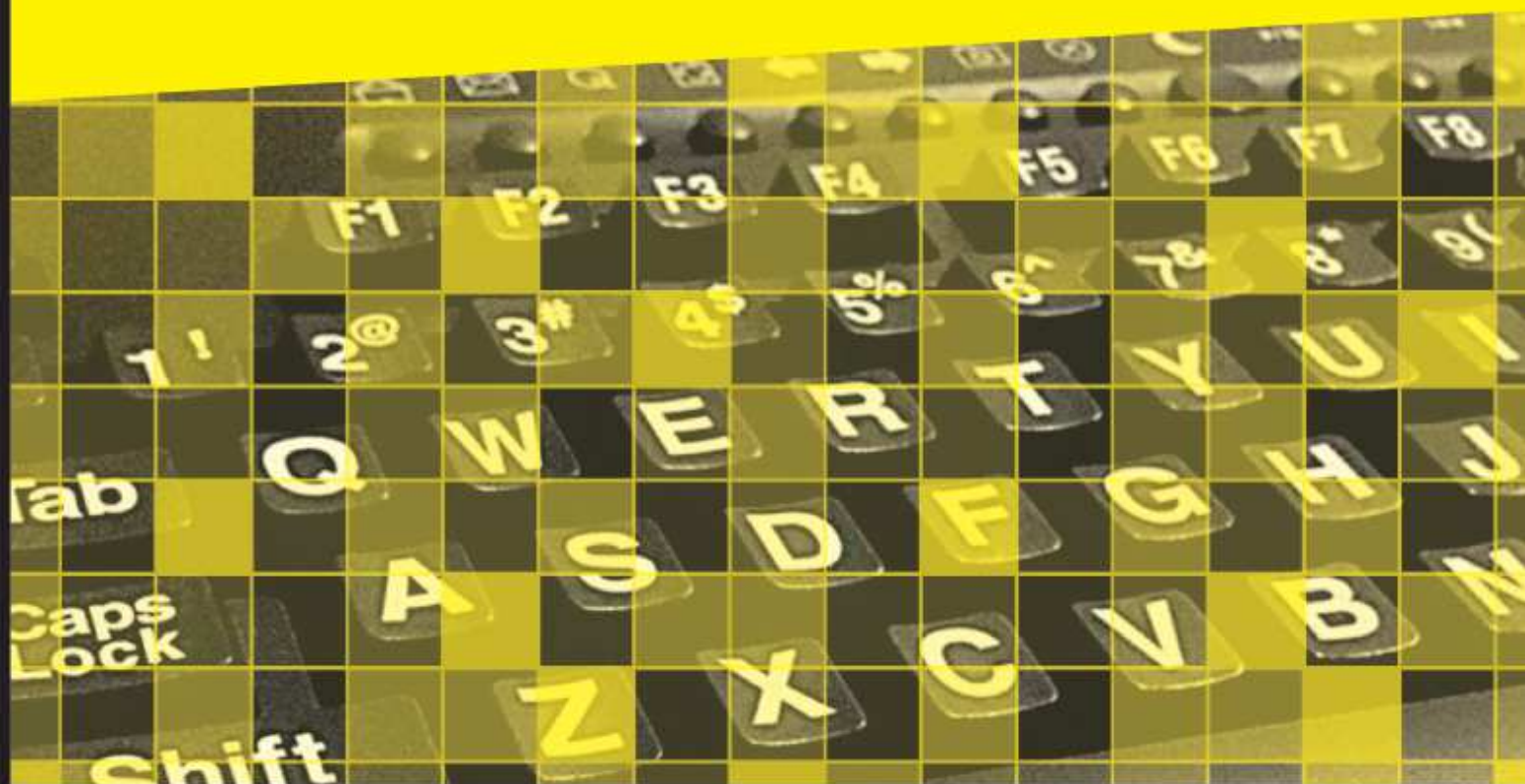


Andrea Rodrigo y Rodolfo Bonino

Aprendo con NooJ

*De la Lingüística Computacional
a la Enseñanza de la Lengua*



Learning with NooJ

From Computational Linguistics
to Language Teaching

Trad. Silvia Reyes

Rodrigo, Andrea

Aprendo con NooJ : de la lingüística computacional a la enseñanza de la lengua / Andrea Rodrigo ; Rodolfo Bonino. - 1a ed. - Rosario : Ciudad Gótica, 2019.
100 p. ; 28 x 20 cm.

