Inversionista</i>	<i>Investidor</i>
<i>Inversión</i>	<i>investimento</i>
<i>Inmensidad</i>	<i>Imensidão</i>
<i>Lavado</i>	<i>Lavagem</i>
<i>Multitud</i>	<i>Multidão</i>
<i>Seguridad</i>	<i>Segurança</i>
<i>Vejez</i>	<i>Velhice</i>

Este fenómeno da lugar a producciones de interlengua de los hablantes de portugués que aprenden ELE/2 en las que puede observarse la ocurrencia de la raíz convergente (en una y otra lengua) combinada a un sufijo derivativo existente en español pero que se presenta inadecuado para esa raíz. Expresiones como *\*manuten-ción*, en lugar de *manteni-miento*, *\*financia-miento* en lugar de *financia-ción*, *\*seguran-cia* en lugar de la palabra española *seguir-dad*, *\*apar-ejo*, en lugar de *apar-ato*, *\*suges-tión* en lugar del vocablo español *sugeren-cia*, *\*investimiento* en lugar de *inversión* o *\*investí-dor* en lugar de *inversión-ista*, entre otros casos, proporcionan datos interesantes sobre las operaciones cognitivas de combinación de raíz + sufijo que podrían activarse en luso hablantes en su proceso de aprehensión de español. La observación y comprensión del funcionamiento de los procesos no solo de derivación sufijal divergentes sino también de palabras con derivación convergente con diferencias ortográficas (*-miento / -mento; -dad / -dade*) pueden responder a operaciones de comprobación de hipótesis basadas en los conocimientos morfológicos en lengua materna y también en la lengua extranjera.

## **5. CONSIDERACIONES FINALES**

A partir del recorrido mostrado por las diferentes corrientes teóricas de adquisición de lenguas (LM y L2) hemos podido determinar que, de acuerdo con cada propuesta, existen diversos factores que influyen en los procesos de adquisición y aprendizaje. Vinculadas a estas corrientes, existen diferentes consideraciones sobre el error en tanto manifestación lingüística en el proceso de adquisición de los aprendientes que se desvía de la las características de la

lengua meta. Existe hoy un consenso entre los estudiosos de Adquisición de Segundas Lenguas en considerar los errores como evidencias de los procesos de aprehensión de un nuevo sistema que no representan un problema ni un desvío y que deben servir como base para la investigación de las operaciones cognitivas que intervienen en estos procesos. Sin embargo, en muchos espacios geopolíticos interculturales que presentan población inmigrante, prevalecen ciertas actitudes de intolerancia ante estas manifestaciones idiosincrásicas propias de la interlengua de los hablantes no nativos. Factores de orden cognitivo, psicológico y social condicionan las producciones de los hablantes de una L2 en contexto de inmersión, y es posible que muchas de las manifestaciones idiosincrásicas respondan a procesos diversos, vinculados tanto a operaciones cognitivas de comprobación de hipótesis como a otros vinculados a la conformación de la identidad social.

Los estudios de lenguas próximas como el portugués y el español en contextos geopolíticos marcados por procesos de integración regional, pueden aportar datos interesantes no solo para la elaboración de material didáctico para fines y públicos específicos sino también para la consideración cuidadosa de los factores involucrados en las manifestaciones de interlengua de alumnos en contextos de inmersión. En ese sentido, la reflexión comparativa entre ambas lenguas en los diferentes niveles de análisis lingüístico fomenta la concientización de aspectos que pueden tanto facilitar el proceso de aprendizaje como consolidar procesos de fosilización, y esta reflexión contribuye a la elaboración de material didáctico que se beneficie de la proximidad entre estas lenguas y proponga abordajes adecuados, constructivos y reflexivos.

La morfología derivativa en clase de ELE/2 resulta de fundamental importancia para dotar al alumno de una mayor autonomía en su proceso de adquisición de la lengua. Entender los procesos de derivación sufijal permitirá, por un lado, comprender la clase gramatical a la que pertenece la palabra con sus consecuentes combinaciones en la cadena sintagmática y por otro, dominar gradualmente los morfemas derivacionales y las diferentes combinaciones posibles para formación de elementos léxicos en la lengua meta. En este proceso de aprendizaje, manifestaciones que presenten divergencias morfológicas en la combinación de raíz + sufijo deben considerarse como indicios de procesos cognitivos de concientización, comparación y elaboración de hipótesis de los aprendientes.

Finalmente, creemos que entender que los procesos de adquisición de una L2 se ven atravesados por factores de diversa índole, y comprender que dichos procesos intervienen en

la conformación de la identidad social del sujeto en la(s) lengua(s) que habla, nos llevará a evitar actitudes de intolerancia lingüística y promover un mayor respeto y entendimiento en contextos interculturales.

## Bibliografía

- Angelucci, T. C. y Pozzo, M. I. (2020). Estudiantes brasileños en la facultad de ciencias médicas de Rosario (Argentina): implicancias interlingüísticas. *Trabalhos em Linguística Aplicada*, 59(1), 807-833.
- Carvalho, O. L. S. y Bagno, M. (2015). *Gramática brasileira para hablantes de español*. Parábola, São Paulo.
- Baralo, M. (1996). Adquisición y/o aprendizaje del español/LE. En: Actuales tendencias en la enseñanza del español como lengua extranjera II. *actas del VI Congreso Internacional de ASELE*. León 5-7 de octubre de 1995. Servicio de Publicaciones, pp. 63-68.
- Benedetti, A. M. (2002). El portugués y el español frente a frente: aspectos fonéticofonológicos y morfosintácticos, *Carabela*, 51, pp. 147-171.
- Chomsky, N. (1959). A review of BF Skinner's Verbal behavior. *Language*, 35 (1), pp. 26-58.
- Chomsky, N. (1965). *Aspects of the Theory of Syntax*. Traducción española: *Aspectos de la teoría de la sintaxis*, Barcelona, Gedisa, 1999.
- Corder, S. P. (1992). Dialectos idiosincrásicos y análisis de errores. *La adquisición de lenguas extranjeras: hacia un modelo de análisis de la interlengua*. Visor, pp. 63-77.
- Infante, U. (1995). *Curso de gramática aplicada aos textos*. Scipione. São Paulo.
- Krashen, S. D. (1985). *The input hypothesis: Issues and implications*. Addison-Wesley Longman Ltd.
- Martín, J. M. (2004). La adquisición de la lengua materna (LM) y el aprendizaje de una segunda lengua (L2)/lengua extranjera (LE): procesos cognitivos y factores condicionantes. Sánchez Lobato, J; Santos Gargallo, I. *Vademécum para la formación de profesores*. SGEL, Madrid.
- McLaughlin, B. (1987). *Theories of second language learning*. Arnold, London.
- McNamara, T. (1997). Theorizing social identity; what do we mean by social identity? Competing frameworks, competing discourses. *TESOL quarterly*, 31(3), 561-567.
- Moreno Fernández, F. (2016). El contexto social y el aprendizaje de una L2/LE. Sánchez Lobato, J. y Santos Gargallo, I. *Vademécum para la formación de profesores*. Enseñar

- español como segunda lengua (L2) / lengua extranjera (LE)*. Sociedad General Española de Librería, Madrid.
- Ortega, S. V. y Fabregat, S. (2005). *Morfología léxica: la formación de palabras*. Gredos.
- Piaget, J. (1984). *La representación del mundo en el niño*. Ediciones Morata.
- Ricciardi, N. (2015). La presencia del error en la clase de español LE en estudiantes brasileños: Algunos análisis. *Tramos y Tramas V: Culturas, lenguas, literatura e interdisciplina. Estudios comparativos*. Laborde Libros, Rosario, pp. 383-391.
- Santos Gargallo, I. (1993). *Análisis contrastivo, análisis de errores e interlengua en el marco de la lingüística contrastiva*. Síntesis, Madrid.
- Skinner, B. F. (1957). *Verbal Learning*. Appleton-Century-Crofts, New York.
- Varela Ortega, S. (2003). Léxico, morfología y gramática en la enseñanza de español como lengua extranjera. *ELUA. Estudios de Lingüística*. Nro. 17, pp. 571-588.
- Vygotsky, L. S., Kozulin, A. y Abadía, P. T. (1995). *Pensamiento y lenguaje*. Paidós, Barcelona.

# La gramática del español. Perspectivas teórico-descriptivas y metodológicas en la enseñanza como L1 y L2

Nora Múgica

Facultad de Humanidades y Artes, UNR

[noramugica4@gmail.com](mailto:noramugica4@gmail.com)

## Resumen

El presente trabajo solo pretende ser un aporte acotado a la propuesta de la enseñanza de gramática en el aula, dirigida, en particular, a aprendientes de español como segunda lengua, y, por extensión, a ciertos aspectos generales que incluyen, también, la enseñanza de español como L1. Los puntos axiales de este son la formación de estructuras por combinación en la que importa la sintaxis, la morfología, el léxico, en busca del significado. Los puntos de vista metodológicos se apoyan en que la actitud esperable del docente y del aprendiente es la reflexión. Se insiste en que el conocimiento gramatical es un proceso con etapas similares a los de cualquier ciencia, y acorde con ello, de formulación intrínseca de hipótesis y corroboraciones. Por lo que el marco general no es el de ver los objetos de estudio como contruidos, y analizarlos desde afuera, sino que sean el input para elaborar cuestiones puntuales y llegar a ciertas generalizaciones.

**Palabras –clave:** gramática, sintaxis, léxico, interfaces, gramática del español/ enseñanza.

## Abstract

The present work is only a limited contribution to the proposal of the teaching of grammar in the classroom, directed, in particular, to learners of Spanish as a second language, and, by extension, to certain general aspects that also include the teaching of Spanish as L1. The axial points are the formation of structures by combination in which the syntax, the morphology, the lexicon matter in search of the meaning. The methodological points of view are based on the fact that the attitude of the teacher and the learner is reflection. It is insisted that grammatical knowledge is a process with stages similar to those of any science, and in accordance with it, of intrinsic formulation of hypotheses and corroborations. Therefore, the



general framework is not to see the objects of study as constructed, and to analyze them from the outside, but rather they are the input to elaborate specific questions and arrive at certain generalizations.

**Key-words:** grammar, syntax, lexicon, interfaces, Spanish grammar / teaching.

## 1. INTRODUCCIÓN

El aporte de este trabajo son consideraciones acerca de la enseñanza de una lengua particular como L1 o como L2, pensando en un aprendiente adulto, en el ámbito universitario; el área de interés es el de la gramática y del léxico, en relación con una teoría lingüística particular, los principios generales de la Gramática Generativa como soporte del estudio gramatical de la lengua. La idea general es que para exponer, describir, explicar, cuestionar aspectos propios del lenguaje es necesario hacerlo desde algún punto de vista. No es meramente mostrar, por ejemplo, una estructura, e interesarse en ver cómo implementar esa puesta en escena; es, en cambio, definir cuál es el qué de la cuestión, cuál es el aspecto seleccionado, qué orientación darle, cómo implementar la exposición para que sea accesible y, por lo tanto, desde qué óptica se la va a plantear. En síntesis, decimos que por detrás de lo manifiesto, de lo que se trabaje en cada clase, hay un punto de vista que he discutido conmigo mismo o con otros si ha sido posible, que lo tengo internalizado, y que, en sentido amplio, se extiende a cómo veo el lenguaje, con qué nociones, principios, denominaciones y terminología voy a hacer mi trabajo en la lengua particular, qué categorías voy a adoptar, etcétera. Lo que decimos, en definitiva, se resume en la relación que establezcamos entre la lengua particular y el sostén teórico, el desde dónde.

Visto desde la perspectiva de la enseñanza, es válido decir que este paso, el de la teoría a la práctica, no es el de una relación directa. Es un proceso con distintas etapas de adecuación y de reciclado, un proceso en el que pueden intervenir distintos agentes para cada etapa, ya que esta adecuación supone, en principio, delimitar el objeto de estudio, y definirlo siguiendo las operaciones que hemos mencionado, buscando las palabras más apropiadas al nivel al que nos dirigimos.

Solo para dar un ejemplo: supóngase que introduzco en mi clase algún tópico que tiene que ver con *la linealidad de constituyentes* en una secuencia del español (un punto relevante en el estudio del español). Mi pensamiento, entiendo, tendrá que haber pasado por el tema general y particular del orden de palabras, del orden de sintagmas<sup>1</sup>, de las distintas posiciones que los

---

<sup>1</sup> La *linealidad* hace referencia al orden de los constituyentes en la oración. Los estudios de lenguas particulares señalan dos órdenes canónicos: lenguas de orden fijo, lenguas de orden libre, a los que no se ajustan stricto sensu. El caso del español es el de una lengua semilibre, en cuanto a que hay flexibilidad en la disposición sucesiva de los sintagmas, no así de las palabras, con ciertas

mismos pueden ocupar en la estructura lineal (definir el centro, las posibles anteposiciones o posposiciones, los efectos de ambigüedad que puedan producirse, los de foco, las distintas lecturas o interpretaciones a que den lugar, teniendo in mente, por cierto, que estoy hablando acerca del español y si está tomado como L2, estoy además, pensando en otra lengua como posible referente.). Naturalmente que lo primero que tengo que tener a mano es qué se quiere decir al decir *linealidad* o *linealización*, por qué es de interés su análisis, qué aspectos conviene tener en cuenta, cómo caracterizarlos, qué variaciones son posibles, cómo una organización lineal no es equivalente a una organización jerárquica, y demás, como se puede mostrar en el ejemplo que sigue,

1 Inteligentemente Juan resolvió el problema

con una doble lectura: (i) Juan fue / se mostró inteligente al resolver el problema (ii) la resolución del problema fue inteligente.<sup>2</sup>

En otros términos, si intentamos orientar al alumno a que busque, reflexione y haga su propio recorrido, se hace indispensable tener en claro cuáles son esos pasos, de qué materiales dispongo para justificarlos, y cómo adaptarlos al nivel de los alumnos de que se trate.

En lo que sigue abrimos la sección 2 destinada a *Preliminares en la relación gramática y léxico*, 2.1. *Acerca de Gramática*; en 2.2. plantemos la relación de *Interfaz léxico –sintaxis*. En 3. van las *Conclusiones*

## 2. PRELIMINARES: RELACIÓN GRAMÁTICA –LÉXICO

Nos referimos en este apartado a tres aspectos que consideramos preliminares del estudio que planteamos, uno al *saber* y al *hacer gramática*; otro, a la *relación gramática –léxico*.

---

restricciones, por ejemplo un SP (=sintagma preposicional) estará precedido necesariamente por la P; una oración relativa o interrogativa tendrá al inicio el marcador correspondiente. Por otra parte, la posibilidad de variación en la colocación de los sintagmas puede también ser una herramienta que conduzca a la ambigüedad de la interpretación o depender de factores estilísticos. Una razón aducida que justificaría la libertad de colocación del español es la característica de omisión de sujeto. En paralelo, una lengua como el inglés, el alemán, tiene reglas de colocación de los constituyentes mucho más marcadas.

2 Desde ya que estas variaciones van a estar sujetas a la raíz léxica del adverbio en –mente. Si cambiamos *inteligentemente* por *adecuadamente*, se reducen las interpretaciones, en la medida que difícilmente pueda entenderse que Juan fue adecuado.

## 2.1. Acerca de la gramática

En esta primera instancia, vamos a considerar una evidencia a favor de una hipótesis simple: *todo hablante tiene, conoce, sabe una gramática*. Qué queremos decir con eso de que *tiene una gramática* Veamos un caso a modo de ilustración. Un hablante dice, escucha, lee:

(2) Mientras iba a la plaza, pensaba en muchas cosas, que creo, hoy ya no las recuerdo.

