



Reconocimiento automático de sintagmas nominales en producciones escritas de aprendientes brasileños de español



Carolina Paola Tramallino
Romina Paola Arnal

Recibido: Mayo 2019 – Aceptado: Junio 2009

Cátedra "Desarrollos Psicológicos Contemporáneos A"
Facultad de Psicología UNR

✉: carotramallino@hotmail.com

1. Introducción

Este trabajo forma parte de un proyecto de investigación radicado en la Facultad de Humanidades y Artes sobre la enseñanza de español como segunda lengua mediante el empleo de herramientas informáticas.

El objeto de estudio radica en la adquisición del español como segunda lengua y el problema de investigación hace referencia a las dificultades en el nivel sintáctico que se detectan en las producciones escritas de estudiantes brasileños que realizan exámenes de proficiencia de español en la Universidad Nacional de Rosario.

Atendiendo al hecho de que Argentina y Brasil pertenecen al MERCOSUR y están involucrados como países hermanos y limítrofes no sólo en convenios económicos y comerciales sino también en lo que refiere a relaciones académicas y profesionales es de suma importancia contribuir en el campo de la enseñanza del español. Para una enseñanza de forma más rápida y eficaz, se hace necesario el uso de herramientas que faciliten el aprendizaje de los alumnos como: el empleo de recursos diversos y nuevas tecnologías, todo lo que puede llevar al éxito de la enseñanza y aprendizaje del alumno para que desarrolle las habilidades comunicativas en el nuevo idioma.

Respecto a las observaciones realizadas en el corpus reunido, encontramos construcciones sintácticas propias de la interlengua de aprendientes de español que difieren de las estructuras del español estándar, por ejemplo cuando se asigna otro género al sustantivo en sintagmas nominales como: "la viaje".

En cuanto al objetivo de la investigación, ésta se propone indagar sobre las problemáticas en el nivel sintáctico observadas en la interlengua de aprendientes brasileños de español y lograr mediante el uso del sistema Nooj que los estudiantes detecten y corrijan formas y estructuras desviadas en sus producciones escritas.

La hipótesis de trabajo es que pueden reconocerse sintagmas propios de la interlengua de aprendientes de español mediante la herramienta informática mencionada.

Para ello es necesario explicar a qué refiere el concepto de interlengua y desde qué enfoque se analiza el corpus.

El término **interlengua** es utilizado por Selinker en 1972, proponiendo un enfoque psicolingüístico y colocando a esta lengua en un lugar intermedio entre la lengua materna y la meta, teniendo elementos comunes con ambas pero definiéndolo como un sistema lingüístico independiente.

Es un sistema lingüístico con una gramática propia que se encuentra en continuo cambio, sus oraciones son idiosincrásicas, no erróneas y es un sistema peculiar de un estudiante individual o de un grupo de estudiantes que posee la misma formación académica.

Con los estudios de la interlengua entonces, el investigador considera toda la producción del aprendiz, es decir, tanto la correcta como la incorrecta.

La relevancia de los estudios de la interlengua, coincidiendo con Alexopoulou (2010: 2)

“... surge de la necesidad de descubrir los principios generales que determinan el aprendizaje de la lengua extranjera, según los cuales el cerebro humano de los aprendientes procesa los datos del lenguaje a los cuales están expuestos.”

Es importante aclarar que la interlengua contiene reglas propias y no debe tomarse como una mezcla entre los dos sistemas lingüísticos. La literatura especializada le otorga dos características que sin embargo parecen contradictorias: sistematicidad y variabilidad.

La primera refiere al conjunto de reglas coherente a partir de las cuales los aprendientes realizan hipótesis, sin embargo éstas son variables, se modifican a medida que el estudiante avanza en la adquisición y pasa a otra etapa. A partir del nuevo estadio se reestructuran esas reglas e hipótesis.

