http://dx.doi.org/10.22168/2237-6321-10esp2116

Localización y contabilización de sufijos nominales en corpus de aprendientes de español como segunda lengua

TRAMALLINO, Carolina Paola; BELTRÁN, Celina; RICCIARDI, Natalia. Localización y contabilización de sufijos nominales en corpus de aprendientes de español como segunda lengua. **Entrepalavras**, Fortaleza, v. 11, n. esp., p. 412-436, ago. 2021. DOI: 10.22168/2237-6321-10esp2116.

Identification and measurement of noun suffixes in text corpus produced by learners of Spanish as a second language

Carolina Paola TRAMALLINO (CONICET-UNR)
carolinatramallino@gmail.com
Celina BELTRÁN (UNR)
cbeltran2510@gmail.com
Natalia RICCIARDI (UNR)
natalia.ricciardi@gmail.com

Recebido em: 30 de set. de 2020.
Aceito em: 22 de out. de 2020.

Resumen: Este trabajo se propone analizar y contabilizar las palabras que poseen sufijos nominales en *corpus* producidos por aprendices de español como segunda lengua. Se considera de suma importancia obtener el índice de sufijación para poder extraer conclusiones en cuanto a la cantidad de nominalizaciones como así también al porcentaje de ocurrencias coincidentes con el español y de formas idiosincrásicas. El marco teórico corresponde a la hipótesis de interlengua, dentro de las corrientes de adquisición de segundas lenguas y a las últimas investigaciones sobre tratamiento automático de errores en *corpus* de español, particularmente los léxicos. La muestra es de tipo textual y está dividida en grupos según el nivel de proficiencia de los estudiantes cuyas lenguas nativas son variadas. Con respecto a la metodología, se utilizará el sistema NooJ para localizar los términos que posean sufijos nominales y a continuación, se emplearán dos técnicas estadísticas para la medición de los datos: el Test no paramétrico de Wilcoxon y la Prueba de Kruskal-Wallis. Los resultados obtenidos mostrarán que existen diferencias significativas respecto del uso y la distribución de

sufijos nominales pertenecientes al español entre los *corpus* que no comparten un mismo nivel de instrucción. En efecto, se hallará un aumento significativo de sufijos nominales coincidentes en la muestra de nivel intermedio con respecto al nivel inicial. Sin embargo, los *corpus* analizados no presentarán diferencias considerables en otros dos casos: sufijos existentes en español pero no combinables con las bases y sufijos inexistentes en dicha lengua.

Palabras clave: Sufijos nominales. *Corpus*. Español como segunda lengua. Medición.

Abstract: This study aims to analyze and to record the number of words with noun suffixes in a corpus produced by learners of Spanish as second language. Obtaining the suffix index is of utmost importance to draw conclusions about both, the number of nominalizations and the percentage of occurrences coincident with Spanish suffixes as well as idiosyncratic forms. The study is based on the theoretical framework of the interlanguage hypothesis within the theories of second language acquisition and the latest research on automatic handling of errors in Spanish corpus, particularly lexicons. The text corpus is divided in groups according to the proficiency level of the students, who have different native languages. Regarding the methodology, the Nooj system will be used so as to identify the terms that have noun suffixes, and then two different statistical techniques will be implemented to measure the data: the non-parametric test of Wicoxon for independent samples and the non-parametric test of Kruskal-Wallis. The obtained results will show substantial differences in the use and distribution of Spanish noun suffixes among the corpus that do not share the same level of proficiency. Indeed, a significant increase in matching nominal suffixes will be found in the intermediate level sample as compared to the initial level. However, the examined corpus will not show considerable differences in two other cases: existing suffixes in Spanish not combinable with bases and non-existent suffixes.

Keywords: Noun-suffixes. Corpus. Spanish as a second language. Measurement.

Introducción

Este artículo se propone localizar y contabilizar las palabras que poseen sufijos nominales en *corpus* de español como segunda lengua mediante un *software* libre. El trabajo forma parte de un proyecto sobre enseñanza de español a partir del uso de tecnología, dependiente del Centro de Estudios de Adquisición del Lenguaje de la Universidad Nacional de Rosario, Argentina. Se considera de suma importancia obtener el índice de sufijación para poder extraer conclusiones en cuanto a la cantidad de nominalizaciones como así también al porcentaje de ocurrencias coincidentes con el español y de formas idiosincrásicas. A su vez, dichas conclusiones serán relevantes para formular hipótesis acerca del funcionamiento del "lexicón" (BARALO, 2001) en el aprendizaje de una lengua extranjera. Este concepto se refiere a la organización mental de los primitivos morfológicos y de su combinación para formar palabras complejas o derivadas.

En el primer apartado, referido al marco teórico, se hará mención a algunos de los estudios emprendidos en cuanto a la adquisición

ISSN 2237-6321

v. 11 (esp.)
412-436
ago.
2021

de la morfología derivativa, para realizar, luego, un breve recorrido por las teorías de adquisición de segundas lenguas hasta llegar a la corriente de interlengua, hipótesis a la que adhiere la presente investigación.

Además, se explicitarán las aplicaciones y los alcances de la lingüística computacional con el objeto de revisar, particularmente, las investigaciones actuales sobre análisis y tratamiento automático de errores en *corpus* de aprendices no nativos de español.

En la sección sobre metodología, se detallarán cuestiones referidas al *corpus* empleado, que reúne producciones de estudiantes que se encuentran en un nivel inicial o intermedio de aprendizaje, cuyas lenguas de origen son variadas, a saber: portugués, inglés, francés, alemán y holandés. Asimismo, se describirá y analizará la muestra, particularmente en lo que respecta al empleo de sufijos nominales en la formación de nombres, lo cual permitirá establecer una clasificación propia de acuerdo con tres casos diferentes: SC (sufijos coincidentes), SI (sufijos inexistentes) y SE (sufijos existentes), que se expondrán con ejemplos.

A continuación, se explicará el funcionamiento de la herramienta informática NooJ, específicamente la función para localizar términos y terminaciones de palabras. Además, se mostrará el análisis estadístico efectuado a partir de dos técnicas adecuadas para muestras independientes, como son el Test no paramétrico de Wilcoxon y la Prueba de Kruskal-Wallis y la obtención de los resultados.

Por último, se expondrán las discusiones y posteriores conclusiones de acuerdo al análisis cuantitativo y cualitativo de los datos para contrastarlas con las conclusiones halladas en los estudios teóricos referenciados.

Marco teórico

Existen diversos estudios sobre la morfología derivativa y sus implicancias en la metodología de la enseñanza de una segunda lengua (WILLIAMS, 1994; SALAZAR GARCÍA, 1994; VARELA, 2003). Estos coinciden en la hipótesis referida a la adquisición de la formación de palabras de manera gradual y basándose (al igual que en la L1 o lengua materna) en la segmentación de las formas morfológicamente complejas (VARELA, 2003). Asimismo, siguiendo a Varela (2003), los modelos de formación más productivos de la lengua que se estudia son los que se adquieren en primera instancia. De esta forma, se refiere a

la capacidad del aprendiz[1] de establecer una correlación entre sufijo y categoría (VARELA, 2003).

En efecto, el aprendizaje léxico implica el conocimiento morfológico de las diferentes unidades léxicas que organizan la estructura de las palabras y la formación de nuevos términos. Agaronian (2019) asegura que es a partir de este conocimiento que los estudiantes pueden establecer relaciones entre las nuevas palabras de la lengua que aprenden y las reglas sujetas a restricciones semánticas, sintácticas y fónicas (AGARONIAN, 2019). En el mismo sentido, Baralo (2001, p. 12) asevera que existe una competencia morfoléxica, (denominada "lexicón", un concepto cognitivo, dinámico y procesual) que se encuentra representada en tres niveles:

> [...] en primer lugar, los datos de entrada de las reglas de formación de palabras, que contienen una lista finita de morfemas (-ble, -mente, -or, -ito, por ejemplo); en segundo lugar, las reglas de formación de palabras y, por último, la salida o producto de estas reglas. Estas reglas contienen especificaciones con respecto a la categoría gramatical a la que pueden adjuntarse [...] (BARALO, 2001, p. 17).