Ese hablante interpreta, sin duda, lo que dice o escucha o lee. El conjunto es una sucesión sonora o gráfica que es interpretable. El conjunto no es simple, es complejo. Tiene la complejidad que emana de una estructura que introduce *mientras* indicando un proceso en el tiempo, al que sigue *iba* y en paralelo, *pensaba*, es decir, va describiendo con las palabras la linealidad del tiempo (que marca el evento). Además, ubica este evento en el pasado, en un pasado que también nos habla de proceso, de transcurrir (porque no nos dice *fue* ni *pensó*). Pero hasta esta altura de la secuencia, no hay definición de quién es el que iba ni pensaba, esto es, de lo que llamaríamos *sujeto* porque nuestros *iba* y *pensaba* son ambiguos en la definición de la persona (¿primera persona, tercera persona? ¿Yo, él/ella?). Además, en la macro combinación, introducimos un *que* y más adelante un *las*, en relación; dos pronombres con pocas marcas gramaticales, el *que*, indeterminado, el *las*, con indicación de género y número, pero los dos entendemos con remisión a *muchas cosas*, es decir, dos constituyentes en relación anafórica con *de muchas cosas*, dos constituyentes que toman los rasgos nominales que *las* visibiliza y *que*, no. Algo más: insertado hay un *creo* que queda como saliendo de la secuencia, fuera de ella; y en fin, decimos *ir a...*, *pensar en ....*

Esto es, hay mucha estructura, hay mucha combinación, una intrincada relación entre lo que llamamos sintaxis, morfología, y léxico. Y en consecuencia, de todo este conjunto, hay un significado. En síntesis, (2) introduce un enunciado complejo que responde a esquemas que bien pueden dibujarse como un diseño gráfico. La pregunta es: ¿por qué un hablante (del español) lo puede interpretar? Porque, decimos, *tiene una gramática, sabe su gramática*, la gramática de la lengua materna, que no representa un saber consciente, sino que, de manera natural, va armando y reconfigurando, en un proceso que parte desde la infancia, un proceso compartido con los demás hablantes de una lengua. Es un conocimiento que sucede como una necesidad biológica. Desde el punto de vista de la enseñanza, (2) apunta a un aprendiente de nivel avanzado. Pero la presencia gramatical estaría de igual manera si en lugar de (2) hubiéramos introducido (3), un enunciado simple

(3) La calle está / apareció / amaneció desierta

Ineludiblemente, el abordaje de (3) toma por centro la relación sintáctico- morfológica de la *concordancia*, que determina la buena formación de la estructura. La concordancia, en este caso, lidera la relación entre el sintagma determinante (SDET) *la calle* en asociación con la forma *está* y el predicado *desierta*, en una cadena bien formada de rasgos nominales, esto es, de género, de número y de persona.

Retomando nuestra hipótesis pivote, la respuesta que la avala es sin duda empírica, no es una demostración científica, pero nos alcanza. Ni bien llegamos a descifrar el entramado, constatamos que la oración propuesta en (2) tiene una estructura compleja, como se dijo, por lo que también estamos ante un conocimiento complejo, que facilita al hablante diferenciar categorías, distinguir sustantivos de adjetivos, verbos y preposiciones; que le permite establecer relaciones, configurar diferentes estructuras combinando constituyentes, y ubicarlas en una determinada posición en la cadena oracional; realizar operaciones para destacar un constituyente sobre otros (construcción periferia izquierda de foco por ejemplo), formación de oraciones relativas e interrogativas entre otras; por último, vincular el aparato formal con el léxico, y definir un significado<sup>3</sup>.

(2) y (3) remiten al aparato formal de la sintaxis que marca reglas y restricciones. La gramática, en resumidas cuentas, es un complejo de articulaciones entre las áreas centrales de la *sintaxis* y la *morfología*,<sup>4</sup> del *léxico* y, como resultado, *del significado*. Y dado que adoptamos la hipótesis del conocimiento de la gramática por parte del hablante, se sigue que la práctica de la enseñanza tiene una base fuerte en la reflexión acerca de este conocimiento. Reflexión que se adecuará a si se trata de enseñanza de español L1 o de español L2. Pero tanto en un caso como en el otro, lo que importa es que el docente tenga un dominio de la gramática como para mostrarla, describir los casos que se planteen y pueda responder satisfactoriamente a los requerimientos de los alumnos.

En relación con lo que se acaba de decir, queda aún por considerar el segundo aspecto que hemos introducido, *hacer gramática, hablar sobre la gramática*, que es más o menos decir *reflexionar, discurrir sobre la gramática*, punto central de nuestro trabajo. ¿Por qué central? Porque, dados nuestros objetivos, diseñar la gramática en el ámbito de la enseñanza del español como primera y segunda lengua, requiere *hacer gramática*, en el sentido de describir

3 Cf. "Gramática del español. Un desafío" en *ELE*, 2016, Cap.1:4.

4 En particular, aludimos a la morfología flexiva que en esencia es también una sintaxis o forma parte de ella.

y explicar la gramática del español considerando las áreas mencionadas, y dentro de lo posible, en contraste con otra /s lengua/s. Esto es, convertir la gramática internalizada en la gramática externalizada.

Esta gramática surge a partir de determinados supuestos teóricos y del relevamiento de las evidencias empíricas. Si en la primera entrada, la de *saber gramática*, recurrimos a un conocimiento intuitivo y usamos este conocimiento para poder expresarnos, en el segundo caso reflexionamos sobre ese conocimiento. La reflexión sobre la gramática supone mínimamente analizar y explicitar qué y cómo se forman las estructuras, qué contenido tienen (sean sintácticas o morfológicas), qué lugares ocupan, cómo se construyen y qué las regula, qué combinaciones podemos hacer para producir un significado y para variar los significados y los aspectos ad hoc vinculados teniendo, en cuenta, además, las variaciones de registros del habla.

Volviendo al caso (2), lo reformulamos en (4) y decimos:

(4) Mientras caminaba por la calle, iba pensando en cuál podría ser mi próximo viaje; tengo muchos lugares sin conocer, me decía, y en efecto es así, pero cuál elegir, y con quién hacerlo.

A lo largo de este texto, este hablante expresa un significado a través de distintas formas; ubica el hecho de pensar en las circunstancias témporo-espaciales de *encontrarse caminando por la calle*, piensa en algo que es el tema central, respecto del cual se formula algunos interrogantes. Si bien las *categorías léxicas*, es decir, los verbos, los sustantivos, algunos adjetivos, *próximos*, *muchos*, que recortan a los respectivos sustantivos *viaje*, *lugares*, expresan los significados fundamentales con sus respectivos núcleos y complementos, la construcción total, con una estructura perfectamente articulada, está sostenida por las *categorías funcionales*<sup>5</sup> que realizan esta articulación y que conforman el entramado general, esto es, la expresión de una relación de simultaneidad entre *caminar* y *pensar*, manifestada por la forma *mientras*, las formas verbales seleccionadas, el imperfecto en *camínaba* y en *iba* dentro de

---

<sup>5</sup> Entendemos por *categorías funcionales* los elementos o constituyentes que establecen una relación. Son por ejemplo, los complementizadores que introducen la subordinación, en este caso, sustantiva, *que*, *si*; los relativos, los indicadores de concordancia, de caso; en gran medida, las preposiciones. Las categorías -F son relacionales, y en número finito. Esto las diferencia de las cat-L (léxicas), que se agrupan en verbos, sustantivos, adjetivos, adverbios, y cuyo inventario no es finito ya que por diferentes razones (históricas, sociales, tecnológicas, derivacionales, entre otras) se van incrementando en el día a día, y el hablante puede ir incorporando también en el día a día, a través de la lectura o del intercambio lingüístico, nuevas unidades a su diccionario mental.

una frase de ir + gerundio (ya que no sería lo mismo si dijera *Mientras caminé por la calle*, por ejemplo), el juego de pasado y de presente (*caminaba / iba pensando - tengo*), los coordinantes con sus propios significados (el *pero* adversativo, el copulativo *y*), y la flexión que hace a la concordancia reflejada en la desinencia verbal y en los grupos nominales. Es decir, una combinatoria de operaciones le han permitido a este hablante hilvanar esta construcción oracional que completa un significado, una construcción en apariencia simple, dijimos, pero, sin embargo, con una carga de funciones y de operaciones que la hace compleja si se piensa en quien aprende la lengua (pensemos simplemente, en cómo produce las dos interrogaciones con la anticipación de *cuál* y de *con quién*, y además, en cuál es la referencia de cada interrogativo de cuya relación depende una interpretación adecuada; en un orden similar, véase qué retoma el pronombre *lo* cliticizado al infinitivo *hacer*, etc.). Pero, además, añadimos un aspecto muy relevante en la enseñanza-aprendizaje del español: en tanto el texto no avanza hasta la forma *mi* (en *mi próximo viaje*) no sabemos si se ha introducido una primera o una tercera persona ya que *caminaba / iba* no tienen definición morfológica de persona gramatical (esto es, *caminaba / iba pensando* puede remitirnos a un *yo* o bien a *él, ella, Juana*, etc., como anticipamos en (1)). La cadena *mi - tengo - me* (decía) desambiguan la lectura.<sup>6</sup> Cada aspecto de esta lectura descriptiva gramatical conforma una etapa particular y específica de la gramática, que puede desglosarse a los fines de elaborarla como material de clase (L1/L2). Es decir, un texto compacto como hemos presentado, puede luego abrirse en los diferentes rubros que hacen a la gramática y a partir de él, captado el sentido general del mismo, ir retrabajando con el alumno cada tema gramatical.

Vista desde la perspectiva de un hablante nativo en dominio de un léxico, de una morfología, de una sintaxis, o sea del dominio que sintetizamos diciendo *de las categorías léxicas y funcionales*, resulta una construcción bien formada porque hay una selección adecuada de los complementos, porque el texto está adecuadamente articulado, porque hay correspondencia temporal, porque expresa en su globalidad un significado. Apuntamos al hecho de que uno de los ejes de trabajo se orienta a destacar entre una construcción bien formada, vs una construcción no bien formada, aunque sea interpretable por re-estructuración (proceso que permite el reordenamiento de la oración a los fines de la interpretación) vs una construcción mal formada.

---

<sup>6</sup> Sobre morfología regular del español, véase [www.espanolonlinea.com](http://www.espanolonlinea.com).

- (5) (i) La noche está llena de estrellas.  
(ii) La noche llena de estrellas está.  
(iii) \* La noche están estrellas llena.

(5i) es una oración bien formada ya que hay convergencia de todos los rasgos nominales que se asocian en cadena, definidos por el N *noche* (número, género, persona, como se ve en la terminación verbal). (5ii) puede ser entendida como una variante de (5i), dentro de un determinado registro de lengua. En cambio, (5iii) lleva el asterisco (notación convencional para indicar construcción no –convergente). La transmisión de rasgos es uno de los factores más frecuentes ya sea para determinar la convergencia como la divergencia. Luego, la construcción de (5iii) ha colapsado.

Este punto puede constituir un material específico de elaboración conjunta teniendo en cuenta una combinación de factores gramaticales en juego.

## 2.2. Léxico y sintaxis: acerca de la relación de interfaz

La denominación *interfaz* es suficientemente transparente. El sentido preciso de interfaz es el de contacto, y de por sí expresa la relación entre al menos dos sistemas (dentro de lo estrictamente lingüístico, interrelación entre dos o más componentes: fonológico, sintáctico, morfológico, léxico). La concepción de interfaz está presente en el ámbito de las variaciones, de los micro/ y macro parámetros tanto en cuanto a la relación léxico-sintaxis, como a la interfaz con la FF (*forma fonética*, interpretación del sonido) y CI (*conceptual –Intencional*, interpretación semántico-conceptual)<sup>7</sup>.

La puesta en escena desde esta perspectiva sería, entonces, analizar el objeto lingüístico en ese preciso lugar de contacto, en nuestro caso, un objeto lingüístico organizado entre la sintaxis y el léxico. ¿Cómo proceder metodológicamente en la enseñanza? Hagamos la siguiente suposición: si constituimos a X en un objeto de análisis, no hay inconveniente alguno de abordarlo, por ejemplo, como objeto sintáctico en cuanto a su propia constitución, es decir, considerando la combinación que le dio lugar; en la medida en que estamos trabajando con una concepción de estructuras a partir de un núcleo, (SX=sintagma X), bien podré analizar la expansión de ese núcleo (en diferentes escalas) para constituir ese objeto de

---

7 FF y CI: condiciones impuestas desde fuera de lo estrictamente sintáctico, por los componentes semántico y fonológico



estudio a partir de las posibles *fusiones*<sup>8</sup> que conforman la organización interna de dicho sintagma; o si interesa, también, por movimiento del sintagma, operación que también genera nuevas combinaciones (formación de relativas, de interrogativas parciales, de oración pasiva). Ahora bien, metodológicamente, no es fácil, diría, separar este objeto sintáctico del léxico, ni de los constituyentes funcionales. Así, pues, si las estructuras y las combinaciones se desarrollan a partir de un núcleo y ese núcleo es léxico, puedo entender que estoy por así decir, justo en la coyuntura de los dos módulos, sintaxis y léxico; tal vez, a riesgo de circularidad, puedo ir de la sintaxis al léxico o del léxico a la sintaxis. Sintaxis y léxico se entrelazan. De qué manera planteo el léxico. Supongamos que tomo el N, per se, fuera de un contexto determinado, y mi planteo básico y general va a considerar los NN contables, los NN de masa y los NN eventivos. De aquí puedo formar paradigmas, ya que, para caracterizarlos, procedemos por selección de rasgos, y así, es posible determinar listas con características similares; no obstante, entiendo, casi por necesidad, no vamos a quedarnos en el armado de meras listas, sino que giraremos hacia los comportamientos secuenciales, y por lo tanto a entrar en el ámbito de la sintaxis. Algo similar va a suceder si tomo el V, o el A. En este último caso, por ejemplo, si el tema de abordaje es el de la sufijación en el campo de la morfología derivativa, supóngase que se tome *-ble* como sufijo formador de A, y se proponga *analizable* como tipo; *-ble* selecciona V para formar la palabra derivada, un V marcado por la condición de tener objeto (=V transitivo); lo relevante de la forma derivada es el aporte que hace *-ble* a la sintaxis. Es decir, el sufijo *-ble* es responsable no solo de que la combinación resultante sea un A, sino además, que tenga comportamiento pasivo más un plus de significado modal. Es decir, en el ámbito de la morfología derivativa, un morfema derivativo, que es un constituyente léxico, va hacia la sintaxis, y la determina. Para corroborar y hacer accesible la interpretación, podemos descomponer el derivado en los significados mencionados en la paráfrasis:

(6) (i) X es analizable

(ii) X puede ser analizado (lectura modal de posibilidad más forma pasiva)

Véase que en lo que antecede, hemos incluido V,A,N pero no P. Lo que hay que saber acerca de P es que la preposición es una categoría léxica y al mismo tiempo, funcional. Es léxica en cuanto a que puede aportar significado, caso de *sobre, hasta, durante, por, con*, etcétera. Pero

---

<sup>8</sup> Fusión o ensamble (merge): operación básica del lenguaje que de dos constituyentes (X,Y) se forma uno (SX, por ejemplo) siendo X el núcleo que rotula o da nombre a todo el sintagma.

a la vez es funcional, en la medida en que reúne dos condiciones: número finito (ningún hablante puede incluir nuevas preposiciones ni generarlas) y además, porque es requisito indispensable en lenguas como el español para formar un SP. Esto es, en la evolución y transformación de la lengua desde el latín al español y lenguas romances, en términos generales, se fue dando un fenómeno de reducción del sistema casual predominante en la declinación de sustantivos y (en concordancia) de adjetivos de la lengua latina, y su sustitución por una forma analítica de P + N. Este hecho lingüístico determinó que la estructura de los adjuntos (o constituyentes periféricos indicativos de tiempo, modo, instrumento, causa, y demás) tuvieran mayoritariamente la forma de P +SDET o P +SN (*por la noche, en Rosario, sobre la mesa...*).

Retomando el tema en cuestión, en síntesis, decimos que *interfaz* indica lugares de contacto y en nuestro caso, contacto entre áreas del lenguaje. Si se piensa en las áreas tradicionalmente definidas como fonología, fonética, morfología, sintaxis, en el orden gramatical, y por otro lado el léxico, el trabajo desde una concepción de interfaz nos conduce a considerar que estas áreas no actúan de manera independiente sino en interacción y que, por lo tanto, nuestro trabajo tiene la opción de ubicarse en ese lugar de contacto. Digo *la opción* porque por otro lado, no se descarta profundizar en uno de los ámbitos indicados.