La sistematicidad entonces está dada por la coherencia interna que presenta esa interlengua en un determinado momento de su desarrollo, se vislumbra en la aplicación de reglas lingüísticas que responden a estrategias y procesos que activan los aprendientes. En esta gramática singular, encontraremos oraciones o construcciones “desviadas” que sin embargo, serán adecuadas desde el punto de vista del aprendiente.

Es importante destacar que ese sistema lingüístico tiene características y peculiaridades propias que coinciden con las halladas en el sistema de todo estudiante que está adquiriendo una segunda lengua, si se toma en cuenta un mismo nivel de aprendizaje.

2. Materiales y método

La metodología de trabajo implicó primero la conformación de un corpus, luego la digitalización de éste y posteriormente la descripción de la sintaxis. Por último la adaptación del programa informático Nooj.

2.1. Análisis del corpus

El corpus reúne producciones escritas de aprendientes brasileños de español como segunda lengua, los cuales se encuentran en un nivel inicial.

Los estudiantes pertenecen a dos grupos diferentes: por un lado a doctorandos de mediana edad que se encuentran realizando el Doctorado en Ciencias de la Educación en la Universidad Nacional de Rosario (UNR) y por otro lado, a jóvenes ingresantes a carreras de grado de la UNR que deben rendir un examen de proficiencia de español.

Las consignas están pautadas por el docente y el contexto de producción entonces es una situación de examen.

2.2. Análisis de los sintagmas propios de la interlengua de aprendientes de español

El sintagma es una estructura intermedia que funciona como constituyente de la oración. El sintagma nominal (SN) es una construcción que se agrupa o articula en torno a un nombre, este sustantivo es el núcleo de dicha estructura.

A continuación se agrupan los ejemplos de asignación incorrecta de género divididos en tres casos:

Ejemplos de tipo I

- (1) "extrañé mucho **la lenguaje**"
- (2) " **la desenvolvimiento**"
- (3) "Aguardo la respuesta sobre **el bonificación**"
- (4) "**Esas hombres**"
- (5) "**La paisaje**"

Ejemplos de tipo II

- (1) "es **un gran oportunidad** para mi carrera de postgrado en tecnología"
- (2) "como debemos vivir para tener **un buen salud**"
- (3) "**el enseñanza medio...**",
- (4) "en **la tema lúdico**"

Ejemplos de tipo III

- "en torno de **lo fuego**"
- "**lo ambiente**"
- "**lo edificio** de la UNR"
- "**lo nuevo edificio**"
- "donde puede notar que **lo departamento**"
- "**Lo chico** Vinicius Alexandre hace un trabajo"
- "**Lo gobierno** realizo buenísimo ideas"
- "por medio de **lo funcionamiento físico**"

En los ejemplos del tipo I y II se le adjudica un género erróneo al nombre, núcleo del sintagma nominal. Por ejemplo, el sufijo nominal -aje posee género femenino en

portugués y por eso se le atribuye el mismo género en palabras del español: "La *paisaje*, la *lenguaje*". Por otra parte se desconoce que el sufijo nominal -ción posee género femenino y por eso hallamos sintagmas del tipo: "el *bonificación*", además de no saber qué género tienen los sustantivos terminados en consonante como "salud" o con la vocal e: "novidades".

En cuanto a los ejemplos del tipo 3, se forma el sintagma nominal con el clítico masculino lo, éste es empleado como artículo, es decir que se emplea otra categoría sintáctica, dada la similitud con el artículo del portugués o.

3. Herramienta informática: sistema Nooj

La utilización de este sistema permite la extracción de información en grandes corpus de textos y requiere de una formalización lingüística por parte de los usuarios.

Incluye herramientas para crear y mantener fuentes lexicales así como gramáticas sintácticas y morfológicas.