Las formas idiosincrásicas en la formación de sustantivos que hallamos en el *corpus* son producto de la aplicación de esas reglas que realiza el hablante no nativo, lo cual puede llevarlo a generar palabras que no coinciden con las que realmente existen en la lengua española.

Respecto a la adquisición de segundas lenguas, en la década del 40 del Siglo XX, en el ámbito de la lingüística aplicada, comienzan las investigaciones que tienen como objeto el estudio de la lengua del aprendiente. A través de un análisis contrastivo entre la lengua nativa del estudiante y la lengua meta, se intentan predecir los errores que este cometerá con la finalidad de evitarlos. De esta forma, se estudian las zonas de contacto y divergencia entre ambos sistemas en los diferentes niveles lingüísticos, como el fonológico, morfológico, sintáctico y léxico para detectar y pronosticar dichos errores.

Sin embargo, al advertir que la fuente de la causa de los errores no respondía únicamente a la interferencia con la lengua materna (entendida esta como el efecto de la lengua nativa sobre la lengua meta) surge un nuevo paradigma que es el del análisis de errores. El primer referente de esta línea investigativa es Corder (1967), quien les atribuye

[1] En este estudio, los términos "aprendiz" y "aprendiente" serán empleados como equivalentes.

ISSN 2237-6321

v. 11 (esp.)
412-436
ago.
2021

a los errores un valor positivo, en lugar de considerarlos como un hecho negativo, debido a que estos aplican como un instrumento mediante el cual los estudiantes pueden deducir las nuevas reglas del sistema lingüístico que están aprendiendo. A su vez, el análisis de errores adopta una perspectiva más empírica, al clasificar los errores efectivamente producidos por los estudiantes y sin ocuparse de predecirlos, tarea que ejercía un lugar central en el análisis contrastivo. El aporte de esta corriente reside en despojar de negatividad al error y tomarlo como un indicador del proceso de aprendizaje llevado a cabo por el aprendiz. En palabras de Torijano (2008), el error es visto como paso de aprendizaje (*learningstep*), el cual tiene la capacidad de mostrar que, en efecto, el alumno está avanzando en su Interlengua hacia la segunda lengua (TORIJANO, 2008).

Acerca de las taxonomías de errores, algunas investigaciones relevantes en este ámbito son las de Liceras (2009), Rigamonti (2006) y Alba Quiñones (2009), quien retoma todas las clasificaciones existentes realizadas hasta ese momento desde distintos enfoques, entre otros estudios.

Hipótesis de la interlengua

La corriente de interlengua implica un giro metodológico ya que se propone el estudio y análisis tanto de las formas idiosincrásicas como de las que coinciden con el sistema lingüístico que se adquiere (ALEXOPOLOU, 2010). En otras palabras, esta línea de investigación toma en consideración la producción total de los estudiantes para demostrar que tanto unas como otras son importantes en el proceso de aprendizaje. El término es empleado, inicialmente, por Selinker en 1972, quien propone un enfoque psicolingüístico y ubica a este sistema en un lugar intermedio entre la lengua materna y la meta, por lo que posee elementos comunes con ambas. De esta forma, sostiene que los estudiantes transitan por etapas o niveles de competencia que van reestructurando ese sistema a medida que adquieren vocabulario y nuevas estructuras del idioma que aprenden. Dicho atributo de variabilidad responde, por lo tanto, a su carácter transitorio. Lo valioso de esta perspectiva radica en la siguiente hipótesis: aunque esta lengua idiosincrásica difiera en cada alumno en particular, presenta zonas de intersección en aprendientes con un nivel de instrucción similar. Corder (1971) había denominado a este sistema un *dialecto idiosincrásico*

o *dialecto transitorio* al que también atribuía como característica principal el poseer una gramática propia que va modificándose. Asimismo, explicaba que podía corresponder tanto a un sistema peculiar de un estudiante individual como a un grupo de estudiantes que compartieran el mismo nivel académico.

Selinker (1972) propone la existencia de una estructura psicológica latente, "la cual es activada por el aprendiz de L2 al momento de expresarse en una determinada situación comunicativa" (TRAMALLINO, 2016, p. 18). En dicha estructura psicológica, se establecen "las identificaciones interlingüísticas, que constituyen el lazo entre los sistemas: la LM, la L2 y IL" (TRAMALLINO, 2016, p. 18).

Para este autor, una cuestión central es el concepto de fosilización, "con éste se refiere a la permanencia en la IL de ciertas reglas gramaticales, vocabulario o aspectos de pronunciación, más allá de la cantidad de instrucción recibida" (TRAMALLINO, 2016, p. 18). Dichos fenómenos persisten en la IL a través del paso del tiempo. La explicación que brinda es que estas estructuras psicolingüísticas se encuentran almacenadas en el cerebro por medio del mencionado mecanismo.

Uno de los principales procesos de fosilización que destaca Alexopoulou (2010, p. 92) tiene que ver con la hipergeneralización del material lingüístico de la L2: es decir, cuando la permanencia es producto de una falsa regularización de reglas de la lengua meta.

La presente investigación adscribe a la teoría de interlengua y por lo tanto, contempla en su conjunto tanto a los sufijos coincidentes con el español como a los idiosincrásicos o propios de la interlengua.

Lingüística computacional

La lingüística computacional es un área interdisciplinaria que toma saberes de la Lingüística, la Informática y la Estadística. Su tarea consiste en crear sistemas informáticos capaces de procesar el lenguaje humano y emular[2] la capacidad lingüística humana. De esta forma, las diversas ramas de la lingüística clásica, tales como la psicolingüística, la neurolingüística, la dialectología, la sociolingüística y la lexicografía, entre otras, se proyectan de forma renovada gracias a los instrumentos digitales que han venido en su

[2] Emular no significa comprender cómo funciona el cerebro humano si no intentar construir sistemas que comprendan y produzcan el lenguaje de manera similar a un humano.

ISSN 2237-6321

v. 11 (esp.)
412-436
ago.
2021

complemento (PARODI, 2004). La mencionada disciplina que cuenta con una trayectoria de varias décadas en los países centrales como Estados Unidos o Francia puede acuñarse bajo el término "tecnologías del lenguaje". Éste refiere a todas aquellas tareas en las que se aplica el conocimiento sobre la lengua para desarrollar sistemas informáticos capaces de reconocer, analizar, interpretar y generar lenguaje (LAVID, 2005). En cuanto a la lingüística de *corpus*, esta se ha convertido en una metodología que se apoya en técnicas estadísticas y computacionales para estudiar datos reales de la lengua, tomando la definición que brinda Parodi (2008).

En lo que respecta a analizadores de errores en *corpus* de español L2, podemos mencionar la investigación llevada a cabo por Campillo Llanos (2014), quien analiza los errores léxicos de hablantes no nativos a través de *ordenador* en estudiantes pertenecientes a nivel intermedio-bajo para extraer frecuencias de uso de categorías y de unidades léxicas. El estudio concluye que los errores referidos a la forma son los más frecuentes en el nivel A1 y se reducen a la mitad en el nivel B, mientras que los de significado léxico persisten de un nivel de aprendizaje a otro. Lo llamativo es que el *corpus* portugués exhibe una mayor cantidad de errores, incluso en la comparación con los *corpora* de otras lenguas romances, como francés e italiano. El autor determina, como posible causa, la proximidad entre la lengua materna y la española.

Por otra parte, Ferreira, Elejalde y Vine (2014) presentan un análisis de Errores Asistido por Computador a partir de *Corpus* de Aprendientes de Lenguas en Formato Electrónico, que se compone de resúmenes realizados por estudiantes de nivel B1. Para ello, diseñan un sistema de etiquetas de anotación de errores y los analizan en diversos niveles de categorización a partir del programa Nvivo 10. Los resultados arrojan que la mayoría de los errores son lingüísticos y entre estos, los más representativos son los gramaticales. Asimismo, los más frecuentes corresponden a los errores de categorías (sustantivos, verbos, artículos, preposiciones, entre otras), que constituyen el 35 %, seguidos por los errores de coherencia textual, que representan el 26 %. En lo que concierne a los errores etiológicos interlinguales, es decir, que tienen su causa en la interferencia con la lengua materna, los de mayor frecuencia son los errores de uso de la L1. En cuanto a los intralinguales (que refieren a cuestiones particulares de la lengua meta), los de mayor aparición son los de sobre-generalización de las reglas pertenecientes al sistema que se está adquiriendo.