Pero también, y en particular cuando se quiere abordar la producción de *significado*, ¿podemos sostener sin más que con apuntar a los factores lingüísticos hemos completado la tarea de desentrañar la conformación del significado? La respuesta es que muchas veces no. Esto es, que no alcanza con lo lingüístico puro, y que se requiere ampliar el campo para dar cabida a lo que hemos considerado *extralingüístico*, esto es, a apelar a conocimientos conceptuales que explicarían restricciones semánticas, lo que es pensar en principios más generales que rigen la interpretación de lo lingüístico- Esto es, el sistema lingüístico y el sistema extralingüístico -conceptual abren otra instancia de interfaz. Esta cuestión no es de reciente planteo, tiene larga trayectoria, en la que se suman trabajos con propuestas desde diferentes posiciones, desde una posición moderada que tiende a centralizarse en lo lingüístico aun cuando atienda también a lo extralingüístico, a una más extrema y radical, que entiende que lo extralingüístico es omnipresente. Solo para citar un abordaje, BertPeeters, 2000<sup>9</sup> alude a la relación entre conocimiento lingüístico y conocimiento no –lingüístico que, como señala, ha interesado siempre a lingüistas, filósofos, psicólogos, profesores de idiomas, semiólogos,

---

9 *Setting the scene: some current milestone in the lexicon- encyclopedia debate*, Peeters 2000, I. Introduction.

científicos cognitivos. Y para una comprensión exacta o más aproximada de lo que está introduciendo, presenta el verbo *comenzar* + *objeto*, como es el caso de *comenzar un libro / el libro* para mostrar cuál/ es son los conocimientos que se activan. Lo que señala el autor es que *comenzar un libro* de parte de Juan (*Juan comenzó el libro*) tiene la ambigüedad que sólo se puede resolver desde lo extralingüístico, de si Juan ha comenzado a leer o a escribir el libro, considerando que *comenzar* da lugar a construcción con infinitivo; de esta manera se puede justificar que se elija solamente *escribir o leer* y no por ejemplo, *pintar*. Es decir, es el objeto *el libro* el que restringe un menú más amplio de acciones referidas a él.<sup>10</sup>

Agregamos que a los fines áulicos, es adecuado trabajar *comenzar el libro*, como podría ser *terminar el libro*, y diferenciarlo, en paralelo, por ejemplo, de *comenzar /terminar el baile*; dos ocurrencias con construcciones en apariencia iguales, pero orientadas a lecturas diferentes dada la naturaleza del N en posición de objeto, en un caso, objeto contable, de individuo, *libro*, en el otro, N que representa un evento. Entonces, se activan conocimientos lingüísticos a la par de conocimientos extralingüísticos. En este punto reside la interfaz de la que hablamos.

Entonces, la *relación léxico –sintaxis* ha sido central en los estudios lingüísticos específicos, teóricos y prácticos, y de interés en el ámbito de la enseñanza, ni bien se fue tomando conciencia de que no alcanzaba con presentar listas de verbos a modo de inventario, por ejemplo, describiendo sus selecciones y posibles comportamientos, que no alcanza ni siquiera compete describirlos desde la transitividad con la simple marca de TR o de INTR, que la presencia / ausencia / estructura del objeto, entre otras cosas, tiene mucho que decir, específicamente en el caso del español y en otras lenguas, en las que son relevantes los indicadores de significado ubicados en los SSNN / SDET.

Cerramos este apartado con una breve referencia al *significado*. Las lenguas están organizadas en torno a sistemas estructurados, o a *un* sistema estructurado que en todo caso incluye microsistemas. Entonces, desde la perspectiva lingüística, entendemos que el significado se construye en base a las estructuras que resultan de las combinaciones propias de la sintaxis / morfología y del léxico. El entramado sintáctico y morfológico unido al léxico constituyen las formas por medio de las cuales se expresa el significado. Las palabras, en mayor o en menor

---

10 Cabe la aclaración de que también puede justificarse esta ambigüedad sin recurrir a lo extralingüístico, simplemente señalando que *comenzar* se construye con infinitivo y que el objeto definir la pieza léxica que puede entrar en la combinación.

grado, tienen significado. Las palabras se combinan y los objetos resultantes de estas combinaciones vuelven a combinarse, según reglas, operaciones y procedimientos que son descritos por la teoría y que se supone que como hablantes los ponemos en funcionamiento aun sin saberlo. Pero la pregunta que me parece convocante y que constituye un desafío es la de la *visibilidad*, esto es, si en realidad se patentiza, si se muestra cada porción y cada relación de significado, si no hay límites que los determinen y que impidan dar cuenta expresa de todo lo que está presente; vale decir, preguntarse si hay parcelas de significado que no tiene realidad material (fonética o morfológica). En síntesis, los componentes de sentido, los constituyentes estructurales, las combinaciones, la visibilidad que se escinde entre lo visible (materializado) y lo no visible, son los puntos axiales para ir buscando por dónde se cuele el significado. Cuando en el quehacer específico, por ejemplo, del estudio de verbos derivados de sustantivos (verbos *denominales*) tipo *salsear*, *enmantecar*, *acaramelar* se analiza la derivación morfológica, las relaciones en la sintaxis y la semántica derivada, se acuerda en que cada serie es representativa de verbos que expresan *cambio*, que por tal razón son transitivos<sup>11</sup>, que tienen por necesidad un objeto afectado sujeto al cambio de posición o de estado, se constituye un significado complejo que incluye la morfología de la palabra, la sintaxis oracional y componentes de sentido que no pueden ser demarcados en una parte del derivado verbal. A esto apuntamos al hablar de visibilidad<sup>12</sup>.

### 3. CONCLUSIONES

En el recorrido realizado hemos definido algunos puntos centrales de las relaciones gramaticales de interfaz entre sintaxis/ morfología, y léxico, en su conjunto dirigido a la búsqueda de la constitución del significado. Se destaca, fundamentalmente, que el significado se construye en la estructura. Hemos hecho referencia a la interfaz conocimiento lingüístico / conocimiento extralingüístico, aspecto presente en la interpretación. Nuestra presentación está dirigida a la enseñanza; así, pues, hemos tratado de que cada avance en la presentación y exposición de cuestiones gramaticales fueran ilustradas con ejemplos del español, pensando en su posible implementación en la enseñanza de español como L1 y como L2. Se ha hecho

---

11 Esta afirmación surge del hecho de que todo predicado de cambio supone un objeto afectado, esto es, un objeto en el que recae el cambio, sea de estado, sea de posición.

12 Una aclaración: hemos utilizado un par de veces la palabra *palabra* por conveniencia expositiva, para no desviar la atención hacia otra problemática que lleva al cuestionamiento de la palabra como un primitivo de la teoría lingüística o un derivado, en el plano de las unidades. Cf. Carlos Piera.

hincapié en la necesidad de que el docente de lengua no sólo haga una eficiente implementación sino que tenga un conocimiento firme de la gramática de esta lengua. De aquí que la primera subsección la hayamos dedicado a dos cuestiones emparentadas, saber gramática y hacer gramática. En síntesis, entendemos que lo que proponemos es una conjunción de gramática, cuestiones gramaticales, formas de implementación. Cada ilustración puede ser centro de un estudio particular.

## Bibliografía de consulta

- Aldenstein, Adelina (ed) (2012). Interfaces semánticas. Serie 2012. SAL, Argentina.
- Baralo, M. (2005), Aspectos de la adquisición del léxico y su aplicación en el aula. *FLAPEI, Congreso Internacional. "El español, lengua del futuro"*. Toledo, pp. 1-13.
- Beloro, Valeria A. (ed.) (2019). *La interfaz sintaxis –pragmática. Estudios teóricos, descriptivos y experimentales*. De Gruyter.
- Bosque Muñoz, I. (2016). La aplicación de la gramática en el aula. Recursos didácticos clásicos y modernos para la enseñanza de la gramática. *RLA. Revista de Lingüística Teórica y Aplicada*. Concepción, 54 (2), pp. 63-83.
- Brucart, José M. (1996). Gramática y adquisición en la enseñanza del español como lengua extranjera. *ASELE*. Actas, Centro Virtual Cervantes.
- Liceras, Juana M. (2007). Enseñanza de lenguas: ¿Qué sucede cuando la gramática ya “te ha crecido”? *TINKUY*, 4: 49-59.
- Mugica, Nora (2007a). Acerca de la tensión norma –variación lingüística. Sintaxis, morfología, léxico. *Revel, Revista Virtual de Estudos da Linguagem*, Brasil.
- Múgica, Nora (2007b). La interfaz léxico –sintaxis –semántica en la derivación con sufijo –izar. *Signo & Señal*, Nro. 15. Inst. de Lingüística, Facultad de Filosofía y Letras, UBA.
- Múgica, Nora (2011). La periferia y las condiciones estructurales. *Actas de las Segundas Jornadas de Adquisición y Enseñanza del Español como Segunda Lengua*. Fac. de Humanidades y Artes, UNR, 9,10 y 11/09 2010.
- Múgica, Nora (2013). Acerca de la estructura argumental, aspectos constantes y variaciones. *Estudios del Español como Lengua Segunda y Extranjera, serie lingüística*. USAL, 243-269.
- Múgica, Nora (2014a). Sobre el orden de las palabras en latín. *Arquías*. Año 1, Nro. 1. People&Anfusse, Montevideo. 45-68.
- Múgica, Nora (2014b). Qué decimos y hacemos cuando hablamos del léxico. *Revista del español como lengua Extranjera*. Nro. 8. USAL.
- Múgica, Nora (dir.) (2017). La gramática del español. Un desafío. *ELE*. Biblos, Buenos Aires.

- Música, Nora y Mangialavori Rasia, Ma. Eugenia (2016a). Estructuración semántica del Trayecto y visibilidad en la interfaz syn-sem. *En torno a las interfaces del lenguaje en la lingüística generativa*, Vol. Tem. de la SAL. Editado por Nora Música, 41-68.
- Música, Nora y Mangialavori Rasia, Ma. Eugenia (2016b). Los verbos de adjetivales. Categorías léxicas y raíces léxicas. *En torno a las interfaces del lenguaje en la lingüística generativa*, Vol. Tem. de la SAL. Editado por Nora Música, 19-40.
- Música, Nora y Mangialavori Rasia, Ma. Eugenia (2017). Léxico y sintaxis en los adverbios españoles: un estudio a partir de la predicación secundaria. *Revista SAGA*, Escuela de Letras, Fac. de Hum. y Artes, UNR.
- Peeters, Bert (2000). Setting the scene. Recent milestones in the lexicon-encyclopedia debate. *The lexicon - encyclopedia interface. (Current research in the semantics / pragmatics interface)*, 5. Bert Peeters, Oxford. Elsevier Science. 1-52.
- Perea, Francisco Javier Siller (2007). Análisis de errores y enseñanza de las preposiciones de la lengua española como LE. *Revista de Didáctica*.
- Verba Hispánica XXVII (2019). Nuevos retos y perspectivas en la enseñanza /aprendizaje del léxico en ELE. *Anuario de la Sección de Estudios Hispánicos*. Facultad de Filosofía y Letras, Universidad de Ljubijana, Eslovenia.

# **Producción escrituraria, argumental y sintáctica, en alumnos universitarios. Un caso representativo: el empleo de “el mismo” y sus variantes con valor anafórico**

**María Cristina Calvi**

Universidad Nacional de Rosario

**María Susana Freidenberg**

Universidad Nacional de Rosario

sufreidenberg@hotmail.com

## **Abstract**

In this work we deal with some problems that the use of “el mismo” (the same) and its variances in genre and number cause in the comprehension and writing of scholarly texts produced by students.

Our hypothesis is that some possible causes of the inappropriate anaphoric use are grammatical factors, generally within the sentence structure: the repetitious “emphatic” use in the writing production; problems related to referentiality and ambiguity; replacement of complex structures by simple ones, among others. Here we analyze the incidence of these factors in the comprehension and writing of scholarly texts produced by the students of the School of Letters, Faculty of Humanities and Arts (National University of Rosario).

Methodologically, mainly within the frame of the Generative Grammar, we describe with analytic purposes different uses of this syntactic construction in texts chosen from the written productions of our database. Further, an explanation is given to surpass the difficulties observed in the sample works. The process of correction on the students’ papers implies to direct attention to the deficit and then to suggest a reformulation taking into account the diagnosis of errors. The second version, after the given explanations about syntactic rules, shows an acceptable improvement: the 80% of the delivered texts show a better comprehension of the writing strategies and self-regulation of the process.

**Key words:** scholarly writing – sentence syntax – uses of “el mismo” (the same) and its variances



## Resumen

En este trabajo abordamos algunos problemas que el uso de “el mismo” y sus variantes en género y número ocasiona en la comprensión y producción de textos en estudiantes universitarios. Hipotetizamos que entre las posibles causas del uso inapropiado con valor anafórico se encuentran factores gramaticales circunscriptos, generalmente, al ámbito oracional: empleo “enfático” reiterativo en el ámbito de la producción; problemas de referencialización y ambigüedad; empleo de estructuras simples en reemplazo de estructuras complejas, entre otras dificultades. Aquí analizamos la incidencia de estos factores en la comprensión y producción de textos en el marco de las actividades curriculares de una cátedra de la Escuela de Letras, Facultad Humanidades y Artes, U.N.R.

Metodológicamente, principalmente en el marco de la gramática generativa, describimos con propósitos analíticos diferentes usos inadecuados de “el mismo” y variantes en las producciones que constituyen nuestra base de datos. Luego, ofrecemos una explicación de las dificultades observadas en los trabajos. El proceso de corrección señala los déficits para guiar a la posterior reformulación del texto atendiendo al diagnóstico de los errores. El resultado de la segunda presentación, subsiguiente a la explicación de los principios sintácticos, arroja valores de mejoramiento muy aceptables: el 80% produce una nueva versión que evidencia mayor comprensión de las estrategias escriturarias y autorregulación del proceso.

**Palabras claves:** escritura académica – sintaxis oracional – uso de “el mismo” y sus variantes

## 1. INTRODUCCIÓN

Ciertamente, este trabajo surge de la fructífera labor que hemos desarrollado junto a la Dra. Zulema Solana en la Escuela de Letras, Facultad de Humanidades y Artes, durante tres décadas de trabajo conjunto. Además integra la agenda de trabajo del PID 1HUM 541 de la Secretaría de Ciencia y Tecnología (UNR) “Variables de incidencia en la comprensión y producción de textos académicos en alumnos del ámbito universitario“, del cual participa la totalidad del equipo docente de la cátedra Lengua Española III.

La problemática que presentamos aborda algunas dificultades relevadas en textos de alumnos universitarios en el empleo de la construcción “el mismo” y sus variantes, en género y número. Se puede observar que estas dificultades proyectan su incidencia en ciertos aspectos de la comprensión y producción de textos académicos, tal como se evidenció en el marco de las actividades de la cátedra.

En una primera instancia, comprobamos que el empleo de estas construcciones, a las que se les atribuye impropriamente valor anafórico, produce redundancia semántica y ambigüedad referencial. Hipotetizamos que entre las posibles causas de este uso inadecuado se encuentran factores de índole gramatical situados, generalmente, en el ámbito oracional: empleo “enfático” reiterativo en el ámbito de la producción; problemas de referencialización y ambigüedad; empleo de estructuras simples en reemplazo de estructuras complejas, entre otras dificultades.

Metodológicamente, principalmente en el marco de la gramática generativa, describimos con propósitos analíticos los diferentes usos inadecuados de “el mismo” en las producciones que constituyen nuestra base de datos. En un segundo momento, se ofrece una explicación que consideramos podría contribuir a superar las dificultades observadas en las producciones muestrales. Estos déficits, como otros relevados en la producción de los textos requeridos por la cátedra, se señalan y analizan en el proceso de corrección brindando posibles alternancias superadoras. En todos los casos se propone la reformulación del texto presentado atendiendo al diagnóstico de los errores. El resultado de la segunda, y a veces de una tercera presentación, posterior a la explicación de los principios sintácticos que regulan la lengua española, arroja valores de mejoramiento muy aceptables: el 80% de los trabajos ofrece una nueva versión que

da cuenta de un mayor conocimiento de las estrategias escriturarias y la autorregulación del proceso.