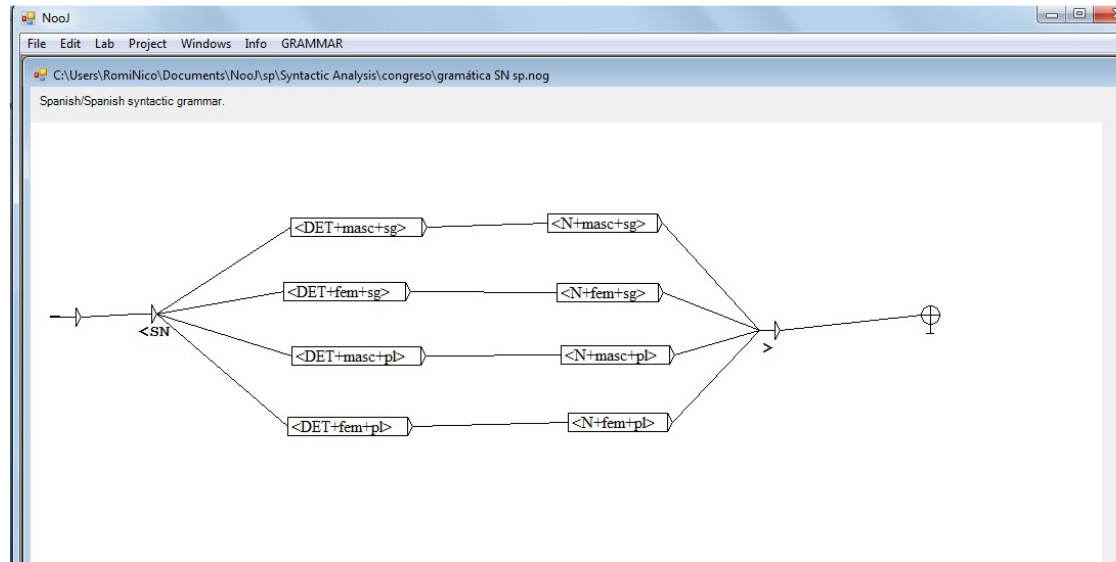
Tanto para que el sistema reconozca, por ejemplo, los nombres en un texto, como para que analice morfológicamente enunciados es necesario crear nuevos textos e ingresar los ítems léxicos, modelos de flexión y categorías con las que debe operar el software. Estos se consignan en tres archivos de diferente extensión: los modelos se ubican en el archivo Gramática («Grammar») que posee extensión .nog, los lemas se ingresan en el archivo Diccionario («Dictionary») con extensión .dic y, por último, las categorías gramaticales con sus rasgos, por ejemplo, Nombre: género/ número, en el archivo Propiedades («Properties ' definition»), que tiene extensión .def

Existen dos tipos de recursos lingüísticos:

- Diccionarios: (archivos con terminación .dic) asocian palabras o expresiones a por ejemplo una categoría, como verbo, a uno o más paradigmas de inflexiones o derivaciones (por ejemplo, cómo conjugar verbos), y también asocian a una o más propiedades sintácticas y semánticas.
- Gramáticas: se usan para representar fenómenos lingüísticos, desde niveles ortográficos y morfológicos a niveles sintagmáticos. Contiene tres tipos de gramáticas:
 - a. Gramáticas derivacionales (o archivos con terminación .nof).
Se emplean para representar las propiedades flexivas, por ejemplo, conjugaciones o derivaciones como las nominalizaciones de las entradas léxicas. Estos modelos pueden declararse de forma gráfica o mediante reglas.
 - b. Gramáticas lexicales, ortográficas, morfológicas o terminológicas (archivos con terminación .nom).
Se usan para representar conjuntos de tipos de palabras y asociarlas con información léxica; por ejemplo: estandarizar la ortografía de palabras o de variantes, reconocer neologismos, asociar expresiones sinónimas.
 - c. Gramáticas semánticas o sintácticas (archivos con terminación .nog)
Son utilizadas para reconocer o anotar expresiones en textos como, por ejemplo, etiquetar tanto a frases nominales como a ciertas expresiones sintácticas o idiomáticas. Se emplean para extraer o desambiguar palabras filtrando algunos ítems léxicos o anotaciones sintácticas en el texto.