Traducción de: Silvia Reyes.
ISBN 978-987-597-398-5

1. Enseñanza de Lenguas Extranjeras. I. Bonino, Rodolfo.
II. Reyes, Silvia, trad. III. Título.
CDD 418.0071

ISBN 978-987-597-398-5

Comunicación con los autores:
andreafrodrigo@yahoo.com.ar
rodolfobonino@yahoo.com.ar

Diseño y Maquetación: Sergio Gioacchini

Editorial Ciudad Gótica
Carrasco 1365 - Cel. 0341- 156 096738 - 2000 Rosario
www.cgeditorial.com.ar / ciudadgotica@hotmail.com

Impreso en Argentina - Printed in Argentina

Primera edición: agosto, 2019

Prohibida la reproducción total o parcial por cualquier medio visual, gráfico o sonoro sin la expresa autorización de los autores.

«Aprendo con NooJ»

De la Lingüística Computacional a la Enseñanza de la Lengua

Andrea Rodrigo y Rodolfo Bonino

Colaboradores

Equipo de investigación dirigido por Andrea Rodrigo (2015-2017)

Investigadores

Silvia Reyes

Luciana Andrín

Alejandra Rodríguez

María Andrea Fernández Gallino

Paula Alonso

Soledad Harriet

Participación de las alumnas:

Malen D'Urso (Instituto Superior del Profesorado N° 28, «Olga Cossettini»)

Camila Ferramondo (Facultad de Humanidades y Artes, UNR)

Comité Científico Asesor

Dr. Max Silberztein, Universidad de Franche-Comté, Francia

Dr. Xavier Blanco, Universidad Autónoma de Barcelona, España

Dr. Gerardo Kahan, Universidad Nacional de Rosario, Argentina

ÍNDICE

Prólogo a cargo de Xavier Blanco	7
PRIMERA PARTE	
Capítulo 1 Presentación	
1- El equipo de investigación	9
2- Estructura del libro	10
3- Qué es la lingüística computacional	10
4- La plataforma NooJ	11
5- Qué hace NooJ	11
6- Diccionarios NooJ	14
7- Gramáticas NooJ	14
Capítulo 2 El Corpus de trabajo	
1- El texto	15
2- Análisis del corpus de textos periodísticos	15
3- Contraste con las producciones de jóvenes, fortalezas y debilidades. Datos que brinda la página	19
4- Análisis de los textos producidos por los jóvenes	20
5- Comparación entre el corpus de textos periodísticos y el corpus de textos producidos por los jóvenes	20
6- Opiniones de los alumnos	21
• Experiencia en el «Taller de lingüística aplicada: ‘La herramienta NooJ y su aplicación en el estudio de la morfología del español» (2017)	21
• Experiencia en el Seminario orientado en Lingüística y Lenguas « <i>Aportes de la lingüística informática en la enseñanza de español como L1 y L2</i> » a cargo de las profesoras Silvia Reyes y Andrea Rodrigo (2017)	22
SEGUNDA PARTE	
Capítulo 3 Aplicación didáctica de NooJ	
Cómo trabajar con NooJ en la enseñanza de español	25
Clase 1	26
Clase 2	35
Clase 3	47
Conclusiones	58
Bibliografía	59
Anexo I	
Textos periodísticos	61
Anexo II	
Textos de jóvenes	67

PRÓLOGO

El libro que ahora tienes en las manos, amigo lector, escrito por Andrea Rodrigo y Rodolfo Bonino es el resultado de varios años de fructífero trabajo y se ha nutrido con los aportes de un equipo de investigación que ha dirigido la Dra. Rodrigo, en torno a la plataforma de ingeniería lingüística NooJ, creada por el profesor Max Silberztein (Université de Franche-Comté).