El apartado 2 realiza un somero recorrido por trabajos que han planteado problemáticas similares. El apartado 3 expone nuestra hipótesis de trabajo y el 4 brinda una descripción de la base de datos que constituye la muestra. El apartado 5 analiza, en sus tres subapartados, los casos problemáticos. Finalmente el apartado 6 presenta una conclusión y el 7 las referencias citadas en el trabajo.

## 2. ESTADO DE LA CUESTIÓN

La preocupación acerca de la comprensión y producción de textos académicos ha ocupado un lugar relevante en la agenda de las investigaciones de las últimas dos décadas que, con diferentes filiaciones teóricas, abordan la problemática de la escritura en el ámbito universitario.

García Negroni (2011 [1]) releva los temas que han sido objeto de estudio en años recientes: enfatizadores, atenuadores y marcadores de actitud; citas convencionales y referencias a trabajos ajenos; formas de metadiscurso textual e interpersonal de los mecanismos argumentativos. En el marco de las investigaciones de la U.N.R. numerosos trabajos dan cuenta de la incidencia de la sintaxis en la comprensión y producción de discursos: Freidenberg, M.S. & Solana, Z. (1999 [2]); Freidenberg, M.S. (2004 [3]); Freidenberg, M.S. (2003 [4]); Calvi, M. C.; Freidenberg, M.S. y Palillo, S. (2016 [5]); Calvi, M.S. y Freidenberg, M.S (2017 [6]).

En cuanto a las operaciones de reformulación parafrásica realizadas por estudiantes universitarios cabe citar los trabajos de Cuenca (2003 [7]); Bach Martorell (2005 [8]); Muñoz, Muñoz y Mihovilovic (2006 [9]).

En lo referente a los sintagmas “el mismo” y sus variantes en género y número, el tema no ha tenido en los estudios la misma preponderancia que otras estructuras también problemáticas. Desde un enfoque normativo, el *Diccionario panhispánico de dudas* (2005 [10]) aporta el encuadre referido a su uso:

A pesar de su extensión en el lenguaje administrativo y periodístico, es innecesario y desaconsejable el empleo de *mismo* como mero elemento anafórico, esto es, como elemento vacío de sentido cuya única función es recuperar otro elemento del discurso ya mencionado; en

estos casos, siempre puede sustituirse *mismo* por otros elementos más propiamente anafóricos, como los demostrativos, los posesivos o los pronombres personales (230)

Asimismo, la *Nueva gramática de la lengua española* (2009 [11]) retoma estos conceptos para los casos en los que “el mismo” se emplea exclusivamente con valor anafórico perdiendo su significado habitual de identidad e igualdad:

(...) Se recomienda no abusar de esta construcción y sustituirla, cuando sea posible, por otros elementos anafóricos como un pronombre personal, un posesivo o un demostrativo (1207).

Esta temática tiene también mención en *La Gramática Descriptiva de la Lengua Española*, en el capítulo de Carlos Peregrín Otero (1999 [12]) que destaca la importancia que la relación gramatical entre *anáfora* y *antecedente* posee en todas las lenguas. En relación a *mism-* da cuenta de su valor *adjetival* y de cómo contribuye a la interpretación semántica de la frase que lo contiene de dos maneras diferentes según esté antepuesto o pospuesto al nombre. Si bien se centra en el valor reflexivo de *sí mismo*, resultan interesantes sus argumentos a favor de que *sí mismo* posee el mismo análisis sintáctico que *el mismo* o *los mismos*.

Se trata de una clase de construcciones que han sido analizadas como reflexivas, pero que en realidad distan mucho de ser reflexivas en sentido estricto (...) (23.4: 1498)

Desde la perspectiva de la enunciación, el trabajo de Vilchez y Figueroa (2009 [13]) focaliza las dificultades de los estudiantes universitarios que al momento de emplear referencias deíctico-mostrativas y marcadores referenciales recurren al reemplazo de valores sintácticos por procesos de recuperación semántica. De manera que proponen a modo de conclusión un enfoque orientado a la enseñanza mediante el ejercicio de la reflexión metalingüística:

Se recomienda que las anáforas, como variables de orden contextual, sean objeto de enseñanza explícita en el aula, ya que en ellas confluyen relaciones de orden fonológico, morfosintáctico, semántico y pragmático, es decir, todos los niveles de organización lingüística. La recomendación de su enseñanza explícita en el aula parte de la comprobación, mostrada en otros reportes, de que tal conocimiento desarrolla intuición lingüística (o, mejor dicho, metalingüística) al descubrirle al hablante que el tiempo lingüístico (recursivo) es diferente al tiempo físico (lineal) y fundamental para alcanzar la textualidad o cierre acabado de los textos.

Los diversos enfoques de este recorrido permiten situar la problemática aquí abordada en el terreno de la producción enunciativa académica dificultada por el desconocimiento de valores

sintácticos referenciales que puestos en relación con otros valores (semánticos, pragmáticos, contextuales) construyen el evento comunicativo.

### **3. ACERCA DE LA HIPÓTESIS DE TRABAJO**

Nuestra hipótesis es que los problemas referidos tanto a la comprensión como a la producción de textos responden, en gran parte, a la escasa práctica escrituraria de los estudiantes universitarios y al desconocimiento de los principios que regulan el funcionamiento del español como lengua materna. Esta hipótesis surge de las encuestas sobre los conocimientos previos al cursado de la asignatura y del relevamiento de datos de trabajos habituales de los alumnos de la cátedra. Las encuestas permiten observar que tanto la carencia de práctica escrituraria como el desconocimiento de la gramática del español tienen su origen en la escuela media y en el enfoque sesgado que se le da al estudio de la lengua.

En cuanto a la producción, el *uso enfático* de la estructura objeto del presente estudio, “el mismo” y sus variantes, densifica la estructura debido a una inadecuada reiteración con pretendido valor anafórico. Por otra parte, en lo referido a la comprensión, los problemas de “el mismo” y sus variantes se circunscriben al ámbito de la referencialización y ambigüedad debido a la pluralidad de antecedentes potenciales que surgen del uso incorrecto de este sintagma al que se le otorga valor anafórico.

### **4. CARACTERIZACIÓN DE LA BASE DE DATOS**

La base de datos, constituida por 60 trabajos de población estudiantil universitaria, de varios años académicos, presenta cierta heterogeneidad problemática en el uso inapropiado del sintagma “el mismo”. Los trabajos que constituyen esta muestra fueron asignados como tarea domiciliaria a modo de inicio de una práctica de investigación lingüística. En general, la producción consta de textos argumentativos o descriptivos explicativos mediante los cuales se propuso exponer, entre otros, el recorrido de lectura de un hipertexto- como por ejemplo la web de la Facultad de Humanidades y Artes- o argumentar sobre las relaciones que pueden establecerse entre los contenidos de los prólogos de dos gramáticas de la lengua española. El plazo de entrega de estos trabajos es de más 60 días desde el momento en que se comunican las consignas, de manera que el trabajo domiciliar se realice en condiciones que posibiliten la escritura, la revisión y una posible autocorrección sin premuras de tiempo.

## 5. ANÁLISIS DESCRIPTIVO DE LOS DATOS

El relevamiento, análisis y clasificación de las dificultades observadas en los datos se condicionan con tres tipos de problemas: 1) Empleo “enfático” reiterativo en el ámbito de la producción; 2) Problemas de referencialización y ambigüedad y 3) Reemplazo de una estructura sintáctica compleja por sintagmas simples referenciales. De cada uno de los problemas detectados hemos seleccionado solo tres casos para acotar la extensión del análisis.

### 5.1. Empleo “enfático” reiterativo en el ámbito de la producción

Este empleo, desaconsejado por el Diccionario Panhispánico de dudas y la RAE, se presenta como el más recurrente entre los trabajos de la muestra. El uso de “el mismo”, que generalmente reitera un sintagma nominal (SN) antepuesto, produce un efecto enfático, de recuperación semántica innecesaria (Vilchez, Figueroa, 2009), como en los casos de (1). De modo que su elisión es necesaria y aconsejable en beneficio de una simplificación sintáctica obtenida mediante la economía léxica que facilita la comprensión.

(1)

- a. Desde mi punto de vista, en lo que concierne al contenido, este sitio web cuenta con la información fundamental tanto para quien va a ingresar a la Facultad, como para quien está cursando una carrera en la misma, y además para quienes trabajan en ella.
- b. Inmediatamente, después de observar la información general que brinda la parte inicial del sitio, continué por el lado derecho del mismo, donde se encuentran los íconos con los principales accesos que generalmente utiliza un estudiante de la facultad.
- c. Por otro lado, sí me parecieron interesantes secciones como la de “Educación Continua”, donde encuentro el calendario de los cursos actualizado, los costos de los mismos, así como qué es el área de capacitación y educación continua y de qué se encarga el mismo.

En (1) a. el sintagma “*la misma*” que cumple la función de complemento del SP, resulta redundante debido a que el propio contexto predicativo de la oración remite al ámbito de la Facultad (de Humanidades y Artes): (...) este sitio web cuenta con la información fundamental tanto para quien va a ingresar a la Facultad, como para quien está cursando una carrera y además para quienes trabajan en ella. En (1) b. se presenta el mismo caso. El sintagma “*el mismo*”, también complemento de un SP, se puede elidir sin modificar el significado, ni dificultar la comprensión. El texto describe el diseño de la pantalla inicial del

sitio web de la Facultad de Humanidades y Artes de Rosario, por tanto los desplazamientos en la lectura no requieren especificar en cada caso cuál es el soporte. Una alternancia posible para evitar dudas podría expresarse como: “*continué hacia la derecha, donde se encuentran los íconos*”. En este caso, observamos además el uso apropiado de la expresión “continué por la derecha”. En (1) c. al uso innecesario de “*los mismos*” se suma la reiteración en la misma oración de la variante en singular. La proposición oracional describe mediante una enumeración los datos que aporta la pestaña “Educación Continua” de la página web de la Facultad lo que torna redundante el empleo del sintagma “*el mismo*”. La elisión de ambas apariciones no modifica el contenido y simplifica la secuencia descriptiva, tal como se observa en la siguiente reformulación: “*(...) donde encuentro el calendario de los cursos actualizado, los costos, así como qué es el área de capacitación y educación continua y de qué se encarga*”.

## **5.2. Acerca de la comprensión**

### *5.2.1. Problemas de referencialización y ambigüedad.*

El empleo de este sintagma se presenta en un rango de apariciones heterogéneas cuyas referencias, debido al supuesto valor anafórico de “el mismo”, resultan difíciles de determinar por la multiplicidad de posibles antecedentes de la oración que lo contiene.

(2)

- a. En el caso de la GDLE, su prólogo está subdividido en pequeños apartados de tipo explicativo que desarrollan las distintas características de la obra, la historia acerca de su composición, los criterios utilizados y la organización de *la misma*.
- b. La sección “Carreras de grado” está organizada por escuelas y distingue entre profesorado y licenciaturas. Ingresé a “Escuela de Letras”, luego a “Profesorado en Letras” y finalmente retrocedí en el navegador para seleccionar “Licenciatura en Letras”, donde pude comparar los objetivos de *las mismas*.
- c. Debido a la complejidad de este recurso (el ejercicio reflexivo), Negroni y las autoras Telma Piacente y Ana María Tittarelli en sus respectivos artículos mediante datos e investigaciones identifican la poca utilización *del mismo* en los trabajos académicos de estudiantes universitarios.

Los casos de (2) evidencian, como en los anteriores, el abandono del valor adjetival de “el mismo” al otorgársele un valor anafórico impropio que genera ambigüedad. En (2) a. el sintagma “la misma” que en este caso cierra la proposición oracional puede referir indistintamente a “la GDLE”, “la obra” o “la historia”, dado que hay concordancia en género y número con los tres SN y cualquiera de las tres interpretaciones podría ser legítima. Pero la ambigüedad que produce esta referencialización impropia, se evita recurriendo a una estructura enumerativa de las características de la edición de la obra. Se podría apelar al uso de los dos puntos y determinantes posesivos, tal como: “(...) *las distintas características de la obra: la historia acerca de su composición, los criterios utilizados y su organización*”. Cabe comentar aquí que el empleo de los dos puntos se constituye como otra de las dificultades en la producción de los enunciados académicos debido al desconocimiento de los valores de los signos de puntuación. En (2) b. se presenta una situación similar a la anterior; el sintagma “las mismas” que también cierra la proposición oracional puede referir indistintamente a “las carreras de grado”, SN en función sujeto en el segmento inicial de la oración, o a una pluralización que incluya “Profesorado en Letras” y “Licenciatura en Letras”. También es posible, forzando la interpretación, incluir “Escuela de Letras”. Si optamos por interpretar que la expresión final a la que se le otorga valor anafórico refiere al antecedente “carreras de grado”, con la que concuerda en género y número, deberíamos reformular la oración reiterando los SN lo cual desambigua al aportar referencias específicas: “*La sección “Carreras de grado” está organizada por escuelas y distingue entre profesorados y licenciaturas. Ingresé a “Escuela de Letras”, luego a “Profesorado en Letras” y finalmente retrocedí en el navegador para seleccionar “Licenciatura en Letras”, donde pude comparar los objetivos de las carreras*”. En (2) c. la relación establecida entre el antecedente y el sintagma “el mismo” posee una interpretación más compleja. La construcción “el mismo” no puede referir a “sus respectivos artículos” pues no hay coincidencia de rasgos en cuanto al número aunque sí al género, pero tampoco es factible una interpretación lógica legítima. La distancia con el antecedente del cual se predica es altamente significativa dado que la expresión es “el ejercicio reflexivo” que se menciona en el inicio del párrafo anterior. En este caso se debió haber reiterado la expresión a modo de aposición dada que la distancia entre la primera aparición del sintagma y su reiteración es muy extensa: “*Debido a la complejidad de este recurso -el ejercicio reflexivo-, Negroni y las autoras Telma Piacente y Ana María*



*Tittarelli en sus respectivos artículos mediante datos e investigaciones identifican la poca utilización en los trabajos académicos de estudiantes universitarios”.*

### *5.2.3. Reemplazo de una estructura sintáctica compleja por sintagmas simples referenciales*

En la muestra se observa una baja frecuencia del empleo de estructuras subordinadas, particularmente de oraciones de relativo, cuando se trata de predicar más de una característica respecto de un SN. En su reemplazo se recurre al empleo inapropiado del sintagma “el mismo” y sus variantes. En otros casos, hay un déficit en la selección de un sintagma simple que determine o referencie al SN al que se alude.

(3)

a. En la (columna) de la izquierda hallamos las últimas cuatro noticias del sitio destacadas, cada una contiene una imagen, título de la nota y la fecha de publicación; tanto la imagen como el título contienen enlaces ocultos hacia la noticia correspondiente, además de un enlace exclusivo para ampliar *la misma*.

b. Del lado izquierdo de la página se encuentra el área de noticias. Allí se pueden leer los resúmenes de *las mismas* acompañados por imágenes ilustrativas, debajo de cada resumen hay un link que dice “Leer más” para acceder a la noticia completa.

c. Al volver otra vez a la opción “Estudiantes”, ingresé a la subpestaña “Calendario Académico”, la cual es de mucha utilidad para los estudiantes, ya que permite organizarse de acuerdo a cómo están distribuidas las actividades y fechas de *las mismas* a lo largo del ciclo lectivo.

En (3) a. el sintagma “la misma ” refiere a “la noticia” como cierre de la oración, un empleo que se reitera en los datos de la muestra y que pone de manifiesto que el autor/ra intuye que es necesario algún elemento sintáctico para explicitar la referencia. Sin embargo no se recurre, como una posible alternativa, a la construcción de una subordinada de relativo que evite este uso: “*En la (columna) de la izquierda hallamos las últimas cuatro noticias del sitio destacadas, cada una contiene una imagen, título de la nota y la fecha de publicación; tanto la imagen como el título contienen enlaces ocultos hacia la noticia correspondiente, además de un enlace exclusivo que amplía la información*”. En (3) b. “las mismas ” refiere a “noticias”, SN situado en la oración anterior. Asimismo la oración se inicia con un pronombre adverbial “allí” que alude a “el área de noticias”. En ambos casos la selección léxico-

sintáctica no parece ser la más apropiada. Si recurrimos al empleo del nexo subordinante “cuyo”, que aporta rasgos de género y número y a la coordinación de dos oraciones subordinadas relativas, se puede evitar ambos empleos poco aceptables: *“Del lado izquierdo de la página se encuentra el área de noticias cuyos resúmenes están acompañados por imágenes ilustrativas y que vinculan al link “Leer más” para acceder a la noticia completa”*. Por último, la oración de (3) c. presenta otro empleo de “las mismas” que puede sustituirse por una expresión adverbial que no modifica el significado pero que evita la estructura a la que se le atribuye valor anafórico: *“(...) ya que permite organizarse de acuerdo a cómo están distribuidas las actividades cronológicamente a lo largo del ciclo lectivo”*.