Los paradigmas inflexión-derivación son formalizados como bibliotecas de gráficos estructurados o reglas basadas en textos. Se realizan mediante grafos.

En la captura de pantalla N°1 que se presenta a continuación, se observa una gramática del sintagma nominal propio del español estándar (SN).



Captura de pantalla N°1:
Gramática de grafos para SN español estándar.

3.1. Adaptación del sistema Nooj

Como se mencionó en el apartado anterior, el sistema Nooj permite a los usuarios crear, mediante grafos, *gramáticas productivas sintácticas*.

Productivas, porque se emplean categorías. Por ejemplo: DET para "determinante"; N para "nombre"; ADJ para "adjetivo".

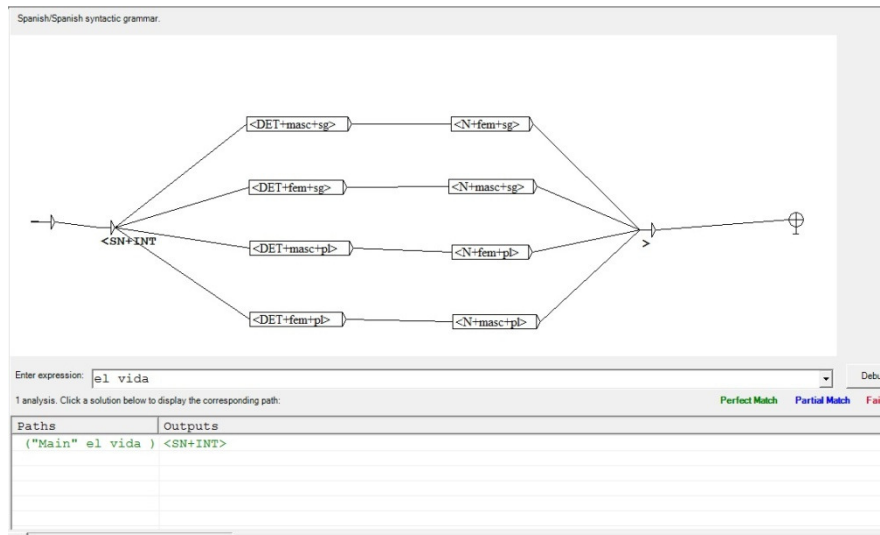
Sintácticas, porque agrupa dos o más palabras y las etiqueta. Por ejemplo: SN para "Sintagma Nominal".

Precisamente, al ser los usuarios los que tienen la posibilidad de crear las etiquetas de las expresiones sintácticas, se pueden realizar adaptaciones para reconocer sintagmas propios de la interlengua.

El programa proporciona el nodo inicial (\triangleright) y el nodo final (\oplus), por lo que son los usuarios los que deben crear los nodos intermedios y unirlos.

Entonces, para crear una gramática del sintagma nominal perteneciente al español estándar, se pueden crear nodos de determinantes y nombres con sus rasgos de género y número y unirlos estableciendo la concordancia entre ellos. Además pueden agregarse nodos de adjetivos y así seguir ampliando el SN.