Respecto de los sistemas que se encuadran en la tipología "tutorial inteligente" (STI), además de dominar un área de conocimiento específico, son metodológicamente adaptativos y dinámicos en la interacción con el estudiante. Estos tienen la capacidad de identificar los errores de entrada y proporcionar una retroalimentación inmediata junto con la fuente de errores. Se los llama inteligentes por la capacidad de analizar gramaticalmente una entrada en lenguaje natural y luego producir un mensaje o enunciado correspondiente a una estrategia de *feedback* correctivo adecuado para la falla localizada del que aprende.

Entre algunas propuestas correspondientes a los últimos años de STI, podemos mencionar "ELE Tutor" (FERREIRA *et al.*, 2012), elaborado para la detección y corrección de errores en textos de aprendientes de español. Perteneciente a este mismo proyecto, se encuadra el trabajo de Kotz Grabole y Ferreira Cabrera (2013), quienes proponen la construcción de un analizador sintáctico computacional para la detección de errores gramaticales. Dichas autoras establecen una jerarquía de equivocaciones teniendo en cuenta dos criterios: la gravedad y la frecuencia. Las más graves refieren a no dar una respuesta acorde al ejercicio propuesto, mientras que las más frecuentes son los errores de concordancia y empleo de verbos (KOTZ GRABOLE; FERREIRA CABRERA, 2013, p. 230).

Metodología

En este apartado se hará referencia a la recolección del *corpus* y a su organización de acuerdo a niveles de instrucción de los sujetos participantes. A continuación, se referirá al análisis cualitativo requerido para poder clasificar los sufijos nominales hallados en tres casos. En la segunda sección, se explicará el funcionamiento de la herramienta empleada para localizar y contabilizar los sufijos coincidentes con el español y los que no corresponden a la lengua meta. Finalmente, se explicitarán las técnicas estadísticas utilizadas para medir los resultados y extraer conclusiones.

CORPUS

La función principal de un *corpus* es la de establecer la relación entre la teoría y los datos. Por lo tanto, la efectividad en la preparación de este reside en el hecho de poder mostrar un fragmento de la realidad lingüística que pretende estudiarse. Para eso, debe cumplir con dos características: primero estar organizado y segundo ser neutro, esto

ISSN 2237-6321

v. 11 (esp.)
412-436
ago.
2021

quiere decir que no posea marcaciones. De esta forma, conseguirá recoger muestras proporcionales de todos los aspectos o niveles que quieran analizarse (TORRUELLA-LLISTERRI, 1999). En este caso, se establece una muestra aleatoria de estudiantes extranjeros de español como L2 que pertenecen a instituciones educativas de la Universidad Nacional de Rosario, Argentina.

El *corpus* empleado es de tipo textual, reúne producciones escritas que han sido digitalizadas y está dividido en dos niveles según el grado de aprendizaje de los sujetos: inicial e intermedio, de acuerdo al Marco Común Europeo de Referencia para las lenguas (MCERL). A su vez, cada muestra está dividida en grupos según la lengua de origen de los aprendientes de español. Se considera que el diseño muestral más apropiado es el estratificado, considerando como estratos a los grupos definidos por la lengua nativa. Las lenguas de origen son variadas: por un lado, románicas (portugués y francés) y por el otro, germánicas (alemán, holandés e inglés).

Estos sujetos son estudiantes jóvenes que se encuentran viviendo en Argentina. Algunos, con el propósito de ingresar a carreras universitarias de Postgrado (*corpus* portugués inicial); otros, para efectuar un intercambio en localidades de Argentina (grupo alemán).

Las consignas de escritura son sugeridas por el docente en las clases de español. Los escritos corresponden al género literario y predominan las secuencias narrativas y descriptivas, puesto que describen la ciudad de Rosario y la ciudad natal, relatan las vivencias de un viaje o bien, detallan la escena que tiene lugar en un cuadro de un pintor argentino. Las indicaciones son del tipo: "*Compare la ciudad de Rosario con su ciudad natal*", "*Cuente cómo fue su último viaje*", "*Describa el siguiente cuadro del pintor Molina Campos*", entre otras.

Los grupos de producciones que tienen como lengua de origen al francés, inglés y al portugués (nivel intermedio) pertenecen a estudiantes de entre 18 y 20 años de edad que están interesados en ingresar a la Universidad Nacional de Rosario (UNR). El requisito es aprobar un examen de certificación de dominio de español y son, precisamente, esas producciones de las evaluaciones de ingreso las que se reúnen como *corpus*. Estos escritos están redactados en primera persona y en ellos se encuentran secuencias argumentativas y narrativas ya que las consignas incluyen: *redactar un mail para reservar habitación de un hotel, escribir en qué gustos coincide y en cuáles no con el personaje, redactar una carta solicitando la solución a un problema*, entre otros.

Para la selección de sujetos, se eligieron a aprendientes que se encuentran en una etapa inicial y en una intermedia, con el propósito de investigar qué sucede en cuanto a la adquisición de la morfología del español, teniendo como hipótesis de partida la manifestación de formas idiosincrásicas propias de la interlengua, particularmente de sufijos nominales que no se corresponden con los de la lengua meta.

Según el MCERL[3], el primer nivel corresponde a un usuario básico. A su vez, este está constituido por dos subniveles: el A1 y el A2. Los estudiantes pertenecientes al *corpus* francés e inglés se encuentran en el subnivel A2, mientras que los aprendices cuya lengua de origen es el portugués recién están alcanzando el A1.

Los otros grupos se sitúan en un nivel intermedio, los estudiantes del grupo alemán se ubican dentro del nivel B1, mientras que los de holandés y portugués corresponden al B2.[4]

Cada grupo está compuesto por una cantidad determinada de textos que han sido numerados y corresponden a producciones de diferentes sujetos. El objetivo es facilitar el empleo de tablas en donde conste tipo de sufijo, grupo según lengua nativa y número de texto dentro de ese *corpus*. La muestra de nivel inicial suma 3900 palabras, y la de nivel intermedio da un total de 4500 palabras.

[3] Disponible en http://cvc.cervantes.es/ensenanza/biblioteca_ele/marco/cvc_mer.pdf.

[4] Para aclarar esta diferencia, veamos cuáles son los conocimientos mínimos que deben poseer los estudiantes encuadrados en este estrato:

A1 interacción escrita: Puede escribir postales cortas y sencillas. Sabe rellenar formularios con datos personales. Dispone de un repertorio básico de palabras y frases sencillas relativas a situaciones concretas.

A2 interacción escrita: Puede escribir notas y mensajes breves y sencillos relativos a sus necesidades inmediatas. Escribe cartas personales muy sencillas. Utiliza algunas estructuras sencillas correctamente, pero todavía comete, sistemáticamente, errores básicos.

B1: Es capaz de producir redacciones sencillas, coherentes y cohesionadas sobre temas que le son familiares o en los que tiene un interés personal. Es capaz de narrar una historia, escribir una descripción de un hecho determinado como un viaje reciente, real o imaginado.

B2 Puede producir textos claros y detallados sobre temas diversos, así como defender un punto de vista sobre temas generales. Puede describir informes, proponer motivos que apoyen o refuten un punto de vista. Sabe escribir una reseña de una película, un libro o una obra de teatro.

ISSN 2237-6321

v. 11 (esp.)
412-436
ago.
2021

Descripción del *corpus*

La metodología de trabajo consiste en analizar cada grupo de textos para extraer conclusiones en cuanto al porcentaje de a) sufijos nominales coincidentes con el español estándar, b) sufijos inexistentes, c) sufijos nominales propios del español pero no combinables con las bases, con la finalidad de determinar si existen o no diferencias significativas entre los grupos. Se etiquetan de la siguiente forma:

a) SC (sufijos coincidentes), b) SI (sufijos inexistentes), c) SE (sufijos existentes).