En este volumen, se hallan varias obras condensadas en una sola. Se trata, por una parte, de un excelente manual del usuario de NooJ, en español y en inglés, puesto que el libro es, en gran medida, bilingüe. Por otra parte, es también una obra de análisis computacional de la lengua española, que contiene numerosos ejemplos comentados de descripción informática de distintos aspectos ortográficos, morfológicos y sintácticos del español, en su variedad rioplatense, que tanto ha aportado y sigue aportando al patrimonio común de nuestro idioma. Constituye, asimismo, un manual de didáctica de las lenguas, ya que presenta importantes elementos, tanto teóricos como prácticos, de aplicación didáctica de NooJ a la enseñanza/aprendizaje de lenguas maternas y extranjeras, que incluyen incluso relatos de experiencias de estudiantes tras los talleres de lingüística aplicada basados en este entorno informático. Es, igualmente, un manual de análisis del discurso que, además, presenta y analiza textos auténticos representativos de una «gramática de los jóvenes», suscitados en el marco de una investigación original y específica. Por último, pero no menos importante, *Aprendo con NooJ* es también un libro de ejercicios que, basándose en objetivos explícitamente formulados, y en una progresión didáctica minuciosamente planificada, permitirá al lector adentrarse con paso seguro en el universo de la lingüística computacional y sus aplicaciones.

Es este libro, además, una obra viva, puesto que se alimenta del magisterio permanente del amplio equipo de profesores e investigadores que lo firman, así como de las experiencias compartidas a través de la página de Facebook «Aprendo con NooJ», de la serie de vídeos, disponibles en YouTube, que conforman el tutorial para el uso de NooJ en español y del módulo «Español de Argentina» para NooJ, disponible en la página web del mencionado programa informático.

No me queda sino celebrar y aplaudir la aparición de esta obra que constituirá, a buen seguro, un hito en los ámbitos de la enseñanza de la lingüística computacional y la didáctica de las lenguas en Argentina.

Xavier Blanco
Universitat Autònoma de Barcelona

Capítulo 1

PRESENTACIÓN

1-El equipo de investigación

En noviembre de 2014, un equipo de trabajo conformado por docentes investigadores de la Universidad Nacional de Rosario y del Instituto de Enseñanza Superior Nro. 28, «Olga Cossettini» presentamos un proyecto de investigación «Herramientas informáticas para la enseñanza de la lengua: una aplicación didáctica en la formación docente», que constituye la base fundamental de este libro. En 2015, dicho proyecto fue acreditado por la provincia de Santa Fe, Secretaria de Estado de Ciencia, Tecnología e Innovación [1].

Nuestra línea de trabajo, apoyada en la lingüística computacional, se enmarca en un abordaje diferente de la enseñanza de la gramática. Nuestro supuesto de base es que mediante la reflexión metalingüística, al poder comprobarse en máquina las hipótesis lingüísticas planteadas, es posible generar un conocimiento explícito de la lengua. Desde nuestra línea de trabajo, pensamos que la lingüística computacional es un recurso eficaz tanto para enseñar la gramática de una lengua como para obtener información significativa acerca del uso que hacen los estudiantes de la lengua escrita, lo que permitirá elaborar material didáctico tendiente a reforzar los conocimientos que tienen y a propiciar la incorporación de otros nuevos.

La aplicación de herramientas informáticas en la enseñanza de lengua constituye una verdadera innovación, ya que hasta el momento no se registran antecedentes, al menos en la provincia de Santa Fe. Con este trabajo intentamos contribuir a la formación de docentes de lengua mostrando algunos de los múltiples usos que se pueden hacer de estos recursos.

Consecuentemente, en primer lugar, nos proponemos dar una visión general de la lingüística computacional tratando de evitar tecnicismos con la finalidad de que sea accesible para estudiantes de la escuela secundaria y de los primeros años del ciclo superior, y para docentes no especializados en el tema. En segundo lugar, mostramos cómo se puede aplicar la plataforma NooJ para comparar textos producidos por jóvenes con textos periodísticos que corresponden a una variedad estándar de español rioplatense, lo que nos permitirá contar con elementos de diagnóstico al elaborar estrategias didácticas, partiendo del supuesto de que exista una «gramática de jóvenes». Por último, planteamos una serie de ejercicios básicos de análisis automático de textos y de creación de diccionarios y gramáticas electrónicas para el español y parcialmente para el inglés. Dichos ejercicios están enfocados sin la utilización de dichos recursos, es decir, cuando los propios estudiantes crean sus propios diccionarios y gramáticas.