Las reformulaciones que se proponen para superar los déficit escriturarios no poseen la condición de únicas, solo ejemplifican cómo con un mayor conocimiento de los valores funcionales del léxico, las posibles alternancias con otras estructuras sintácticas, principalmente las complejas, es factible producir un texto con calidad y precisión académica.

## **6. CONCLUSIONES**

Esta presentación avala la hipótesis sobre la incidencia que la sintaxis posee en la comprensión y producción de textos escritos. En primer lugar, a partir de la observación detallada de las producciones de los alumnos fue posible visualizar redundancias, contradicciones y otros mecanismos que el uso inadecuado del sintagma “el mismo” y sus variantes, produce en la comprensión y producción de un texto. En un segundo momento, a partir de la descripción de los problemas, se pudieron establecer generalizaciones que, en gran medida, posibilitaron tipificar las ocurrencias de los sintagmas objeto de estudio. Concluido este análisis, se abordaron en las clases los temas problemáticos emergentes para luego exponer los fundamentos teóricos-metodológicos que vehiculizan la etapa explicativa y posterior corrección. Como lo recomiendan Vilchez y Figueroa (2009), es necesaria la enseñanza explícita de algunas estructuras. El diagnóstico de los errores consiste en esa explicitación que promueve la reflexión metalingüística de los principios específicos de la lengua española, eje fundamental de la cátedra.

La última etapa se aboca al proceso de re-trabajo por parte de los alumnos. En particular, se enfatiza en lograr la precisión sintáctica en la construcción de los enunciados explicativos-argumentativos. El proceso de revisión y corrección concluye con la entrega de una segunda,

y a veces una tercera, versión. La reformulación pone al descubierto la importancia que el conocimiento de la estructura gramatical de un idioma, incluida L<sub>1</sub>, posee para el aprendizaje de la lengua escrita, el mejoramiento de la comprensión lectora, la comprensión de cuestiones normativas, la enseñanza del sistema lingüístico en general y el desarrollo de otras habilidades cognitivas.

## Referencias

- [1] García Negroni, M. M. (Coord.) (2011). *Los discursos del saber. Prácticas discursivas y enunciación académica*. Editoras del Calderón, Buenos Aires.
- [2] Freidenberg, M. S. y Solana, Z. (1999). Incidencia de la estructura sintáctico-morfológica en la comprensión lectora. *Encrucijada del Lenguaje: Oralidad, Lectura y Escritura, Serie Lingüística y Educación*, (1): 166-186.
- [3] Freidenberg, M. S. (2004). Algunos aportes para optimizar la comprensión de textos, *Revista de Letras*, 9: 87-111, Volumen de Estudios Lingüísticos.
- [4] Freidenberg, M. S. (2003). El componente lingüístico en la comprensión de textos. *Revista de Letras*, 8: 63-80, Volumen de Estudios Lingüísticos.
- [5] Freidenberg M. S., Calvi, M. C. y Palillo, S. (2016). Producción de textos de elaboración compleja, *X Jornada de Ciencia y Tecnología de la Universidad Nacional de Rosario*, UNR Editora.
- [6] Freidenberg M. S., Calvi, M. C. y Palillo, S. (2017). Producción de textos de elaboración compleja. *Valencias en el uso del paréntesis, XI Jornada de Ciencia y Tecnología de la Universidad Nacional de Rosario*, UNR Editora.
- [7] Cuenca, M. J. (2003). Two ways to reformulate: a contrastive analysis of reformulation markers. *Journal of Pragmatics*, 35 (7):1069-1093.
- [8] Bach Martorell, C. (2005). Los marcadores de reformulación como localizadores de zonas discursivas relevantes en el discurso especializado. *Debates terminológicos*, 1: 1813-1867. Red Iberoamericana de Terminología.
- [9] Nuñez, P. Muñoz, A. y Mihovilovic, E. (2006). Las funciones de los marcadores de reformulación en el discurso académico en formación, *Revista Signos*, 39 (62): 471-492. Pontificia Universidad Católica de Valparaíso.
- [10] Real Academia Española (2005). *Diccionario Panhispánico de Dudas*. Santillana, 429-430.
- [11] Real Academia Española (2009). *Nueva gramática de la lengua española*. Espasa, 1207.

- [12] Peregrin Otero, C. (1999). Pronombres reflexivos y recíprocos, en Demonte, V. y Bosque, I. (Coord.), *Gramática Descriptiva de la Lengua Española*, Espasa, 1 (23): 1427-1518.
- [13] Vilchez, M. y Figueroa, V. J. (2009). La deixis y la foricidad como señales de anclaje enunciativo. *Revista Núcleo*, 21 (26). [http://ve.scielo.org/scielo.php?script=sci\\_arttext&pid=S0798-97842009000100007](http://ve.scielo.org/scielo.php?script=sci_arttext&pid=S0798-97842009000100007)

## **CAPÍTULO III**

### **Análisis del Discurso**

# **Lenguaje Cliché y disposición escénica: la perspectiva de Alfred Lorenzer y Víctor Klemperer**

**Diego Alberto Beltrán**

Universidad Nacional de Rosario. Facultad de Humanidades y Artes. Instituto Universitario Italiano de Rosario. Facultad de Psicología

Rosario, Argentina

[diegoabeltran@yahoo.com.ar](mailto:diegoabeltran@yahoo.com.ar)

## **Resumen**

A los 16 años mis ojos se detienen en un libro de Jean Paul Sartre situado en un estante de su biblioteca; “Huracán sobre el azúcar”. Le pregunto “¿de qué trata?”... y me habla sobre el viaje de Sartre a Cuba en 1960, de cómo analiza la revolución en marcha y de los eventos de la misma... Recuerdo el entusiasmo de su explicación que se traslada a mi lectura y entonces las escenas políticas enterradas por la dictadura salen al dominio del día. La luz gris-densa cotidiana se transforma en otra cosa benéfica. Aquel es el momento de la mayoría intelectual...gracias por todo...

## 1. INTRODUCCIÓN

Entendemos por lenguaje cliché, en un nivel sub-teórico, a aquellas frases, sintagmas o palabras que una persona repite con cierta compulsión sin reflexionar demasiado en ellas. Este artículo es una línea derivada de un proyecto de investigación denominado “Los enunciados cliché, el lenguaje sustitutivo y la declinación del sujeto reflexivo: un estudio instrumental de caso”. El objetivo general del proyecto y el de este mismo escrito es analizar las expresiones del lenguaje que impiden la reflexión sobre las propias acciones y las de los demás específicamente en el plano ideológico-político y en diferentes contextos históricos. Nos centramos por un lado en el lenguaje utilizado durante el Tercer Reich. No podemos hablar de argot o jerga del Tercer Reich porque no sólo la usaban los cuerpos militares y políticos de los nazis o sus simpatizantes; sino también las propias víctimas del régimen o quienes estaban en su periferia. Por otro lado, nos enfocamos en un abordaje psicoanalítico muy particular realizado por Alfred Lorenzer en la suposición de que las frases cliché de la política puedan tener algún nexo con el concepto de *cliché* y *lengua privada* acuñados por este psicoanalista alemán. El desarrollo del proyecto aún está en su fase de experimentación y creación de teoría. Para evidenciar una de las vertientes del proyecto que desarrollaremos aquí partiremos de un episodio muy particular de la política argentina que ha sido tomado en broma pero que deja indicios importantes del rumbo que deseamos tomar. En una escena política de la pandemia hay una manifestación en contra del gobierno de Alberto Fernández y un periodista es interceptado por una de las manifestantes con la que mantiene el siguiente diálogo:

Periodista: [comentando un reportaje anterior] algunas de las opiniones de las personas que se van arrimando...

Señora: no queremos ser Valenzuela queremos ser libres

Periodista: ¿no quiere ser Venezuela?

Señora: [con énfasis] Valenzuela

Periodista: ¿Valenzuela?

Señora: [nuevamente con énfasis] Valenzuela...queremos ser libres, la libertad se defiende con todo...la bandera [exhibiendo una bandera argentina] está adelante...

Periodista: la bandera primero...muchísimas gracias[1].



El periodista y los manifestantes exaltados por el peligro de que “Argentina se transforme en Venezuela” están en una escena política y la señora parece estar en una escena relacionada a un país, ente o persona quizás real si se refiere al apellido de una persona (aunque el contexto en el cual se lo decía no indicaba tal cosa) y definitivamente inexistente si se refiere a un país (aunque hay una ciudad paraguaya de ese nombre de alrededor de siete mil habitantes). Este “caso” es anfibio en el sentido de su doble dimensión personal-subjetiva y política:

- (a) por un lado la manifestante parece manejarse con un término con un significado sólo comprensible para ella, dicho de otra manera, parece emplear un término proveniente de un reservorio semántico de habla privada.
- (b) por otro lado ese habla privada es enunciada en un escenario político público y conectada a una serie de opiniones políticas referidas a la relación entre Argentina y Venezuela.

En (a) nos encontramos con un fenómeno interpretable desde la perspectiva de Alfred Lorenzer. La señora enuncia un significante de origen desconocido que reitera convencida (seguramente varias veces en ese escenario) relacionado a un lenguaje privado y anclado a una escena olvidada y excomulgada de la conciencia; tenemos aquí a un cliché individual [2]. En (b) nos encontramos con un imaginario de la coyuntura política mundial que muestra una falla tectónica: los significantes Venezuela, Valenzuela y el peligro de que Argentina se transforme en alguno de ellos remite a un quiebre o declive pronunciado de las reglas para emplear, articular y verificar enunciados o conjunto de enunciados que se refieran a las cosas (en este caso a las cosas políticas)[3]. Los objetivos de este artículo son:

- Analizar la potencialidad heurística del concepto cliché enunciado cliché para el abordaje de la violencia creciente del lenguaje político actual a través de su uso peculiar en la Alemania nazi.
- Delinear los posibles factores que determinan la efectividad e intensidad de este lenguaje muchas veces homologado o lindante al concepto de *posverdad*.
- Determinar algunos rasgos estructurales de larga duración de este tipo de lenguaje hablado en los escenarios políticos de muchos países europeos y del continente americano.

## 2. CLICHÉ Y ESCENA PRIVADA

Para Alfred Lorenzer, en todo relato del analizante frente al analista se actualizan tres tipos de escenas; las vividas en el pasado, en el presente y en la situación analítica. En el caso de determinadas patologías, por ejemplo neurosis, el lenguaje utilizado por el analizante tendrá ciertas características particulares y un sentido solo comprensible para este y difícil de decorticar para el analista. El paciente estará hablando una *lengua privada* frente al lenguaje público que habla la mayoría. La diferencia lengua privada/lengua pública tiene su anclaje en la diferenciación entre el símbolo y el cliché realizada en «El lenguaje destruido y la reconstrucción psicoanalítica». La capacidad para simbolizar es exclusiva de la especie humana y se caracteriza por poder evocar imágenes, pensamientos o conceptos sin la presencia de un estímulo inmediato a diferencia de los registros mentales necesitados de esta inmediatez y que compartimos con las otras especies [4]. Los símbolos pueden ser «catectizados», es decir, se les adhiere energía de los procesos inconscientes:

1. Los representantes son símbolos, es decir, productos de un proceso de formación de símbolos.
2. Los representantes son elementos de los procesos de catexis; en ellos, en cuanto mundo interior del paciente, se desenvuelven las catexis pulsionales.

Resulta natural combinar ambas posiciones y concluir:

3. Las catexis pulsionales se desenvuelven en símbolos [5].

Hay un tipo de representantes, los «representantes inconscientes», que Lorenzer denomina clichés. Estos últimos son representantes simbólicos formados durante el proceso de socialización aunque “‘ex - comunicados’ en el curso de la represión” y, por ende, desplazados de la comunicación lingüística pero vigentes en cuanto a la energía pulsional y su capacidad para afectar conductas[6]. Las formas de subjetividad son determinadas por los procesos de socialización (situaciones de interacción) dentro de los cuales se forma el lenguaje (en un sentido ontogenético) a través de la conformación del símbolo y el cliché. El cliché designa una situación de interacción o escena:

Así el niño en presencia de la madre que pronuncia la palabra madre, registra que *madre* no designa solamente a la persona de la madre, sino también a la propia escena de interacción del niño con la madre... Se construye así el cliché que designa una imagen una disposición escénica[7].

A medida que el niño se desarrolla; los clichés surgidos de las distintas situaciones de interacción se van transformando en símbolos lingüísticos. Cada cliché se produce a través de interacciones particulares de cada individuo: es decir que tienen un sentido específico para cada persona y sólo es comprendido por ella. Por lo tanto, Lorenzer plantea que los clichés constituyen una *lengua privada*. Las características del cliché son:

- igualdad entre símbolo y objeto (es como si la palabra *perro* ladrara)
- automatismo de acción/compulsión a la repetición (el cliché no pasa por el proceso de pensamiento; un individuo repita sin darse cuenta de que está repitiendo).
- Irreversibilidad del cliché: dado que no hay tiempo en el inconsciente no es posible establecer diferencias; lo diferente se transforma en lo igual
- no sufre el efecto de otras representaciones; una vez en el inconsciente las representaciones adheridas al cliché actúan como si las otras no estuviesen y tienen carácter alucinatorio [8].

En el proceso normal de desarrollo infantil los clichés se tornan símbolos. Es decir; dejan de representar una disposición escénica específica y pierden, en consecuencia, su status de lenguaje privado. Esta pérdida va a permitir la inserción del individuo en la cadena significante posibilitando su participación en una lengua pública. Si el pasaje de la lengua privada a la pública no es fluido o se da en forma incompleta genera incomunicación, distorsión del lenguaje y neurosis. A diferencia del cliché, en el símbolo hay autonomía entre símbolo y objeto, se puede introducir la reflexión y el pensamiento y existe adaptación, flexibilidad y reversibilidad [9].

## **2.2. El pequeño Hans y la privatización del significado**

En esta historia clínica Freud se ocupa de la fobia a los caballos sufrida por un infante que tiene miedo a que lo muerda un caballo; para Lorenzer el lenguaje “como instrumento verbal del entendimiento está falseado” [10]. Los contenidos reprimidos por el sujeto, además de ser excluidos de la comunicación y de “la ensambladura de símbolos objeto de entendimiento intersubjetivo previo” son reincorporados con un rodeo en la forma subrepticia de un “lenguaje privado seudocomunicativo” [11]. El falseamiento se produce cargando o sobrecargando de significado al significante *caballo* y vaciando de sentido al de *padre*. Esta circulación de significado unidireccional es posible por dos escenas: la del caballo de tiro que

cae y patalea y la protagonizada por su amigo Fritzl, que se lastima y se cae cuando ambos jugaban al juego del caballo en el que Hans hacía de jinete. Para Freud hay una fluida vía asociativa que va de Fritzl al padre y además Hans acepta la hipótesis de Freud en relación a que detrás del miedo a que el caballo muerda y/o se caiga está el miedo al padre. De hecho Hans desarrolla un juego actuando la identificación padre-caballo al morder al padre en tono juguetón. La fobia a los caballos no lleva a una “virtud traumática” o a una suerte de disposición traumática sino al “significado anterior del caballo como objeto de la predilección y del interés”[12]. Es el azar el que liga el miedo al padre con el miedo al caballo generando la posterior desimbolización de los significantes *caballo* y *padre*. La privatización del lenguaje al nivel intermedio entre el individual y el público, es decir, la región del segmento o grupo social; es abordable desde el instrumental teórico de Lorenzer.