En el siguiente cuadro, se muestran ejemplos del caso b, es decir, sufijos inexistentes (SI) en el español que fueron hallados en las muestras del *corpus*:

Cuadro 1 – sufijos inexistentes en español

LENGUA MATERNA	SUFIJO	EJEMPLO	NIVEL
PORTUGUÉS	-I	FILOSOFI	A1
PORTUGUÉS	-DADE	NOVIDADE	A1
PORTUGUÉS	-DADE	UNIVERSIDADE	B2
PORTUGUÉS	-DADE	MOVILIDADE	B2
PORTUGUÉS	-IÑO	VECIÑO	B2
ALEMÁN	-EN	NADIEN	B2
INGLÉS	-TION	INFORMATION	A2
HOLANDÉS	-TION	ESTATION	B2
HOLANDÉS	-EN	PARTICIPANTEN	B2

Fuente: elaboración propia.

El cuadro que sigue exhibe ejemplos del caso b) SE (sufijos existentes), pero no combinables con las bases:

Cuadro 2 – sufijos existentes en español, pero no combinables con las bases.

LENGUA MATERNA	SUFIJO	EJEMPLO	NIVEL
PORTUGUÉS	-e	Maestre (maestro)	A1
PORTUGUÉS	-eza	Defeza (defensa)	A1
ALEMÁN	-e	panique (pánico)	B1

ALEMÁN	-encia	Presistencia (pesidencia)	B1
ALEMÁN	-al	Tipical (típica)	B1
HOLANDÉS	-tor	Asesinator	B2
HOLANDÉS	-tad	Identitad	B2
HOLANDÉS	-orio	Armatorio	B2
HOLANDÉS	-es	Figures	B2
PORTUGUÉS	-ancia	Segurancia	A2
PORTUGUÉS	-tad	Universitad	A2
PORTUGUÉS	-e	Nasale	B2
PORTUGUÉS	-dad	Dificuldad	B2
INGLÉS	-e	Persone	A1
FRANCÉS	-a	Responda	A1
FRANCÉS	-os	Teuistos	A1

Fuente: elaboración propia.

El caso de sufijos que no pueden combinarse con ciertas bases, por ejemplo: *segurancia, armatorio, asesinator*, *tipical* (Ver Cuadro 2) tienen su causa en la aplicación incompleta de las reglas gramaticales del español que determinan la falsa elección de una forma gramatical, de acuerdo a la taxonomía de errores presentada por Ferreira Cabrera; Elejalde Gómez (2020). En tanto, la ocurrencia de sufijos que no pertenecen al español, como por ejemplo: *novidades, infomation, nadien.* (Ver Cuadro 1) responden a la transferencia directa del léxico.

Para localizar palabras formadas con sufijos nominales del español, se utilizará el sistema NooJ. En cuanto a la localización de sufijos inexistentes, se deberá prestar atención a las palabras resultantes del primer análisis morfosintáctico que arroje la herramienta, analizadas como "desconocidas", es decir, no consignadas en los diccionarios. A continuación, será necesario realizar un análisis cualitativo que permita determinar cuáles corresponden al empleo de un sufijo inexistente en la lengua española.

Sistema NooJ

Este programa de libre acceso y disponible on line es una herramienta para el tratamiento de lenguas naturales desarrollada por Max Silberztein[5] (2003) en 2002. Contiene módulos para más de veinte lenguas, entre ellas el italiano, inglés, portugués, griego, francés.

[5] Software NooJ disponible en: http://nooj-association.org

ISSN 2237-6321

v. 11 (esp.)
412-436
ago.
2021

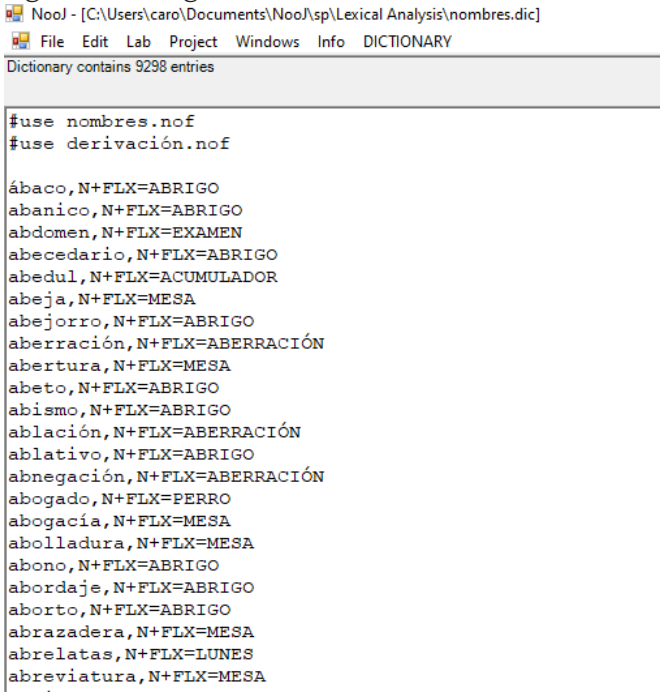
No solo localiza términos en grandes *corpus* de textos, sino que además, indica el capítulo en el que se encuentra y exhibe las palabras del contexto anterior y posterior al término encontrado. Además, puede buscar una categoría con un rasgo morfológico determinado o rastrear todas las variantes morfológicas de un lema. También permite mostrar todas las ocurrencias o sólo cien de ellas, mediante una indicación del usuario, lo cual es fundamental para trabajar con datos estadísticos.

Diccionarios

El *software* cuenta con dos tipos de recursos, los diccionarios y las gramáticas, que pueden ser morfológicas o bien, sintácticas. En el módulo *Spanish* (Argentina) correspondiente al español[6], los diccionarios de las palabras variables, como nombres, adjetivos y verbos, se encuentran asociados a gramáticas morfológicas en donde se declaran los modelos flexivos (TRAMALLINO, 2013).

En el diccionario correspondiente a nombres, por ejemplo, se encuentra consignada la siguiente información lingüística: la etiqueta morfosintáctica N (nombre) y el modelo de flexión que sigue ese sustantivo, como se observa en la siguiente captura de pantalla:

Figura 1 – Fragmento diccionario de nombres



Fuente: Diccionario Nombres en Módulo español (Argentina), sistema NooJ.

[6] Spanish Module (Argentina) disponible en: http://www.nooj-association.org/index.php?option=com_k2&view=item&id=6:spanish-module-argentina&Itemid=611

En la gramática morfológica dependiente de las entradas del diccionario "Nombres", se confeccionan los modelos flexivos de acuerdo a los comandos que trae el sistema. Entre ellos, se pueden mencionar: <B> borra caracteres de izquierda a derecha. <A> quita la tilde a la primera vocal acentuada que encuentra de izquierda a derecha. <Á> agrega tilde aguda a la primera vocal que encuentra de izquierda a derecha. <C> cambia minúsculas por mayúsculas y viceversa. <D> duplica un carácter. Por ejemplo, para formar la variante femenina singular del nombre *perro*, debe confeccionarse el modelo de la siguiente forma:

PERRO = <B>a/fem+sg, el cual indica que debe borrar la "o" y agregar la "a" con la etiqueta femenino singular.

La imagen que se encuentra a continuación presenta algunos de los modelos flexivos creados para los sustantivos del español que se declaran en el archivo correspondiente:

Figura 2 - modelos flexivos correspondientes al diccionario de nombres

```
ABRIGO = <E>/masc+sg | s/masc+pl;
MESA = <E>/fem+sg | s/fem+pl;
ACCIONISTA = <E>/masc+sg | <E>/fem+sg | s/masc+pl | s/fem+pl;
PERRO = <E>/masc+sg | s/masc+pl | <B> a/fem+sg | <B> as/fem+pl;
ACUMULADOR = <E>/masc+sg | es/masc+pl;
PARED = <E>/fem+sg | es/fem+pl;
BACHILLER = <E>/masc+sg | <E>/fem+sg | es/masc+pl | es/fem+pl;
ACREEDOR = <E>/masc+sg | es/masc+pl |  a/fem+sg |  as/fem+pl;
LUNES = <E>/masc+sg | <E>/masc+pl;
CRISIS = <E>/fem+sg | <E>/fem+pl;
SILLÓN = <E>/masc+sg | <A> es/masc+pl;
ABERRACIÓN =<E>/fem+sg | <A> es/fem+pl;
ANFITRIÓN = <E>/masc+sg | <A> es/masc+pl | <A>  a/fem+sg | <A> as/fem+pl;
BARÓN = <E>/masc+sg | <A> es/masc+pl | <A>  esa/fem+sg | <A> esas/fem+pl;
AVESTRUZ = <E>/masc+sg | <V> es/masc+pl;
CRUZ=<E>/fem+sg | <V> es/fem+pl;
APRENDIZ = <E>/masc+sg  | <E>/fem+sg  | <V> es/masc+pl | <V> es/fem+pl;
CONDE = <E>/masc+sg | s/masc+pl | sa/fem+sg | sas/fem+pl;
ABAD = <E>/masc+sg | es/masc+pl | esa/fem+sg | esas/fem+pl;
MARQUÉS = <E>/masc+sg | <A> es/masc+pl | <A>  a/fem+sg | <A> as/fem+pl;
EXAMEN = <E>/masc+sg | es <L5> <Á>/masc+pl;
RÉGIMEN = <E>/masc+sg | es <L7> <A> <R2> <Á>/masc+pl;
ALTIBAJOS = <E>/masc+pl;
COSQUILLAS = <E>/fem+pl;
AGUAFIESTAS = <E>/masc+sg | <E>/fem+sg | <E>/masc+pl  | <E>/fem+pl;
```

Fuente: Gramática morfológica Nombres en Módulo de Español (Argentina), sistema NooJ.

Localización de términos en NooJ

Este programa, como ya se mencionó, posee entre sus recursos la posibilidad de localizar no solo términos o estructuras sintácticas, sino también terminaciones de palabras en grandes *corpus* de textos, a través de la pestaña *Locate*. Para efectuar la búsqueda de sufijos, se debe indicar de la siguiente forma:**<UNK+MP="ción$">**.

ISSN 2237-6321

v. 11 (esp.)
412-436
ago.
2021

Esto significa: UNK: palabra desconocida. Es importante aclarar que, si antes se analizó el texto con un diccionario, se debe consignar la etiqueta correspondiente a la categoría. Por lo tanto, para los sufijos coincidentes, habrá que colocar N (nombre). En tanto, para que reconozca a los sufijos existentes, pero no combinables con esas bases, habrá que emplear la etiqueta UNK (palabra desconocida); +MP=: indica que contiene el elemento que está a continuación del signo igual. También puede realizarse una búsqueda especificando que "no contenga" alguna secuencia, a través de: -MP; "ada$": es la secuencia de caracteres que se quiere localizar. El signo $ indica que se encuentre al final. Si no se le coloca el signo $ rastreará la secuencia en cualquier parte de la palabra.

Las Figuras 3 y 4, que se presentan a continuación, son capturas de pantalla que muestran la lista de ocurrencias de nombres con el sufijo –ción, en el *corpus* portugués B2 y con el sufijo –idad, en el *corpus* portugués A1:

Figura 3 - localización de sufijo -ción

NooJ - [Concordance for Text Untitled]

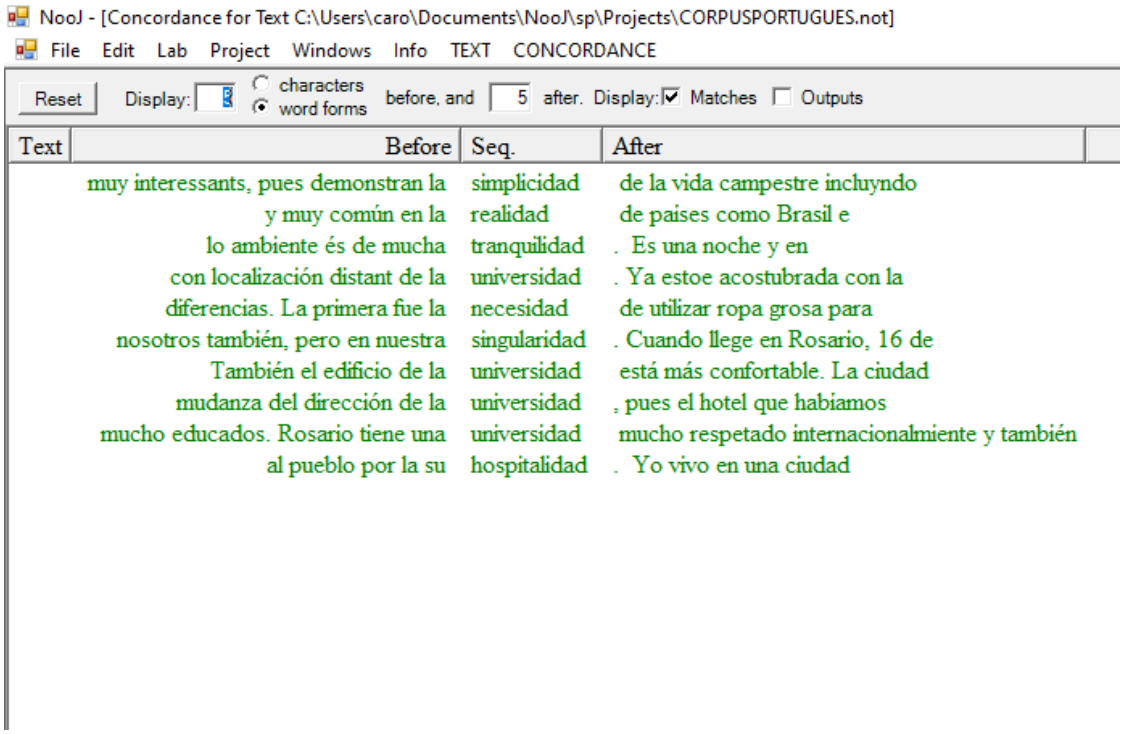
File Edit Lab Project Windows Info TEXT CONCORDANCE

Reset Display: 3 characters word forms before, and 5 after. Display: Matches Outputs

Text	Before	Seq.	After
	la movilidadsustentable, así reduciendo la	contaminación	al medioambiente, proponiendo también a
	Santiago, está sin servicios de	colección	de residuos hacen en la
	que son responsables de no	colección	de la basura, son muchos
	sea destruido. Leei en el	DiarioNación	sobre un modo que podriamos
	que yo haço reclamación para	restitución	de la energia electra. Sin
	y tampoco no desta ninguna	satisfación	. Mis vecinos también ya hicieron
	situación como esa. Es mui	dificilisituación	. Yo e mis vecinos esperamos
	su aparejo de respiración. Intentei	comunicación	por teléfono muchas veces, pero
	de pedilos un poco de	comprención	. Los vecinos más grandes, los
	Dejo abajo mi teléfono para	comunicación	y una respuesta positiva por
	terminando un poco con la	contaminación	y gasto de energia. Creo
	que tomen decisionesimportantes sobre esta	dramáticasituación	. Gracias por el tiempo. Saludos

Fuente: elaboración propia.

Figura 4 - localización de sufijo –idad



Fuente: elaboración propia.

Los nombres que posean sufijos que no corresponden al español deberán deducirse a través de un análisis cualitativo luego de efectuar el análisis automático de los textos. Esto se realiza haciendo *click* en *Unknows* a continuación del análisis morfosintáctico. De esta forma, el software arroja la lista de palabras que no están declaradas en los diccionarios. Del total de esas ocurrencias, se extraen los términos que contienen sufijos nominales inexistentes en español (SI).

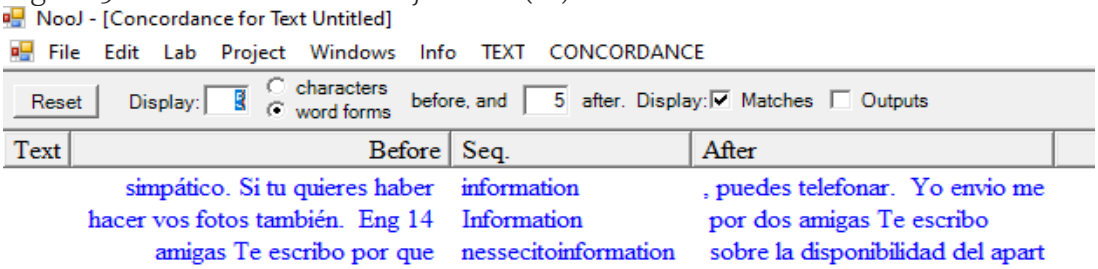
Posteriormente, se efectúa la búsqueda colocando ese sufijo no existente, para corroborar y contabilizar la cantidad de veces que aparece en cada una de las muestras.