Pretendemos brindar una herramienta de análisis que permita partir de un estado de situación. Es decir, cómo contestaríamos a las preguntas ¿qué producen los jóvenes?, ¿qué conocimientos gramaticales han podido construir? Entendemos que solamente desde este diagnóstico es posible proyectar materiales de enseñanza, en este caso, para la enseñanza de Español, aunque también se proporciona un ejercicio como muestra para la enseñanza de Inglés como lengua extranjera. Se pretende enriquecer el conocimiento gramatical del que se dispone partiendo de las fortalezas, es decir, de aquellas categorías lexicales o estructuras sintácticas que se encuentran presentes para contribuir a desarrollar las que se encuentran ausentes o poco desarrolladas. Por ejemplo, si las

[1] Código 2060-005-14, Resolución 117/2015, 18 de marzo 2016 a 18 de septiembre 2017. <http://www.santafe.gov.ar/index.php/web/content/download/219348/1141506/file/Resultados%20Convocatoria%202014%20%20Programas%20de%20Promoci%C3%B3n%20de%20las%20Actividades%20Cient%C3%ADfico-Tecnol%C3%B3gicas%20y%20de%20Innovaci%C3%B3n.pdf>

categorías o las estructuras nominales y verbales son las más desarrolladas cómo lograr que a partir de allí puedan darse las demás categorías o estructuras.

Partimos de algunas apreciaciones en torno al sintagma verbal núcleo en francés, expuesto por Gabriel Bès (1999), como marco teórico que da sustento a este libro. Al considerar las categorías morfosintácticas que forman parte de los sintagmas núcleos, se determina su comienzo, su finalización y su núcleo con restricciones de orden muy precisas, ya que se trata de bloques inseparables. En efecto, la definición de sintagmas núcleos se da como sigue:

- * *la casa* (sintagma nominal núcleo)
- * *muy hermosa* (sintagma adjetivo núcleo)
- * *tan valientemente* (sintagma adverbial núcleo)
- * *de su interés* (sintagma preposicional núcleo)
- * *escuchaba* (sintagma verbal núcleo)

Estas primeras unidades integran unidades mayores hasta conformar finalmente el nivel oracional:

- * *la casa muy hermosa* (Sintagma Nominal)
- * *escuchaba tan valientemente* (Sintagma Verbal)
- * *eran de su interés* (Sintagma Verbal)

Se entiende que dentro del procesamiento del lenguaje natural (NLP), el análisis en sintagmas núcleos permite reducir la ambigüedad de la categorización morfosintáctica desde el momento en que es factible concatenar las expresiones internas. Las hipótesis lingüísticas en torno a las expresiones del español son validadas al ser «puestas en máquina», es decir, al cotejar diccionarios y gramáticas diseñados en NooJ con textos reales. Por último es importante señalar que para la determinación de las diferentes categorías: adjetivo, nombre, verbo, se registran las diferentes combinaciones en que se presentan dichas categorías en los textos, ya que se parte de una concepción de gramática dependiente de contexto.

2-Estructura del libro

Se exponen algunas cuestiones generales en torno al uso de NooJ enmarcándonos en la Lingüística computacional y en las investigaciones en torno al Procesamiento automático de lenguas naturales (Natural Language Processing), Primera parte, para luego entrar en ejercicios de aplicación didáctica, Segunda parte.

3-¿Qué es la lingüística computacional?

Las relaciones entre lingüística y computación son muy variadas; por este motivo, los teóricos han propuesto distintas denominaciones para las interdisciplinas que involucran ambos campos de conocimiento (lingüística computacional, lingüística informática, informática lingüística, procesamiento automático del lenguaje, etc.). Las diferentes denominaciones pretenden dar cuenta de los distintos objetos de estudio que se pueden abordar interdisciplinariamente y de los objetivos que se persiguen. Dado que se trata de un campo de estudio muy amplio, versátil y dinámico las propuestas son variadas y, en muchos casos, contradictorias. Por lo tanto, preferimos explicitar nuestros objetivos a debatir acerca de cuál es el nombre que resulta más adecuado.

Desde nuestra perspectiva, la lingüística computacional es un área interdisciplinar que postula formalizaciones del lenguaje humano procesables mediante programas informáticos, sin presuponer que los procesos automáticos sean análogos a los procesos mentales de los hablantes.