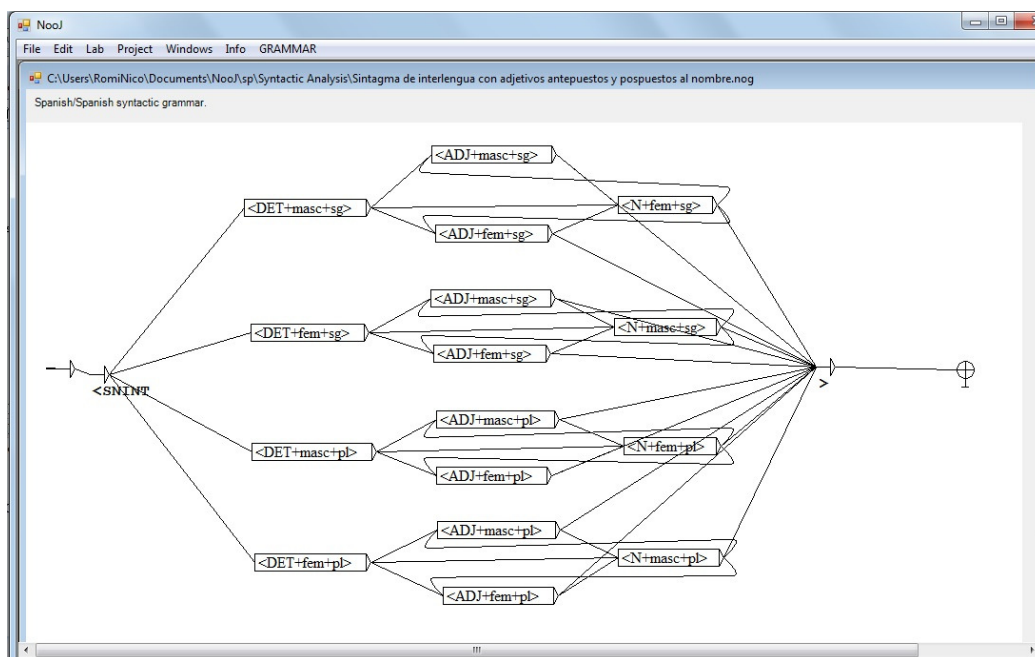
De la misma manera, se puede crear una gramática con la etiqueta SN INT (Sintagma Nominal de Interlengua) en la que los nodos de DET y N estén unidos de tal manera que no haya concordancia de género entre ellos, aunque sí de número. (Como se muestra en la captura de pantalla N° 2)



Captura de pantalla N°2: Gramática sintáctica productiva tipo I

Nooj también permite verificar si la gramática está bien creada, con la acción "Show Debug", que se aprecia en la ventana inferior de la captura de pantalla N° 2, donde se lee que el sintagma "el vida" es reconocido como SN INT.

Esta gramática permite reconocer los ejemplos que se han clasificado como del tipo I. Por otro lado, una gramática para un sintagma nominal de interlengua más amplio, que incluye adjetivos, permite identificar los ejemplos de tipo II.



Captura de pantalla N° 3: Gramática sintáctica productiva tipo II.

En esta gramática los nodos se unen de tal manera que los SN INT pueden estar formados por DET, ADJ, N o por DET, N, ADJ, en los cuales, lo que va a fallar siempre es la concordancia de género entre DET y N, mientras que el ADJ puede concordar con el N o con el DET indistintamente.

Esta gramática presenta los nodos unidos de tal manera que también puede identificar los ejemplos de tipo I, por lo que reemplaza a la anterior.

Por último, los ejemplos del tipo III necesitan una gramática distinta. En estos se observa la confusión entre el artículo masculino singular y el clítico (CL) "lo".

A continuación se muestra la gramática mencionada en la captura de pantalla N° 4:

The screenshot shows the NooJ software interface. The main window displays a syntactic grammar diagram with nodes: <SNINT>, <CL+c3+3era+sg+masc+ac>, <N+masc+sg>, and <ADJ+masc+sg>. Below the diagram, the 'Enter expression:' field contains 'lo nuevo edificio'. The analysis results show a 'Perfect Match' for the path ('Main' lo nuevo edificio) with the output <SNINT>.