Las siguientes capturas de pantalla exhiben la localización de palabras con el sufijo *-tion* (*corpus* inglés) y *–dade* (*corpus* portugués nivel intermedio), respectivamente:

ISSN 2237-6321

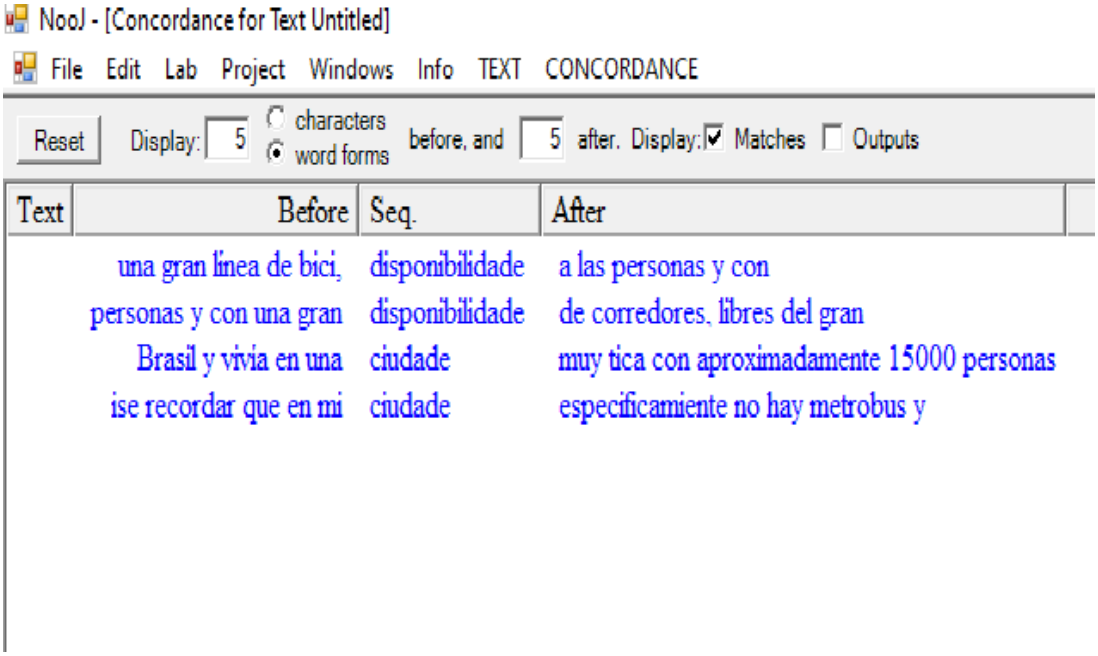
v. 11 (esp.)
412-436
ago.
2021

Figura 5 - Localización de sufijo *–tion* (SI)



Fuente: elaboración propia.

Figura 6 - Localización de sufijo *–dade* (SI).



Fuente: elaboración propia.

Análisis estadístico

Dado que el número de palabras por texto difiere en cada grupo de nivel y lengua, se efectúa la comparación sobre los sufijos en términos relativos al número de palabras, esto es, se calcula el porcentaje de sufijos coincidentes respecto a la cantidad de términos. La tabla que se encuentra a continuación expone un fragmento de las bases agrupadas que contabilizan la cantidad de SC, SI y SE por texto en cada uno de los grupos de lenguas nativas que componen los niveles de aprendizaje inicial e intermedio. Esta tabla se realiza mediante el programa Excel:

Tabla1 - Fragmento de Tabla Bases agregadas.

INTERMEDIO	TEXTO	CANTIDAD PALABRAS	SC	SI	SE
Portugués	1	184	17	2	1
Portugués	2	124	23	0	0
Portugués	3	169	34	0	1
Portugués	4	127	17	0	0
Portugués	5	391	24	0	1
Portugués	6	70	8	1	0
Portugués	7	186	20	0	1
Portugués	8	77	15	0	0

Fuente: elaboración propia.

A continuación, se aplica la siguiente regla:

$$\frac{\textit{Cantidad de sufijos coincidentes x 100}}{\textit{Cantidad de palabras en el texto}}$$

De esta forma, se realiza una descripción de las variables mediante el cálculo de medidas descriptivas por grupo y nivel y luego se utilizan técnicas de inferencia estadística para llevar a cabo las comparaciones de interés.

Las técnicas estadísticas empleadas son el Test No paramétrico de Wilcoxon para muestras independientes y la Prueba no paramétrica de Kruskal-Wallis.

El test de Wilcoxon para muestras independientes es la alternativa no paramétrica al test basado en la estadística t-Student para comparar dos medias poblacionales, mientras que el contraste de Kruskall-Wallis es la alternativa no paramétrica al análisis de la variancia, es decir, sirve para contrastar la hipótesis de que k>2 muestras han sido obtenidas de la misma población. Asimismo, en el caso de hallar diferencias, se procede a efectuar las comparaciones múltiples de a pares de grupos. Estas pruebas son adecuadas cuando se busca contrastar poblaciones cuyas distribuciones no son normales. La única exigencia sobre los datos se refiere a la aleatoriedad en la extracción de las muestras, pero no hacen referencia a ninguna de las otras condiciones necesarias para la validez de los resultados de una prueba paramétrica. Se basan en el uso de los rangos asignados a las observaciones (BELTRÁN-BARBONA, 2016).

ISSN 2237-6321

v. 11 (esp.)
412-436
ago.
2021

Finalmente, se compararon los sufijos coincidentes respecto a los niveles y dentro de cada muestra, respecto a las lenguas. Para la obtención de resultados, se exhibirá, en primer lugar, el análisis descriptivo y en segundo lugar, las medidas descriptivas para el número de palabras y porcentaje de sufijos según nivel de aprendizaje.

Los Cuadros 3 y 4 que siguen muestran el análisis descriptivo:

Cuadro 3 - Análisis descriptivo nivel inicial

NIVEL	Variable	n	Media	D.E.	Mín	Máx	Mediana
INICIAL	CANTIDAD DE PALABRAS	53	72,87	39,26	21,00	201,00	67,00
INICIAL	% SC	53	7,08	7,90	0,00	42,86	4,82
INICIAL	% SI	53	0,32	1,20	0,00	7,84	0,00
INICIAL	% SE	53	0,47	0,98	0,00	4,76	0,00

Fuente: elaboración propia.

Cuadro 4 - Análisis descriptivo nivel intermedio

NIVEL	Variable	n	Media	D.E.	Mín	Máx	Mediana
INTERMEDIO	CANTIDAD DE PALABRAS	28	160,18	106,14	17,00	507,00	138,50
INTERMEDIO	% SC	28	9,14	5,42	0,00	20,12	9,29
INTERMEDIO	% SI	28	0,21	0,50	0,00	1,79	0,00
INTERMEDIO	% SE	28	0,57	1,09	0,00	4,44	0,00

Fuente: elaboración propia.

Los cuadros que se presentan a continuación exponen las medidas descriptivas para el número de palabras y porcentaje de sufijos según lengua para cada nivel:

Cuadro 5 - Medidas descriptivas para palabras y porcentaje de sufijos de nivel inicial

GRUPO	Variable	n	Media	D.E.	Mín	Máx	Mediana
INGLÉS	CANTIDAD PALABRAS	18	77,28	45,44	35,00	188,00	57,50
INGLÉS	% SC	18	3,33	2,62	0,00	9,09	2,34
INGLÉS	% SI	18	0,26	0,84	0,00	3,39	0,00
INGLÉS	% SE	18	0,11	0,33	0,00	1,27	0,00
FRANCÉS	CANTIDAD PALABRAS	14	66,93	23,25	31,00	128,00	69,50
FRANCÉS	% SC	14	4,26	2,66	0,00	9,68	3,77
FRANCÉS	% SI	14	0,00	0,00	0,00	0,00	0,00
FRANCÉS	% SE	14	0,42	0,96	0,00	3,33	0,00

PORTUGUÉS	CANTIDAD PALABRAS	21	73,05	43,12	21,00	201,00	67,00
PORTUGUÉS	% SC	21	12,16	10,33	2,44	42,86	7,78
PORTUGUÉS	% SI	21	0,58	1,73	0,00	7,84	0,00
PORT	% SE	21	0,80	1,26	0,00	4,76	0,00

Fuente: elaboración propia.