Podríamos inferir que, a nivel mundial, el significante *justicia* ha sufrido un proceso de desimbolización pasando de significar la práctica de un procedimiento racional-legal para llegar a la verdad sobre un hecho e impartir la ley a ser visto como una práctica social que incluye “linchamientos y tentativas de linchamientos; en los casos de vindicación o justicia por mano propia; los escraches públicos; los incendios intencionales de viviendas; los escraches en las redes sociales o con carteles en los lugares de trabajo” o también en los *leading case* de los que se ocupan los medios masivos transformados en tribunales televisivos. Este nuevo (o no tanto) significado del término justicia que adquiere resonancias pre-modernas o directamente arcaicas es analizado por Esteban Rodríguez Alzueta para la *justicia vecinal* en la Argentina [13]. El vaciamiento de sentido o desimbolización del término justicia es aprovechado por otros significantes como *justicia vecinal*, *por mano propia*, *expeditiva* que toman la carga semántica de *justicia* aunque modificada: se produce entonces una balcanización del término por medio de la cual justicia pierde fuerza semántica y adquieren representatividad los significantes parciales mencionados. La disposición escénica de *justicia* está anclada a las escenas de sus términos derivados y fragmentados: grupos de personas indignadas que piden una ejecución expedita de la ley a partir de una certeza no apoyada en datos ni en dispositivos de indagación. Estaríamos, en este ejemplo, no en una privatización del significado sino, más bien, en una balcanización del mismo.

### 3. ENUNCIADO CLICHÉ Y ESCENA HISTÓRICA

Víctor Klemperer, ciudadano alemán de origen judío y docente universitario; utilizó el período del gobierno nacionalsocialista para realizar un estudio etnográfico de la terminología por medio de la cual se expresaban los Aparatos de Estado, los simpatizantes nazis y, en general, toda la sociedad alemana (incluso las propias víctimas del régimen nazi). Klemperer es expulsado de la universidad, se le prohíbe acceder a libros de las bibliotecas públicas, es forzado a trabajar en una fábrica y se lo recluye junto con su familia en una *casa de judíos*. Puede eludir el campo de concentración por estar casado con una esposa “aria” y por la misma razón está última puede proporcionarle documentos y textos extrayéndolos a su nombre de la bibliotecas. Una tranquilidad de lo que podríamos llamar “ánimo científico” le permite a Klemperer analizar expresiones, denominaciones y sintagmas que fueron acuñados, popularizados o resignificados durante el Tercer Reich. A esta forma particular de usar la lengua alemana la denomina *Lingua Tertium Imperii* entendiéndola por ella a la lengua(o jerga) utilizada durante el Tercer Reich:

Observaba cada vez con mayor precisión cómo charlaban los trabajadores en la fábrica y cómo hablaban las bestias de la Gestapo y cómo nos expresábamos en nuestro jardín zoológico lleno de jaulas de judíos. No se notaban grandes diferencias; de hecho, no había ninguna. Todos, partidarios y detractores, beneficiarios y víctimas, estaban indudablemente guiados por los mismos modelos. Trate de captar esos modelos, cosa sumamente fácil en cierto sentido, pues todo cuanto se decía y se publicaba en Alemania respondía a las normas del Partido; lo que de alguna manera se desviaba de la forma permitida no llegaba al dominio público, libros, periódicos, formularios y escritos oficiales, todo flotaba en la misma salsa parda, y la absoluta uniformidad del lenguaje escrito explicaba también la homogeneidad del lenguaje hablado [14].

Klemperer plantea 3 tesis básicas:

- La LTI era hablada por víctimas y victimarios; es decir, hegemonizaba la forma de expresarse (y por ende la forma de pensar la propia situación individual y social) de todo el mundo.
- El habla del Reich estaba orientada u organizada por “modelos” o formas de habla específicos.

- El aparato partidario del Reich manejaba y proponía estos modelos que explicaban la homogeneidad absoluta del lenguaje escrito que; a su vez, producía como efecto la uniformidad del lenguaje hablado.

A dichas tesis podemos agregar una proposición teórica nuestra:

- Los modelos o formas de habla pueden verse como grupos de algoritmos que indican con precisión que enunciados discursivos o sintagmas utilizar según los contextos y problemas por los que atraviesa el hablante.

#### 4. GEFOLGSCHAFT Y GEFOLGSCHAFTSSAAL

Durante el siglo XIX se produce una polémica historiográfica. De un lado estaban los germanistas (nacionalistas y románticos) y del otro los romanistas (cosmopolitas e individualistas). El punto de discusión era sobre los orígenes germánicos o románicos del feudalismo. Ambos grupos estaban en lo que podríamos llamar *meta-paradigma* que analizaba el problema como una cuestión de desarrollo jurídico unilineal desde los orígenes de las relaciones sociales feudales. Los germanistas situaban en el Origen a la *Gefolgschaft* (Séquito/Comitatus) de las tribus germánicas. Los romanistas situaban este Origen en el *Patrocinium* o Estado de Patronato[15]. A finales del siglo XI había dos tipos de unidad de señorío que, junto con otras tres formas de organización política; constituían “una estructura laxa, compleja y traslapante” [16]. Una variante de unidad de señorío surgió cuando los colonos (y otrora integrantes de las tribus nómadas de Asia occidental que invadieron el Imperio Romano) “se encomendaban” a los individuos poderosos y como contraprestación recibían alimentos, ropa y protección contra sus enemigos. La persona que se encomendaba era “el hombre” del señor poderoso. El encomendado podía vivir en el aposento de su señor o también trabajar tierras que el señor le otorgare. Los germanistas creían que en el origen de esta unidad de señorío estaba la *Gefolgschaft* el *Comitatus* descrito por el historiador romano Cayo Cornelio Tácito en su obra etnográfica *La Germania*. Este nos habla de unidades políticas construidas con un tipo de lealtad superpuesta a la lealtad de las unidades familiares [17]. En la sección XIII de *La Germania* nos encontramos con la descripción de un rito de iniciación que subtrae al adolescente del ámbito del *Oykos* y lo sitúa en un locus guerrero que, desde el concepto de Estado Nación moderno, podemos verlo como simultáneamente público y privado. El rito de iniciación, que consiste en entregar dos armas

(frámea y escudo) al joven por parte de algunos de los jefes, el padre o algún pariente; le otorga el estatus jurídico de ciudadano pues la frámea “es para ellos su toga” y el escudo “el primer honor de la juventud” [18]. La relación social concreta en la que se ven inmersos estos nóveles guerreros es la del *Comitatus* en la que grupos de estos jóvenes siguen a un señor. Es decir forman parte del séquito de este señor: aquí encontramos el núcleo de dicha relación social; fidelidad del seguidor y protección o apadrinamiento del señor. Es difícil ver el *Comitatus* como un rito de iniciación ciudadana dado el vínculo de sumisión que aparece en el novel guerrero. De todas formas Tácito subraya el pasaje, vía rito iniciático; de lo familiar privado (*Oykos* griego) al ámbito público:

... *ante hoc domusparsvidentur, moxreipublicae*[19].

...antes formaban parte de una familia; después ya son de la república una ilustre cuna[20]

...antes de esto se les veía como parte de la casa, ahora, de la colectividad [21]

Este pasaje adquiere sentido hermenéutico si tomamos la totalidad del escrito monográfico que nos muestra una sociedad fragmentada similar a *Los Nuer* de Evans Pritchard.

#### **4.1. El rito de iniciación: de la escena arcacia a la escena fabril**

*La Germania* nos describe una región delimitada por los ríos Rin y Danubio que separa a los germanos de los galos, los retos y los panonios (región de la Galia, el Tirol y el norte de los Balcanes); las montañas y el “mutuo recelo” los separa de los sármatas y los dacios (zona de la actual Polonia y República Checa); lo que resta del territorio queda enfrentado al océano. En esta región con una cantidad considerable de tribus dispersas; cada una de ellas está formada por unidades familiares (*sippe*) y por el *comitatus* o séquito de jóvenes guerreros (una parte plebeya y otra noble) que siguen a un *señor*. Cada *señor* compite con los otros por tener la comitiva más grande y los integrantes de cada séquito compiten entre sí por escalar posiciones en un tipo de relación relativamente estratificada. De todas formas; la misión de cada séquito es añadir su propia gloria a la gloria del *caudillo*. Existe una relación de liderazgo aunque no de masas sino un liderazgo mucho más pequeño que podríamos llamar *liderazgo de horda*:

En la batalla es una vergüenza para el señor ver su valor superado, y para los seguidores no adecuarse al del señor. Pero lo realmente infame para ellos es retirarse del combate sobreviviendo

al mismo: defenderle, protegerle y añadir las propias gestas a su gloria es el principal deber. Y es que los señores luchan por la victoria, y sus seguidores, por su caudillo [22]

Cum ventum in aciem, turpe principivirtutevinci, turpe comitatuivirtutem principis non adaequare. Iam vero infame in omnem vitam ac probro sum superstitem principis uo ex aciere recessisse. Illud defendere, tueri, sua quoque fortia facta gloriae eius adsignare praecipuum sacramentum est. Principes pro victoria pugnant, comites pro principe [23]

El liderazgo de los *principes* (plural del vocablo latino *princeps*) estaba basado en el ejemplo dado en el campo de batalla y en la admiración que suscitaba en sus seguidores y desembocaba en una relación de fidelidad, sumisión “micropolítica” de cada integrante del *comitatus* hacia el *señor* o *princeps* e intercambio de servicios y favores (protección del señor, reparto generoso del botín de guerra y lucha militar por la gloria del señor por parte de la comitiva). En el rito de iniciación juega un papel importante lo que podríamos llamar *parentesco político*. En el locus “estatal” o de “gobierno central” de los germanos la asamblea o *concilium* “constituye el consenso más honorable” [24]: “Honoratissimum ad sensum genusest armis laudare” [25]. En ella se asiste armado y si el discurso del rey adquiere aceptación y es exitoso los asambleístas “agitan sus frámeas” a modo de aplauso [26]. En una asamblea de estas características se lleva a cabo el rito de iniciación a lo que anacrónicamente podemos llamar *ciudadanía guerrera* y simultáneamente a un estado de adultez para sumergirse en relaciones sociales más allá del núcleo familiar primario. El hecho de que las armas las pueda entregar un pariente ficticio o un *princeps* o señor poderoso que adopte al joven guerrero y lo sitúe en un vínculo de parentesco-sumisión evidencia el carácter arcaico de este *lazo social pre o protopolítico*. Cuando un germano de la monografía etnográfica de Tácito dice *comitatus* o más bien *gefolgschaft* y plantea enunciados con esta o palabras similares nos encontramos con un discurso que regula un específico tipo de relaciones laboral-militares: saqueo de otras tribus guerreras, agricultura que en términos antropológicos es de “saqueo del suelo” (o de quema de bosques y vegetación para cultivar y abandonar luego en busca de nuevas tierras para aplicar el mismo sistema), caza y ganadería. El traslado de un término utilizado por ciertas tribus en la periferia del imperio romano en el 98 DC a su uso en un contexto capitalista de primera mitad del siglo XX crea una *armadura ideológica*. Es decir; se exhuma una terminología y un discurso o representación social ad hoc sin conexión concreta con las relaciones laborales obrero-patronales para vivir las mismas como Don Quijote vivía su relación con los molinos de viento: personas y cosas eran



transfiguradas creándose una realidad alternativa con efectos concretos sobre la “realidad real”:

¡Séquito ¿Quiénes eran, de hecho, las personas allí reunidas? Eran obreros y empleados que cumplían con ciertos deberes a cambio de un salario determinado. Entre ellos y sus patronos todo estaba regulado legalmente; es posible, aunque fuera innecesario y quizá incluso molesto, que existiera cierta relación afectiva entre los jefes y algunos de ellos. Sea como fuere, el comportamiento de todos estaba regulado por la ley fría e impersonal. En la *Gefolgschaftssaal*, sin embargo, eran sacados de la claridad de las normas legales y disfrazados y transfigurados mediante una única palabra: *Gefolgschaft*. Esta los revestía de antigua tradición alemana, los convertía en vasallos, en hombres armados y obligados por un juramento de lealtad, en miembros del séquito de los nobles caballeros. ¿Era tal disfraz un juego inofensivo? En absoluto. Deformaba una relación pacífica para darle un cariz bélico; paralizaba la crítica; conducía por vía directa al principio inherente a la frase que ostentaban las pancartas: ‘¡Fuhrer, ordena y te seguiremos!’ [27]

El capítulo XXXIII de la obra de Klemperer citada inmediatamente más arriba tiene el título de *Séquito*. Esta palabra parece contener y configurar otras palabras del mismo tipo que, a su vez, posibilitan o sugieren determinadas prácticas sociales transfiguradas. En todas ellas el leitmotiv es lo afectivo, instintivo y/o pulsional. En la fábrica en la que trabajaba Klemperer (y suponemos que en todas las empresas del Reich, dadas las disposiciones de la Ley para la ordenación del trabajo nacional de Enero de 1934) el significante *séquito* está anclado a una disposición escénica particular: una sala con un púlpito y un piano en la pared transversal, una gran mesa en herradura en el centro y un perchero en la mitad de la pared longitudinal. Esta disposición se completa con un reloj eléctrico al igual que en todas las oficinas de la fábrica. En los festivos días de utilización de la sala, el púlpito del orador estará cubierto con la bandera de la cruz gamada. Sobre una tarima se colocará un enorme retrato de Hitler con banderas similares que lo flanquean. A la altura del revestimiento de madera de la pared, y en la misma ratio estética, colocan una guirnalda adornada de banderitas con la cruz gamada [28]. Podríamos pensarlo como una *disposición escénica extra-contextual* en la cual, por un lapso de tiempo determinado, los obreros “arios” son poseídos por los guerreros de *La Germania* de Tácito transfigurando su diaria tarea: el trabajo es también combate así como el verdadero trabajo de los germanos que emergen del multiseccular texto etnográfico es la guerra. Este significante (*séquito*) con su disposición escénica ad hoc nombrada, a su vez, con otro significante similar (*Sala del Séquito*) pueden crear un tipo de *legitimidad de mando*

parecido al tipo ideal carismático de Max Weber aunque más primario y circunscripto a un ámbito “micro-político”.

Ahora bien; cuando el lenguaje de la *Germania* de Tácito adviene a los restos de la República de Weimar nos situamos en lo que podemos llamar *patología quijotesca*: actores obreros y patronales de la primera mitad del siglo XX se invisten de la dignidad simbólica y atributos heroicos y se visten con la *armadura lingüística* necesaria para transfigurar las relaciones sociales producto de una sociedad diferenciada ocupacionalmente y referenciada en representaciones mentales del trabajo de tipo moderno-industrial. Lo que ven si levantan la visera de la armadura será algo comparable a lo que ve y se representa Don Quijote.