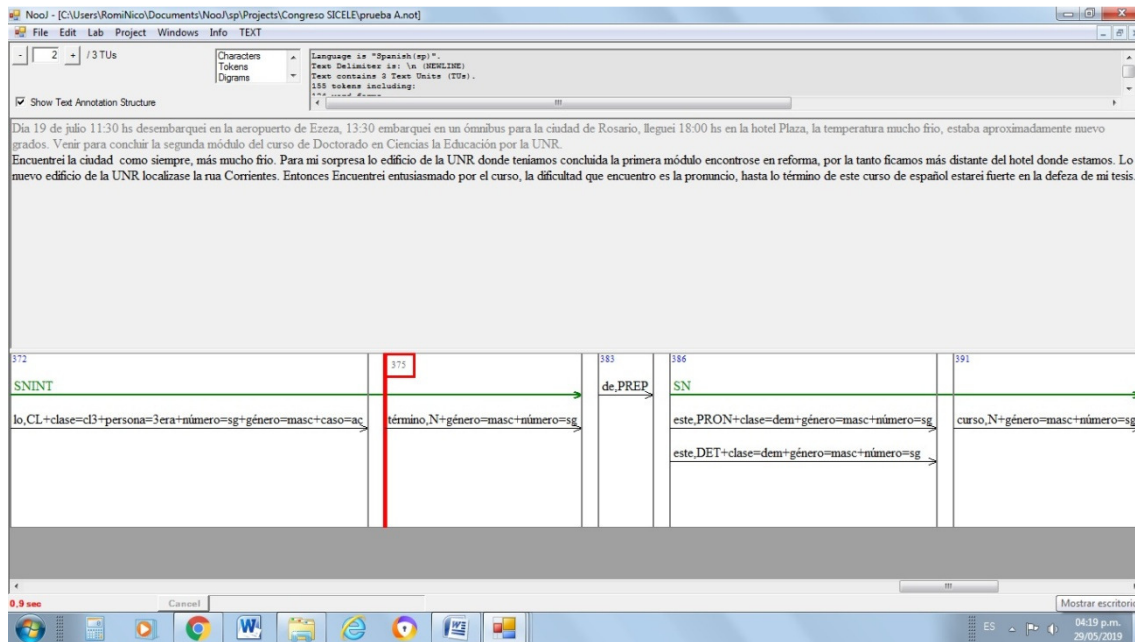
Paths	Outputs
("Main" lo nuevo edificio)	<SNINT>

Captura de pantalla N° 4: Gramática sintáctica productiva tipo III.

Una gramática como la que se observa en la captura de pantalla N° 4 permite identificar SN INT de CL y N ("lo edificio"), de CL, N, ADJ ("lo edificio nuevo") o de CL, ADJ, N ("lo nuevo edificio").

4. RESULTADOS Y DISCUSIÓN

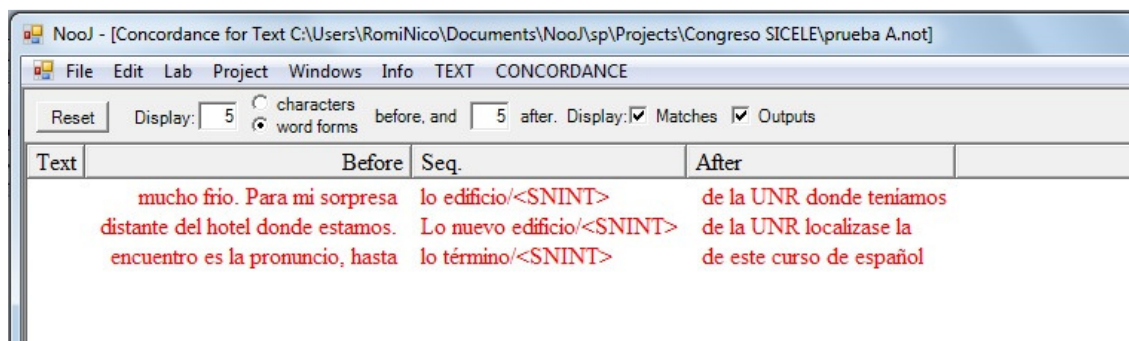
Una vez creadas las gramáticas, se pueden reconocer o analizar estas estructuras en un texto. Para esto, se deben asociar estas gramáticas así como los diccionarios de categorías léxicas que se necesitan para que el programa reconozca las palabras que debe analizar (N, ADJ, DET, CL).