Cuadro 6 – Medidas descriptivas para palabras y porcentaje de sufijos de nivel intermedio

GRUPO	Variable	n	Media	D.E.	Mín	Máx	Mediana
ALEMÁN	CANTIDAD PALABRAS	9	86,00	60,24	17,00	223,00	81,00
ALEMÁN	% SC	9	4,49	3,30	0,00	11,11	3,70
ALEMÁN	% SI	9	0,14	0,41	0,00	1,23	0,00
ALEMÁN	% SE	9	0,73	1,47	0,00	4,44	0,00
HOLANDÉS	CANTIDAD PALABRAS	5	213,40	165,55	112,00	507,00	154,00
HOLANDÉS	% SC	5	10,08	5,42	3,16	15,75	11,38
HOLANDÉS	% SI	5	0,36	0,80	0,00	1,79	0,00
HOLANDÉS	% SE	5	0,90	1,70	0,00	3,90	0,00
PORTUGUÉS	CANTIDAD PALABRAS	14	188,86	83,35	70,00	391,00	185,00
PORTUGUÉS	% SC	14	11,80	4,73	5,78	20,12	11,09
PORTUGUÉS	% SI	14	0,21	0,46	0,00	1,43	0,00
PORTUGUÉS	% SE	14	0,35	0,40	0,00	1,33	0,31

Fuente: elaboración propia.

Análisis de los resultados

En la comparación de niveles, se empleó el test de Wilcoxon para comparar el porcentaje de sufijos coincidentes en los niveles INICIAL e INTERMEDIO. Se encontraron diferencias significativas (p=0.013) que presentan una marcada la diferencia a favor (mayor) del nivel intermedio, en el que se observa una mediana de 9.3, mientras que en el inicial es de 4.8.

En cuanto al porcentaje de palabras formadas con sufijos nominales respecto al total de palabras de cada *corpus*, se obtuvo el 6,4 % en el nivel inicial y el 9,6 % para el nivel intermedio.

En la comparación del porcentaje de sufijos coincidentes dentro de cada estrato, se hallaron diferencias significativas en ambos casos. En el nivel INICIAL, se registraron diferencias entre las lenguas (p=0.0001)

ISSN 2237-6321

v. 11 (esp.)
412-436
ago.
2021

determinándose, que el portugués presenta un porcentaje de sufijos coincidentes significativamente mayor al inglés y francés (Medianas: 2.3, 3.8 y 7.8 respectivamente para inglés, francés y portugués). En lo que refiere al nivel INTERMEDIO, también se registraron diferencias entre las lenguas (p=0.0085). En este caso, el alemán muestra un porcentaje de sufijos coincidentes significativamente menor al holandés y portugués, entre los cuales no se hallaron discrepancias (Medianas: 3.7, 11.4 y 11.9 respectivamente para alemán, holandés y portugués).

Frecuencia de sufijos

La lengua española es una de las más productivas, como lo manifiesta Lang (1992), al comparar los procesos de derivación del español con los del inglés y el francés. Este autor propone dos variables para establecer la productividad de un afijo. El primero tiene que ver con la frecuencia con que es utilizado y el segundo, con el grado de transparencia que posee la relación entre la base y el afijo.

En cuanto a la frecuencia de sufijos nominales, el sufijo –ción es el que tuvo mayor cantidad de ocurrencias, con el 16,5 %. (Algunos ejemplos: educación, localización, dirección, sensación). A continuación, con el 11 % el sufijo -o (por ejemplo: gasto, encuentro, saludo, regreso) y con el 8 % el sufijo -a (ejemplos: danza, reforma, reserva). Estos son seguidos de –ado, con un 6 % (por ejemplo: traslado, cuidado, empleado).

Por lo tanto, se observa que son los derivados verbales los más utilizados, y dentro de esta categoría léxica, los verbos de la primera conjugación los elegidos para formar los nuevos términos, como por ejemplo: regresar, trasladar, emplear.

En cuanto a los sufijos nominales menos frecuentes (apenas con el 0,001%), se encuentran los que derivan nombres a partir de otros nombres o de adjetivos: -ismo (automovilismo), -logía (tecnología), -ura (escultura), -icia (justicia), -ivo (colectivo).

Es probable que la mayor presencia del sufijo –ción esté dada por el conocimiento del significado que se le atribuye como propio de los nombres y por el grado de transparencia de la base que elige. Lo mismo sucede con los otros sufijos nominales más empleados en el *corpus*: -o, -a, -ado que derivan sustantivos a partir de bases verbales.

Discusión y conclusiones

Como primera conclusión, el análisis estadístico de los datos obtenidos determina una mayor presencia de sufijos nominales en el nivel intermedio con respecto a lo que sucede en el nivel inicial. Esto se manifiesta concretamente en las medias: 4,8 para el nivel A frente a 9,3 para el B.

Es factible que el hecho de que los estudiantes empleen y produzcan nuevos términos se deba al conocimiento del léxico que va aumentando a medida que se avanza en el proceso de adquisición. Asimismo, dentro de la clasificación establecida para los tipos de sufijos nominales presentes en el *corpus*, se observa un aumento pronunciado en la cantidad de sufijos coincidentes (SC) con la lengua meta en el paso de un nivel a otro. De esta forma, mientras que la mediana de SC es de 4,8 para el nivel inicial, para el intermedio es de 9,29. En respuesta a esto, dentro de los grupos del nivel B, no se observan discrepancias con relación al holandés (10 %) y portugués (12 %) que pertenecen al estrato B2, pero sí las hay con relación al alemán, que corresponde a un B1 y, por lo tanto, evidencia una media inferior de sufijos coincidentes con el español (4,5%).

En cuanto a los grupos que integran el nivel inicial, el portugués es el que difiere del resto en tanto manifiesta no solo una mayor cantidad de sufijos coincidentes sino también un aumento de sufijos inexistentes y no combinables (1,4% frente a 0,34% y 0,36% del francés e inglés, respectivamente). Creemos que en estos resultados influyen distintas cuestiones. Por un lado, la situación contextual en la que los sujetos realizan la producción textual ya que los estudiantes brasileños responden a consignas de escritura durante sus clases de español mientras que los aprendientes cuyas lenguas de origen son el inglés y francés lo hacen con el objetivo de aprobar un examen de proficiencia. Esta diferencia podría otorgar una mayor libertad a la hora de introducir nuevas formas léxicas. Por otro lado, la mayor cantidad de sufijos inexistentes del grupo portugués puede explicarse por el grado inferior de instrucción recibida (A1) con respecto a los grupos inglés y francés que pertenecen al estrato A2. A esto podemos agregar, en coincidencia con las observaciones de Campillo Llanos (2014) que la mayor proximidad de la lengua portuguesa con la española puede favorecer el proceso productivo en sujetos luso hablantes, pero también interferir negativamente, en las primeras instancias de aprendizaje.

ISSN 2237-6321

v. 11 (esp.)
412-436
ago.
2021

Respecto de los sufijos existentes en español, pero no combinables con determinada base (SE), hallamos que el porcentaje aumenta del nivel inicial al intermedio apenas de 0,47% a 0,57% sobre el total de sufijos nominales. Este hecho puede interpretarse como resultado de las estrategias de comunicación que llevan a los aprendientes

> a producir elementos léxicos que no existen en la lengua, pero que han sido construidos respetando las restricciones impuestas por las reglas de buena formación morfoléxica y conceptual (BARALO, 2001).

Por otra parte, casi en la misma proporción, disminuye la cantidad de sufijos inexistentes (SI) en la lengua española de 0,32% en el *corpus* A 0,21% en el B. Por consiguiente, se obtienen porcentajes similares de formas idiosincrásicas en ambas muestras, y estas pequeñas diferencias entre los diferentes niveles de instrucción no son significativas como ocurre con los SC.