## 5. CONCLUSIÓN

El cliché de Klemperer es un cliché con una disposición escénica no vivenciada ni ex - comunicada por el inconsciente individual. Es una escena del pasado cuasi-mítica/fundacional imaginada por el grupo-masa o por los Aparatos de Estado del nazismo. A diferencia del anterior, el cliché de Lorenzer corresponde a una disposición escénica vivenciada, olvidada y ex – comunicada de carácter individual. La escena tipo desde la cual se enuncian los términos y sintagmas arcaicos klemperianos de carácter incontaminado por la historia. Es el símil de la escena del Origen descrita por Martín Heidegger en su *Discurso del Rectorado* de 1933. Hay una instancia de comienzo traicionada por un tiempo histórico inauténtico: dado que el orden causal del pasado es una suerte de mal paso histórico, el trabajador de una fábrica, el docente colega de Heidegger o cualquier miembro de la sociedad alemana de la década del 30 y 40 del siglo XX puede retornar al comienzo histórico mencionado por medio de una forma de lenguaje peculiar. El lenguaje cliché, en el sentido de una frase o palabra repetida irreflexivamente, no proviene de una escena olvidada como en el caso de Lorenzer si no, al contrario; la enunciación de la frase crea la escena que debieron vivenciar sujetos proto-históricos situados en el origen del tiempo. En este sentido el cliché klemperiano es performativo: el cliché es el que ancla la escena para sí. No es una escena alucinada sino una escena “realizativa” o, utilizando el término derivado del inglés, performativa. Los términos utilizados para referirse a las escenas representadas son de carácter performativo. En *El sacramento del lenguaje. Arqueología del juramento* Giorgio Agamben parte de la siguiente proposición teórica que unifica su escrito: frente a la tesis tradicional que conceptualiza a la historia humana arcaica como realizada en dos estadios sucesivos, el primero religioso y el

siguiente jurídico, el filósofo del derecho analiza al juramento como una institución simultáneamente jurídica y religiosa[29]. Es decir, en el juramento nos encontramos con la dimensión performativa del lenguaje y con el indicio de una estructura originaria de este último en la cual “el nexo entre las palabras y las cosas no es de tipo semántico-denotativo sino performativo, en el sentido de que, como en el juramento, el acto verbal realiza el ser”[30]. Esta estructura originaria es previa o contemporánea a otra estructura de tipo asertórico en la cual el valor de verdad de un enunciado proferido por un sujeto es independiente del mismo y se mide desde parámetros de carácter lógico y objetivo como el principio del tercero excluido o el de la adecuación entre las palabras y la realidad que estas intentan designar [31]. Agamben también plantea una hipótesis tenebrosa en relación a la estructura asertórico-denotativa del lenguaje: esta quizá no forme parte de un carácter original y eterno de la lengua humana, sino, más bien, un producto de la historia que puede llegar a desaparecer algún día [32]. El auge de las fakenews que hacen ganar elecciones o plebiscitos sobre alguna cuestión fundamental de los Estados-Nación muestra el posicionamiento cada vez mayor de la estructura performativa o veridiccional del lenguaje tal cual como la nombra Agamben siguiendo a Foucault.

## Referencias

- [1] Video recuperado de <https://www.youtube.com/watch?v=lorFrndURCI>
- [2] Lorenzer, A. (1987). *El lenguaje destruido y la reconstrucción psicoanalítica*. Amorrortu editores. 1ra. edición en castellano, Buenos Aires.
- [3] Agamben, G. (2010). *El sacramento del lenguaje. Arqueología del juramento*. Adriana Hidalgo editora. 1ra. edición, Buenos Aires.
- [4] Lorenzer, A. (1987). *El lenguaje destruido...* p. 99.
- [5] Lorenzer, A. (1987). *El lenguaje destruido...* p. 100.
- [6] Lorenzer, A. (1987). *El lenguaje destruido...* p. 101.
- [7] Campos de Araujo, M.G. (1988). Psicanálise: ciencia natural ou Hermenéutica?- A concepcao de Alfred Lorenzer. *Revista de Psicologia*, Fortaleza. 6 (2): 103-109, Jul/Dez.
- [8] Campos de Araujo, M.G. (1988). Psicanálise...
- [9] Campos de Araujo, M.G. (1988). Psicanálise...
- [10] Lorenzer, A. (1987). *El lenguaje destruido...* p. 113.
- [11] Lorenzer, A. (1987). *El lenguaje destruido...* p. 113.
- [12] Freud, Sigmund. (1992). *Obras Completas*. Tomo X: Análisis de la fobia de un niño de cinco años a propósito de un caso de neurosis obsesiva. 3ra. reimpresión. Amorrortu. Buenos Aires. Citado por Lorenzer. Documento recuperado de <http://bibliopsi.org/docs/freud/10%20-%20Tomo%20X.pdf>.
- [13] Rodríguez Alzueta, A. (2019). *Vecinocracia. Olfato Social y Linchamientos*. Editorial Estructura Mental a las Estrellas, La Plata.
- [14] Klemperer, Víctor (2007). *LTI Apuntes de un filólogo*. Editorial Minúscula. Barcelona. pp. 26-27.
- [15] Berman, Harold J. (1996). *La formación de la tradición jurídica de occidente*. FCE, México. pp. 312-318.
- [16] Berman, Harold J. (1996). *La formación de la tradición...* p. 312.

- [17] Bermúdez, Marcial Tenreiro. (2007). Tácito Germania 13.1. Armas, Jóvenes y Ritos de Paso. Valores Simbólicos de la Guerra en un Ritual Germano. *Actas del VI Encuentro de Jóvenes Investigadores*. Editado por Echeverría, F.; Montes, M<sup>a</sup>.Y. y Rodríguez, A. pp. 121-134.
- [18] Tácito, Cayo Cornelio. (2012). *La Germania*. Universidad de Chile. Facultad de Derecho. p. 10.
- [19] Tácito, Cayo Cornelio. (2012). *La Germania*. p. 10.
- [20] Tácito, Cayo Cornelio. (2012). *La Germania*. p. 10.
- [21] Tácito, Cayo Cornelio (2010). *La Germania*. Editorial Alderabán. Cuenca. Estudio preliminar, texto latino, traducción y notas de Juan Luis Posadas. p. 47.
- [22] Tácito, Cayo Cornelio (2010). *La Germania*. p. 47. Sección XIV.
- [23] Tácito, Cayo Cornelio (2010). *La Germania*. p. 11. Sección XIV.
- [24] Tácito, Cayo Cornelio (2010). *La Germania*. p. 46. Sección XI.
- [25] Tácito, Cayo Cornelio (2010). *La Germania*. p. 9. Sección XI.
- [26] Tácito, Cayo Cornelio (2010). *La Germania*. p. 46. Sección XI.
- [27] Klemperer, Víctor (2007). *LTI Apuntes de un filólogo*. p. 343.
- [28] Klemperer, Víctor (2007). *LTI Apuntes de un filólogo*. p. 341.
- [29] Giorgio, A. (2010). *El sacramento del lenguaje. Arqueología del juramento*. 1ra. edición. Buenos Aires.
- [30] Agamben, G. (2010). *El sacramento del lenguaje*. p. 86.
- [31] Agamben, G. (2010). *El sacramento del lenguaje*. p. 86.
- [32] Agamben, G. (2010). *El sacramento del lenguaje*. p. 86.

# **Análisis de la argumentación en el alegato final del juicio a las Juntas Militares en Argentina**

**Carolina Paola Tramallino**

UNR, CONICET

Rosario, Argentina

carotramallino@hotmail.com

## **Abstract**

The present work aims to make an approach to the critical analysis of discourse taking as a document an exhibition produced in Buenos Aires, Argentina, in 1985. It is the final plea pronounced by the prosecutor Julio Strassera, in the trial of the Military Juntas that took place during the presidency of Raúl Alfonsín. The starting point is to discover the traces of the enunciation perceived in the statement, and to analyze the discourse not only from a semantic point of view; by establishing the 'system of repression' as a reference, but also from an argumentative point of view; by pointing out the position of the speaker in the face of the actions of the accused, for whom he demands a sentence. It will be shown how the speaker establishes an opposition between the previous regime, sustained by the military government since 1976 and the current democratic regime, as a result of the assumption of the first constitutional president in 1983

**Keywords:** Critical discourse analysis, Argumentation, Reference.

## **Resumen**

El presente trabajo se propone realizar una aproximación al análisis crítico del discurso tomando como documento una exposición producida en Buenos Aires, Argentina en 1985. Se trata del alegato final [1] pronunciado por el fiscal Julio Strassera, en el juicio a las Juntas Militares, ocurrido durante la presidencia de Raúl Alfonsín. Tomará como punto de partida poder descubrir las huellas de la enunciación percibidas en el enunciado, y analizar el discurso tanto desde el plano semántico estableciendo el “sistema de represión” como referencia como desde el plano argumental, señalando la posición del enunciador frente al accionar de los acusados, para quienes exige una condena. Se mostrará cómo el locutor establece una

oposición entre el régimen anterior, sostenido por el gobierno militar desde 1976 y el régimen democrático actual, como resultado de la asunción del primer presidente constitucional en 1983.

**Palabras claves:** Análisis crítico del discurso , Referencia, Argumentación.

## 1. INTRODUCCIÓN

En presente trabajo se propone realizar una aproximación al análisis del discurso, tomando como documento, una exposición producida en Buenos Aires en 1985. Se trata del alegato final<sup>1</sup> pronunciado por el fiscal Julio Strassera<sup>2</sup> en el juicio a las Juntas Militares ocurrido durante la presidencia de Raúl Alfonsín<sup>3</sup>.

El punto de partida será descubrir las huellas de la enunciación percibidas en el enunciado [2] y analizar el discurso desde el plano semántico, en el que se erige el “sistema de represión” como referencia [3]. Con respecto al plano argumental, se señalará la posición del enunciador frente al accionar de los acusados, para quienes exige una condena.

En este recorrido se mostrará cómo el locutor establece una oposición entre el régimen anterior, sostenido por el gobierno militar desde 1976, época a la que le otorga lo sombrío, la oscuridad, la opresión y el silencio; y el régimen democrático actual, como resultado de la asunción del primer presidente constitucional en 1983, que se perfila como el momento para hablar de aquello, para contar lo que sucedió, para descubrir y sacar a la luz lo que trató de ocultarse.

De esta manera representará las dos argentinas a través de metáforas y comparará el pasado del país con un infierno. Por lo tanto, empleará un registro religioso, no solo para retratar los

---

1 La extensión del discurso llevó a elegir un fragmento que está publicado en *Memoria histórica para adolescentes. Dictadura y derechos humanos en Argentina* por Fabián Pico.

2 Julio César Strassera, abogado de profesión, trabajó en el Poder Judicial desde 1965. Durante la última dictadura se desempeñó como juez de sentencia en el fuero ordinario. En 1983, con la reapertura constitucional, fue nombrado fiscal de la cámara en lo Criminal y Correccional de la Capital Federal y estuvo a cargo de la fiscalía en el juicio a las Juntas militares en 1985. Al finalizar el juicio, fue designado por el gobierno de Raúl Alfonsín como embajador para Derechos Humanos en Ginebra. El gobierno de Carlos Menem lo confirmó en ese cargo pero Strassera renunció tras los Indultos.

3 Raúl Alfonsín: político radical. Líder de las corrientes “renovadoras” del radicalismo durante la década de 1970. Asumió en 1983 e incorporó a su campaña electoral las demandas de los organismos de derechos humanos, fue el que impulsó el Juicio a las Juntas militares en 1985. Ante la presión militar y las rebeliones de los “carapintadas”, su gobierno sancionó las leyes de Punto Final y Obediencia Debida (1986-1987) que otorgaron impunidad a los represores. Debó entregar el poder a Carlos Saúl Menem antes de concluir su mandato.



aspectos condenables de la represión, sino también para situarse como un enviado de Dios que tiene la misión de impartir justicia en la tierra.

El juicio a las Juntas Militares es efecto del contexto histórico, social y político que atraviesa las últimas décadas de la Argentina, siguiendo a Arnoux & Bonnin [4] desde fines de la década del 70, respecto de los discursos elegidos, "en los objetos de estudio que se fueron privilegiando incidieron los propios contextos políticos de violencia, dictaduras, transiciones a la democracia, sucesivas crisis del sistema neoliberal (...)" . (Arnoux & Bonnin, 2020: p. 6)

En primer lugar, se mencionarán los hechos que dieron lugar al juicio en cuestión, cuya importancia lo ha convertido en un hecho histórico que marcó un hito, no sólo en Argentina sino en Latinoamérica y el resto del mundo. A continuación, se examinará el alegato, objeto de análisis, para descubrir cuáles son los principales argumentos dados por el fiscal, con los cuales culpabiliza a los acusados.

## **2. EL REGISTRO RELIGIOSO**

Desde la primera frase exhibe su posicionamiento como argentino, pero sobre todo, pone en primer lugar la obligación que tiene como fiscal de denunciar lo que no corresponde a la ley: "La comunidad argentina en particular, pero también la conciencia jurídica universal me han encomendado la augusta misión de presentarme ante ustedes para reclamar justicia". Si atendemos a los distintos significados del sintagma "la augusta misión", podemos deducir que le otorga un carácter místico al papel que le ha tocado como representante y como censor de la causa. Debe responder al pueblo y a Dios ya que "misión" no solo refiere a un encargo sino que también es la peregrinación que hacen los religiosos predicando el Evangelio, mientras que "augusta", además de significar una comisión otorgada a un diplomático, hace alusión al que merece o infunde veneración.

Por lo tanto, podemos arriesgar que el enunciador inscribe su discurso en un registro religioso además de jurídico. Refuerza esta observación cuando declara en nombre del pueblo que el sadismo "no es una ideología política ni una estrategia bélica", sino una perversión moral. De esta manera, sostiene que con el juicio Argentina recuperará su autoestima y su fe en los valores, que han sido dañados y aclara que solicita "el castigo", lo cual se opone a la venganza incontrolada.

Continuando con el registro religioso relata lo que vivían los familiares de las víctimas, a partir del momento en que éstas eran secuestradas, comparándolo con un calvario. Explica que comenzaban por ir a la comisaría en donde no eran aceptadas las denuncias y cierra con una analogía: “Esta era sólo la primera estación de un *calvario*...”. Emplea el término *calvario* como figura del sufrimiento moral, recordando el vía crucis realizado por Jesús antes de llegar al lugar en donde sería crucificado.

Podemos inferir que dota de significados alegóricos a los significantes que utiliza en su exposición, buscando provocar una interpretación religiosa y conmover al auditorio, sobre todo a los jueces, que son los que determinarán la sentencia.

### **3. REFERENCIA: SISTEMA DE REPRESIÓN**

El accionar de los acusados es tan novedoso por lo atroz que no puede definirse ni compararse con regímenes políticos anteriores. Emplea el siguiente el paradigma designacional para referirse al accionar:

“... esta **forma** de delincuencia...”

“... el mayor **genocidio**...”

“... una verdadera **subversión** jurídica...”

“... el **terrorismo** de Estado...”

“El **carácter** arbitrario e indiscriminado de la represión...”

“... un **sistema** de represión...”

“... esa feroz **represión**...”

“... este cobarde **sistema** de represión...”

“... este **método** perverso...”

“... el **proceso** represivo más violento...”

“... **esto**...”

En cuanto a las anáforas gramaticales fieles, tenemos las encabezadas por el artículo *un* o el demostrativo *este*. Respecto de las anáforas léxicas, se encuentra la repetición del mismo término, conformando una repetición parcial: “sistema de represión” y el uso de sinónimos:

“método” y “proceso”, que constituyen anáforas infieles, y en donde hay sustitución por calificación valorativa (“perverso”, “más violento”)

Además emplea hiperónimos: “terrorismo de Estado”, “genocidio” y el retome por metonimia: “subversión jurídica”. [5]. Utiliza también una pronominalización a través del pronombre neutro “esto”.

En todas las designaciones se destaca el sistema de represión, núcleo de los sintagmas nominales, al que le atribuye valoraciones negativas como *cobarde*, *violento*, *feroz*, y al que distingue de otros mediante los demostrativos *esa*, *esta*, *este* y los determinantes *el* o *un*. Por lo que el “terrorismo de Estado” se transforma en el “mayor genocidio” de la historia argentina.

### 3.1. Elementos cohesivos

Los conectores metatextuales que sirven a la cohesión en su enunciación son empleados para explicar el *modus operandi* de los acusados: en primer lugar el secuestro y luego la negación de la detención para impedir la investigación, a través de los ordenadores: “**Primero** el secuestro (...); **segundo**, la mentira,...”

Con ellos también desmantela las contradicciones propias del que esgrime una mentira, mostrando el diferente uso que dio el gobierno de facto al argumento de “la guerra sucia”, en distintos momentos del proceso.

Podemos ver un paralelismo establecido mediante el conector **no solo, sino** referido al accionar de los acusados (1), a la situación de la víctima (2) y a la posibilidad de fundar la paz que tiene el pueblo argentino (3):

(1)

**No sólo** negaban los hechos

**No sólo** ordenaron realizar acciones indignas de las fuerzas armadas

**sino que** convertían al que reclamaba en un subversivo.

**sino que** cuando debieron afrontar la responsabilidad por el mando, negaron...

(2)

Situación que **no sólo** no decide

**No sólo** los secuestrados fueron las víctimas

**sino que** incluso no puede llegar a comprender.

hubo mucho más.