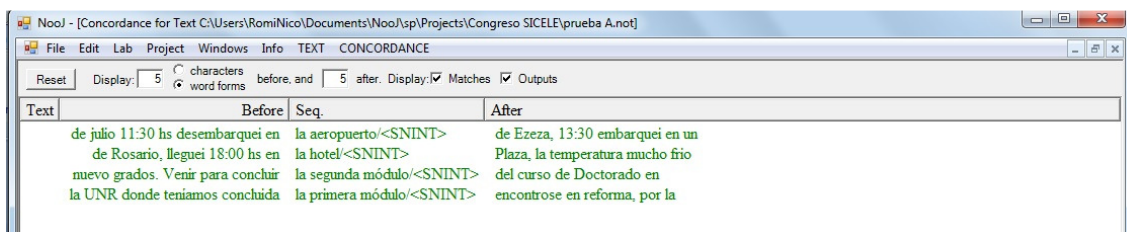
Captura de pantalla N° 5: análisis de sintagmas nominales de interlengua.

El texto que se muestra en la captura de pantalla N° 5 pertenece al corpus de los cursos de español de doctorandos brasileños. Se observa cómo la herramienta analiza y etiqueta tanto al SN coincidente con el español estándar "este curso de español" así como al SNINT "lo término".

Además, NooJ posibilita la observación de todas las secuencias del texto que reconoce la gramática, como puede observarse en las capturas de pantalla N° 6 y 7 que siguen a continuación:



Captura de pantalla N° 6: reconocimiento de sintagmas de interlengua con clítico "lo" en un texto.



Captura de pantalla N° 7: Reconocimiento de sintagmas de interlengua en un texto.

De esta manera, en grandes corpus, los investigadores tienen la posibilidad de extraer datos estadísticos con respecto a la cantidad de estructuras, como en este caso, o de términos o categorías.

Respecto a la medición de los resultados de acuerdo al índice de cobertura y precisión, la precisión se define como el cociente entre el número de ítems correctamente analizados y el número total de ítems analizados, mientras que la cobertura es el cociente entre el número total de ítems y el número de ítems correctamente analizados.

Para los ejemplos del tipo I el reconocimiento automático arroja una cobertura del 100 por ciento y una precisión del 100 por ciento.

En cuanto a los ejemplos del tipo II y III obtenemos los mismos resultados: una cobertura de la totalidad de los sintagmas y una precisión del 100 por ciento.

Por lo tanto, el programa se constituye en una herramienta válida, a partir de la creación de gramáticas sintácticas productivas, para el reconocimiento de sintagmas nominales pertenecientes al español estándar y de sintagmas nominales propios de la interlengua de aprendientes de español como segunda lengua.

5. Referencias

- (1) ALEXOPOLOU, A. "La función de la interlengua en el aprendizaje de lenguas extranjeras" en *Revista Nebrija de Lingüística aplicada*, 2010, 9.
- (2) BOSQUE, I. - GUTIÉRREZ-REXACH. *Fundamentos de sintaxis formal*. Akal. Madrid, 2009.
- (3) SELINKER, L. "Interlanguage", en *International Review of Applied Linguistics*, num 10.3 pp.209-31. 1972.
- (4) SILBERZTEIN, M., *Nooj Manual*, [en línea], Disponible en <<http://www.nooj4nlp.net/NoojManual.pdf>>. 2003. Fecha de consulta: diciembre de 2018.
- (5) ARNAL, R.- TRAMALLINO, C. "La enseñanza de español como I2 a doctorandos brasileños mediante la aplicación de herramientas informáticas" (en prensa). en Libro electrónico de las XII Jornadas de Ciencia y Tecnología 2018 SeCyT-UNR. Libro de trabajos ampliados. UNR Editora, Rosario. 2018.