Concluimos, por lo tanto, que estos fenómenos subsisten en gran medida en la interlengua a través del paso del tiempo e incluso de un nivel al otro, en coincidencia con estudios anteriores (CAMPILLO LLANOS, 2014). Estas formas idiosincrásicas responden a las tácticas que emplea el estudiante para encontrar el término léxico que necesita, como explica Baralo (2001, p. 33) "con estas estrategias llega más allá del límite de lo que sabe y arriesga".

Como discusión, esperamos incrementar los *corpus* de trabajo y analizar producciones de aprendices que se encuentren en niveles avanzados de aprendizaje para poder establecer si estas formas idiosincrásicas, propias de los primeros estadios en la adquisición del español como segunda lengua acontecen y en qué medida, o si por el contrario, se han superado.

En síntesis, este artículo cumplió con el objetivo propuesto de localizar y contabilizar los sufijos nominales a través del sistema NooJ, tanto para las palabras coincidentes, declaradas en el diccionario correspondiente a Nombres como para las formas propias de interlengua, rastreadas como palabras desconocidas en el análisis automático del *corpus* y contabilizadas, luego, a partir de su terminación.

De este modo, permitió contribuir al ámbito de los estudios de interlengua en lo que refiere a la adquisición de la morfología derivativa del español, por medio del análisis estadístico mediante dos pruebas diferentes para comparar el empleo y la frecuencia de sufijos nominales en muestras textuales pertenecientes a estudiantes de nivel inicial (A1 y A2) e intermedio

(B1 y B2). Consideramos que la investigación no se agota con este aporte, por el contrario, se espera ampliar las muestras en cuanto a lenguas nativas y nivel de proficiencia para reforzar las hipótesis esgrimidas.

Referencias

AGARONAIN, E. **Los procesos de formación de palabras en español: análisis y propuesta didáctica en la enseñanza de e/le.** 2019. Tesis doctoral. Universidad de Valladolil, España, 2019. Disponible en: https://core.ac.uk/download/pdf/286337786.pdf. Acceso en 15 ago. 2020.

ALBA QUIÑONES, V. El análisis de errores en el campo del Español como Lengua Extranjera: Algunas cuestiones metodológicas. **Revista Nebrija de Lingüística Aplicada**, v.5, n.3, p. 1-16, 2009.

ALEXOPOLOU, A. La función de la interlengua en el aprendizaje de lenguas extranjeras. **Revista Nebrija de Lingüística aplicada**, n.9, 2010.

BARALO, M. El lexicón no nativo y las reglas de la gramática en **Estudios de Lingüística,** Pastor Cesteros, S. y Salazar García, V. (eds). Universidad de Alicante, España, 2001. Disponible en: https://rua.ua.es/dspace/bitstream/10045/6689/1/EL_Anexo1_02.pdf. Acceso en: 10 nov. 2020.

BELTRÁN, C.; BARBONA I. **La estadística en la investigación. Elementos básicos y aplicaciones.** Rosario: Ediciones del Revés, 2016.

CAMPILLO LLANOS, L. **Errores léxicos en el español oral no nativo**: análisis de la interlengua basado en corpus. 2014. Disponible en: https://rua.ua.es/dspace/bitstream/10045/48502/1/ELUA_28_04.pdf. Acceso en: 10 sept. 2020.

CORDER, S. P. **The significance of Learner's errors.** IRAL, 5, p. 161-170, 1967. (Traducido en Liceras 1992, p. 31-40).

CORDER, S. P. **Idiosincratic Dialects and Error Analysis**. IRAL, v.9, p. 161-170, 1971. (Traducido en Liceras 1992, p. 61-67)

KOTZ GRABOLE, G.; FERREIRA CABRERA, A. La precisión gramatical mediada por la tecnología: El análisis y tratamiento automático de errores. **Literatura y Lingüística**, n. 27, p. 219-242,2013.

FERREIRA, A. et al. La Arquitectura de ELE-TUTOR: Un Sistema Tutorial Inteligente para el Español como Lengua Extranjera. **Revista signos**, n.45, v.79, p. 102-131, 2012. Disponible en: https://dx.doi.org/10.4067/S0718-09342012000200001. Acceso en: 20 mayo 2020

FERREIRA, A.; ELEJALDE GOMEZ, J.; VINE, A. Análisis de Errores Asistido por Computador basado en un Corpus de Aprendientes de Español como Lengua Extranjera. **Revista signos**, n.47, v.86, p. 385-411, 2014 Disponible en: https://dx.doi.org/10.4067/S0718-09342014000300003 Acceso en: 11 mayo 2020.

FERREIRA CABRERA, A.; ELEJALDE GOMEZ, J. Propuesta de una taxonomía etiológica para etiquetar errores de interlengua en el contexto de un corpus escrito de aprendientes de ele. **Forma. func**., v.33, n.1,. p. 115-146, 2020. DOI: https://doi.org/10.15446/fyf.v33n1.84182. Acceso en: 04 de agosto de 2020.

ISSN 2237-6321

v. 11 (esp.)
412-436
ago.
2021

LANG, M.F. **Formación de palabras en español:** morfología derivativa productiva en el léxico moderno, Madrid: Cátedra, 1992.

LAVID, J. **Lenguaje y nuevas tecnologías.** Nuevas perspectivas, métodos y herramientas para el lingüista del siglo XXI. Madrid: Cátedra, 2005.

LICERAS, J. M. La interlengua del español en el siglo XXI. **Revista Nebrija de Lingüística Aplicada**, n.5, p. 36-49, 2009. Disponible en: https://www.nebrija.com/revistalinguistica/files/articulosPDF/articulo_5316fc5066e5c.pdf. Acceso en: 20 jun. 2020.

PARODI, G. Textos de especialidad y comunidades discursivas técnico-profesionales: una aproximación basada en corpus computarizado. **Estud. filol.**, Valdivia, n. 39, p. 7-36, sept. 2004 . Disponible en <https://scielo.conicyt.cl/scielo.php?script=sci_arttext&pid=S0071-17132004003900001&lng=es&nrm=iso>. Acceso en: 09 oct. 2020. http://dx.doi.org/10.4067/S0071-17132004003900001.

RIGAMONTI, D. **Problemas de lingüística de la adquisición y enseñanza del E/ELE a italófonos**.2006. Disponible en: https://www.ledonline.it/ledonline/rigamonti/Rigamontiproblemas.pdf. Acceso en: 23 nov. 2019.

SALAZAR GARCÍA, V. Aprendizaje del léxico en un currículo centrado en el alumno. En: Miquel, L. y N. Sans (eds.): **Didáctica del español como lengua extranjera.** Madrid, Fundación Artilibre, 1994. p. 165-197.

SELINKER, L. Interlanguage. **International Review of Applied Linguistics**, n. 10, v.3, p.209-231, 1972.

SILBERZTEIN, Max. **Nooj Manual**, [en línea], 2003. Disponible en: <http://www.nooj4nlp.net/NooJManual.pdf>. Acceso en: 20 abr. 2020.

TRAMALLINO, C. P. Análisis morfológico con herramientas informáticas. Reconocimiento de nombres en textos de español con el sistema NooJ. **Revista Lingüística y Literatura,** n. 63, p. 33-48, 2013.

TRAMALLINO, C. P. **Análisis morfológico automático de la interlengua de aprendientes de español como L2.** Tesis doctoral. Universidad Nacional de Rosario, Argentina, 2016. Disponible en: http://hdl.handle.net/2133/14768. Acceso en: 9 oct. 2020.

TORIJANO, J. A. El estudio de los determinantes en aprendices luso hablantes de español, **DICENDA. Cuadernos de Filología Hispánica**, n. 26, p. 235-257, 2008.

TORRUELLA J.; LLISTERRI J. Diseño de corpus textuales y orales. En: BLECUA J. M., CLAVERÍA G. SÁNCEZ C. TORUELLA, J. (editores). **Filología e informática**, Univ. Autónoma de Barcelona, Editorial Milenio, p. 45-77, 1999.

VARELA, S. Léxico, morfología y gramática en la enseñanza de español como lengua extranjera. **ELUA. Estudios de Lingüística**, n. 17, p. 571-588,2003.

WILLIAMS, E. "Remarks on Lexical Knowledge", **Lingua**, 92, p . 7-34, 1994.