(3)

**No** en el olvido

**No** en la violencia

**sino** en la memoria

**sino** en la justicia.

Además, utiliza los siguientes conectores:

“**Sin embargo**, existían grupos organizados...” se vale de un conector contraargumentativo en el comienzo de una oración asertiva para exhibir que el gobierno por más empeño que ponga en rechazar las acusaciones, no puede ocultar la evidencia que prueba su accionar represivo.

“... sitúa el centro de la suerte de la víctima fuera de ésta, **pero** continúa considerándola responsable...” Utiliza el conector de oposición para demostrar lo injusta situación en que se halla el secuestrado, lo arbitrario de dicha privación de la libertad.

“**Por eso** hoy se hace necesario averiguar la verdad...” con este conector consecutivo señala la necesidad de juzgar a todos los que han violado la ley.

“Quiero utilizar una frase que no me pertenece, **porque** pertenece ya a todo el pueblo argentino...” Con el conector causal indica que renuncia a la originalidad para cerrar el alegato debido a que recurre a una frase que resume el pedido del pueblo: “Nunca más”.

### 3.3. Uso de pronombres indefinidos

Se observa un marcado uso de pronombres indefinidos que como afirma Ducrot [6] no son indeterminados, sino que su función consiste en determinar un concepto, delimitando una parte de su extensión. En el discurso analizado hallamos los siguientes:

1. “...juzgar a *todos* los que hayan violado la ley...”

Se refiere a no hacer distinciones ni perdonar por el rango o jerarquía dentro de la Institución de las Fuerzas Armadas.

2. "... *todos* estábamos en libertad condicional..."

"... le colocaron una gran capucha a *toda* la sociedad..."

Aquí designa al conjunto de la sociedad, como víctima de ser sospechada y perseguida, afectando tanto a docentes, como empleados, curas, niños, estudiantes y demás.

3. "... *todo* acto de solidaridad era sospechado de subversivo." Significa que cualquier acción que se emprendía era reprimida.

4. "...: *todos* los detenidos son subversivos..." desenmascara el sofismo utilizado por los acusados con el que se amparaban para castigar y matar.

5. "Y con *toda* esta mentira sobre la campaña antiargentina,..."

"... y que en *todas* las guerras se producen episodios crueles..."

"*Ninguno* de los documentos preliminares del "Proceso" habla de guerra..."

"... en *todos* los combates las bajas sólo hallaron..."

Con los indefinidos refuta el argumento sostenido por los acusados, respecto de la existencia de una guerra sucia, con la que justificarían los procedimientos ilegales utilizados.

6. "...*cierta* clase de hechos..."

"... especialmente el homicidio en *todas* sus formas..."

"Pero no sólo los secuestrados fueron las víctimas, hubo *mucho* más."

"Pero hay *algo* peor aún."

Con los pronombres y el adverbio de cantidad "mucho" avala su pedido de condena para castigar el accionar de los militares.

7. "...porque pertenece ya a *todo* el pueblo argentino:..." Cierra el discurso empleando una frase que corresponde a todos, con este término abarca tanto a las generaciones presentes como a las futuras y pide: "Nunca más".

#### 4. LA ARGUMENTACIÓN

"La capucha procura la desesperación, la angustia y la locura... Encapuchado tomo conciencia de que el contacto con el exterior no existe. Nada te protege. La soledad es total. El sólo hecho de no poder ver va socavando la moral, disminuyendo la resistencia."

(Testimonio de Lisandro R. Cubas).

En cuanto a la argumentación, sus dos argumentos fuertes para condenar a los acusados son: haber operado en la clandestinidad e implantando el miedo en la sociedad y segundo, haber falseado documentos y desconocido procedimientos:

“La combinación de clandestinidad y mentira produjo efectos que trastornaron a la sociedad argentina”. “Primero el *secuestro* y las tremendas consecuencias sobre la víctima (...), segundo, la *mentira*, el gobierno rehúsa reconocer toda detención o arresto y niega la necesidad de proceder a una investigación.”, “*ferocidad y mentira* son las dos *notas* del sistema de represión que los acusados implantaron durante años en la Argentina”.

Otra vez coloca en el lugar de núcleos de los sintagmas nominales que funcionan como sujeto de las oraciones, a sustantivos que no son agentes porque quiere resaltar el accionar de los militares. El término “notas” además de ser lo que sobresale, lo notorio; refiere al verbo *anotar* que representa lo que se escribe, el legado que se deja a las generaciones venideras. Por lo tanto, podemos inferir el propósito comunicativo del enunciador: aunque hayan tratado de ocultar y quemar la evidencia para no dejar rastros, haciendo desaparecer incluso a los cadáveres de secuestrados, pese a eso, no podrán borrar las huellas dejadas en la memoria del pueblo argentino.

Como muestra de la ferocidad con la que actuaron los enjuiciados, indica el secuestro de miles de personas que “... fueron robadas, torturadas y asesinadas...”. Además establece una analogía respecto de la forma de gobernar equiparándola con la un personaje histórico: “Se substituyó (...) la sentencia razonada por el gesto neroniano del pulgar para abajo”, refiriéndose al emperador romano Nerón, que fue acusado de un despotismo delirante y hasta de matricidio. Incluso hace referencia al nazismo: “Como se dijo haciendo referencia al régimen nazi...”, sirviéndose de la autoridad polifónica [7] para afirmar la siguiente proposición: “...una vez que se convence a la sociedad de que una minoría (...) puede equipararse a una sabandija, el paso que hay que dar para llegar al propósito de exterminio no es ya demasiado grande...”

Las siguientes aserciones apoyan los argumentos esgrimidos:

Tabla 1

<b>Ferocidad y clandestinidad</b>	<b>La mentira, el invento, el engaño</b>
Sustituyó la denuncia por la delación.	Documentos falsos
_____ el interrogatorio por la tortura.	Negar la detención
_____ la sentencia razonada por el gesto Neroniano.	Negar la actuación y prácticas realizadas por los acusados
Secuestraron a miles de personas. Torturaron y mataron.	Afirmaron que las denuncias internacionales eran fruto de una campaña antiargentina organizada por subversivos.
Ocuparon las casas y mantuvieron a los parientes de los buscados como rehenes.	Considerar culpable a la víctima de una conducta que no puede comprender.
Feroz represión que constituyó la negación de la justicia.	Convirtieron al que reclamaba o se hacía eco de las denuncias en un subversivo.
Amordazaron a la prensa internacional.	Simular enfrentamientos

Tabla N°1: Aserciones que apoyan los argumentos.

El locutor introduce la trama narrativa cuando relata los pasos que seguían los familiares de las víctimas una vez que eran secuestrados, para señalar que en todos los casos ocurría lo mismo y que de ninguna institución a la que se dirigían, recibían respuesta.

El hecho de que se haya habilitado una oficina para la búsqueda de personas desaparecidas en el Ministerio del Interior es otro dato que prueba que operaron a través de la mentira, por eso emplea el modalizador “con cinismo” para presentar ese suceso.

Otra justificación que da es que los habeas corpus y las medidas judiciales eran frenados mediante “informes *falsos*” de los diferentes comandos y de la policía.

Cierra con una metáfora: “Era un chocar permanente con puertas cerradas”, que refiere al encierro, la opresión, la oscuridad y clandestinidad con la que se condujeron. Es importante mencionar que una de las prácticas era buscar a la víctima por la noche en su propia casa, “entraban por la fuerza, aterrorizaban a padres y niños, (...) se apoderaban de la persona buscada, la golpeaban brutalmente, la encapuchaban y finalmente la arrastraban a los autos o los camiones...” [8]. Los operativos se realizaban a altas horas de la noche o de la madrugada. Generalmente, en el domicilio irrumpía una patota. Previo al arribo de la patota se producía el apagón o corte de la energía eléctrica en la zona del operativo. A esta tarea de privar la visión de las víctimas, se la llamó “tabicamiento” en el lenguaje de los represores,

porque se les colocaba un “tabique” o elemento para que no pudieran ver, ello se efectuaba en el mismo lugar donde se los secuestraba, así, los detenidos arribaban al sitio de detención encapuchados.

Las puertas cerradas simbolizan la falta de libertad por eso el fiscal expresa: “Todos estábamos en libertad condicional.”, aludiendo a la imposibilidad del pueblo de desarrollar cualquier actividad ya que esto podía resultar sospechoso. Aquí el enunciador enumera, entonces, acciones cotidianas que él concibe como “solidarias” y sin embargo eran sospechadas de subversión: “Enseñar a leer... dar catequesis... pedir... atender...”.

El locutor continúa así con la metáfora de la clandestinidad y compara la costumbre de las patotas de colocarles una capucha a las víctimas con la acción psicológica esparcida en la sociedad, la cual también se siente encapuchada y paralizada por el temor. Por lo tanto, asimila el proceso vivido con un “descenso a zonas tenebrosas del alma humana...”.

Lo que destaca como política más grave llevada a cabo por los enjuiciados es el haber negado su compromiso en las órdenes impartidas, no haber afrontado la total responsabilidad que les cabía en esos hechos repudiables, por lo tanto, utiliza el recurso estilístico de la anáfora que enfatiza mediante la repetición y la negación: “... *negaron* sus órdenes, *negaron* conocimiento de lo actuado por sus subordinados; *negaron* conocimiento de los secuestros...”

A esta actitud contraponen los hechos que evidencian la realización de esas órdenes mediante un conector contrastivo: “*Sin embargo*, existían grupos organizados que cumplían un horario especial, cuya tarea era interrogar y torturar, y la realizaban en unidades militares...”

#### **4.1. Refutación del argumento Guerra sucia**

Como refutación de un argumento contrario, el fiscal probará que la supuesta “guerra sucia” que utiliza el adversario como justificación, nunca existió.

Emplea el organizador discursivo *en primer lugar* con el que establece un ordenamiento y centra su interés en demostrar que no hubo tal guerra aducida por el régimen militar. Esta será su hipótesis, a partir de la cual dará las razones necesarias para probarla.

El primer subjetivema del que se sirve para calificar el argumento esgrimido por los oponentes, es *deleznable*. Rechaza que “*esto*” (aquí el pronombre demostrativo señala la imposibilidad de nombrar el terrorismo instaurado) haya sido una guerra, a la que además ellos llaman “no convencional” para justificar los “inhumanos procedimientos utilizados”,



para esconderlos detrás de una generalización absurda: “en *todas* las guerras se producen episodios crueles”.

Como razón primera, para rechazarlo, Strassera postula que en ninguno de los documentos liminares del Proceso se menciona una circunstancia de guerra.

Para demostrar la cambiante actitud del gobierno ante los foros internacionales, introduce un relato y mediante organizadores discursivos establece en secuencias temporales las distintas posiciones, apoyadas con fechas: *Primero* negaron la existencia de los hechos que se denunciaban, *en una segunda etapa* de 1978 a 1981, destacaron que la Argentina era víctima de una campaña internacional dirigida por la subversión desde el exterior y *finalmente* el gobierno habló de una “guerra no declarada”.

Luego, mediante la repetición de preguntas retóricas que comienzan con la construcción: “¿Qué clase de guerra es ésta...?” hace reflexionar al auditorio sobre lo lejos que está el concepto de guerra con respecto a los sucesos acontecidos durante esos años. Las pruebas son claras: no están documentadas las distintas operaciones, no existen partes de batalla ni listas de bajas, los enfrentamientos son simulados y hay bajas solo del lado de los enemigos de las fuerzas legales.

A continuación, sigue con las preguntas retóricas que no esperan respuesta al interrogante y en las que utiliza el paralelismo: “¿Se puede considerar acción de guerra...” o “¿Es una acción de guerra...”, los hechos de secuestro en horas de la madrugada, la tortura y asesinato, el ocupar las casas y “mantener a los parientes de los buscados como rehenes”, el tener como objetivo militar “los niños recién nacidos”. Para finalizar, niega rotundamente a través de la doble negación: “No señores jueces, esos no fueron episodios no queridos pero inevitables”.

En cuanto al uso de marcadores discursivos, utiliza “y como tal” para asimilar la violación de los principios establecidos en el convenio de Ginebra, con un crimen de guerra.

Luego, a través del ordenador “Por otra parte”, descalifica el argumento por el cual la guerra entablada haya sido “sucias” y “no convencional” .Aduce que tales calificativos no poseen entidad jurídica, por lo cual concluye, a través del conector “por lo tanto”, que no son importantes para aplicar el derecho.

Otorga una concesión al argumento del oponente y acepta por vía de hipótesis la teoría de la guerra, pero para develar que esos hechos exceden las necesidades del combate, por lo tanto,

se convierten en crímenes de lesa humanidad y no son permitidos ni aun en la guerra. El conector “ni siquiera” invalida el alcance de la concesión.

## **5. CONCLUSIÓN**

El terrorismo de Estado incidió de forma negativa en la transmisión de valores y experiencias que se produce de una generación a la otra. Durante esos años se instalaron hábitos vinculados al miedo y al egoísmo, ejemplificados en las frases: “No te metás” y “Por algo será”, que se refieren a alejarse de la política y de lo público en el primer caso, y a la justificación de la desaparición de personas en el segundo. Luego, con la recuperación de la democracia, la frase era “Yo no sabía nada”, que conlleva una actitud de no responsabilidad colectiva frente a lo sucedido. En oposición a esta actitud, el enunciador aclara que la función de la Comisión no es juzgar sino investigar lo ocurrido durante esos años y para eso confía en los jueces. Esto se evidencia al analizar la puesta en funcionamiento de su enunciación, a través de las numerosas marcas o huellas de subjetividad que el locutor imprime en la exposición.

El yo nos plantea entonces, la necesidad de hablar, de contar, de no callar ni ocultar, de liberarse de esas prohibiciones tan arraigadas en la sociedad. Se posiciona como un mensajero de la sociedad encargado de lograr que se haga justicia. Habla en nombre del pueblo argentino y reviste su misión de un carácter religioso. De esta manera, focaliza el accionar de las Fuerzas Armadas y lo ubica en primer lugar en la mayoría de las proposiciones, para mostrar las ilegalidades y abusos cometidos. A su vez, emplea frases impersonales para no nombrar al oponente, como una forma de desprecio hacia este.

Toma como referencia el terrorismo de estado, del que se analizó el paradigma designacional, construido con una fuerte valoración negativa.

Ese período padecido por todos los ciudadanos, que es único en la historia del país por lo cruel, es también lo que no se puede expresar con palabras, por eso recurre a numerosos pronombres indefinidos y lo compara con las disposiciones establecidas por dictadores históricos como Nerón y Hitler.

En síntesis, el análisis del alegato permitió develar la manera en que Strassera descubre y derriba el discurso proclamado por los enjuiciados a través de una argumentación sólida y acentuada por distintas estrategias como son el uso de preguntas retóricas, la refutación y concesión de argumentos contrarios para poder invalidarlos. Se mostró cómo mediante los

marcadores discursivos logra reforzar los datos y razones expuestas y encamina las conclusiones extraídas en su función de fiscal. Estas consideraciones son presentadas al auditorio y con ellas consigue condenar a los culpables de la etapa más oscura acaecida en la historia política de Argentina.

## Referencias

- [1] Pico, F. (2006). *Memoria histórica. Dictadura y derechos humanos en Argentina*. Homo Sapiens, Rosario.
- [2] Benveniste, E. (1975). *Problemas de lingüística general*, Siglo veintiuno, México D.F.
- [3] Arnoux, E. Narvaja de (2006). *Análisis del discurso*. Santiago Arcos, Buenos Aires.
- [4] Arnoux, E. y Bonnin, E. (2020). *Política y Discurso*. Disponible en: <http://www.unsam.edu.ar/escuelas/humanidades/centros/celes/Papeles-de-trabajo-del-CELES-vol-2-2020.pdf> Fecha de consulta: junio de 2020.
- [5] Calsamiglia Blancafort, H. y Tusón Valls, A. (1999). *Las cosas del decir- Manual de análisis del discurso-* Ariel, Barcelona.
- [6] CONADEP (1984). *Nunca Más*, Buenos Aires.
- [7] Ducrot, O. (1994). *El decir y lo dicho*. Edicial, Buenos Aires.
- [8] Ducrot (1994). *El decir y lo dicho*.
- [8] CONADEP (1984) , *Nunca Más*.