Aunque nuestras formalizaciones se desarrollan en función de los recursos que ofrece un programa en particular, la preexistencia del programa no es una condición indispensable para la investigación; de hecho, el programador trabaja a partir de determinadas hipótesis sobre la naturaleza formal del lenguaje. En síntesis, la relación entre lingüística e informática no es unidireccional en ninguno de los dos sentidos (de la informática a la lingüística ni viceversa) sino dinámica: los programas se

desarrollan a partir de determinadas hipótesis lingüísticas, las funcionalidades del programa condicionan el tipo de gramática que se crea y, a su vez, los programas se enriquecen a medida que las gramáticas plantean nuevas exigencias para una mejor formalización.

4-La plataforma NooJ

A pesar de que en el proyecto original se procuraba partir de herramientas informáticas, (concretamente XFST y NooJ), muy pronto decidimos dejar de lado la herramienta XFST, ya que sólo contábamos con una versión de prueba y con escaso margen para introducir textos y nos centramos en la herramienta NooJ, devenida en plataforma. Varias ventajas nos llevaron a tomar esta decisión:

- ♦ que está disponible en línea y sin ningún costo, proponiéndose solamente si se lo utiliza para fines académicos,
- ♦ que no se trata de una caja negra, sino de una caja vacía en que es posible bien crear textos, diccionarios y gramáticas o bien valerse de los recursos disponibles, en este caso, los referidos al Módulo Español Argentina (hoy en pleno desarrollo), por lo tanto, es totalmente interactiva,
- ♦ que está proyectada para diversas lenguas, lo que lleva a considerar la multiplicidad desde una base común,
- ♦ que es posible contar con el asesoramiento del creador de NooJ, el Dr. Max Silberztein, quien dirige el proyecto NooJ¹. Nuestro proyecto, por lo tanto, se articula con dicha propuesta,
- ♦ que en virtud de poder contar con el Módulo Sp Barcelona gracias a la gentileza del Dr. Xavier Blanco, es factible completar y mejorar los diccionarios y gramáticas del Módulo Sp. Argentina.

Particularmente, la utilización de la herramienta NooJ, nos permite estar a la vanguardia de las investigaciones en lingüística computacional, entendiendo que se trata de una herramienta que permite formalizar distintos niveles lingüísticos: ortografía, morfología y sintaxis.

Su implementación en la formación docente contribuye a que los alumnos puedan ser conscientes de la necesidad de formalización lingüística y de las ventajas que supone.

5-¿Qué hace NooJ?

La función básica de NooJ, que lo hace apto para el tratamiento de muchísimas lenguas naturales, es reconocer caracteres del sistema Unicode e identificarlos como letras, cifras, signos de puntuación y espacios en blanco. Cualquier cadena de caracteres correspondientes a letras que se encuentren entre caracteres con otros valores (espacios en blanco, signos de puntuación, números) se reconoce como forma de palabra, independientemente de cuáles sean esos caracteres. Esta función le permite aportar información estadística aún sin contar con ninguna información lingüística específica:



Figura 1 Análisis de secuencias inexistentes

La figura 1 muestra el análisis de una serie de cadenas de caracteres que no coinciden con palabras de ninguna lengua. En el recuadro superior derecho de la figura, podemos observar que en el texto se

1. ELLIADD Laboratory, Université Franche-Comté (Francia)

reconocen 2 unidades textuales (*Text Units*), 36 unidades lingüísticas (*tokens*) que incluyen 15 formas de palabras (*word forms*), 16 cifras (*digits*) y 5 signos de puntuación (*delimiters*). Como en este caso el texto es muy breve, es posible reproducir manualmente el análisis automático. Así vemos que:

- los cambios de línea (*return* o *enter*) separan unidades textuales,
- las letras se cuentan como *tokens* cuando se encuentran entre caracteres no alfabéticos (cifras, comas, guiones), un *token* alfabético puede estar constituido por una o más letras,
- las cifras y los signos de puntuación se cuentan individualmente como *tokens*, estén o no separados por otros caracteres,
- los espacios en blanco no se cuentan como *tokens*.

En el recuadro superior del centro de la Figura 2, se ve una ventana que permite obtener otros datos estadísticos del texto:

Text has 31 characters.			
Freq	Char	Type	Unicode
16		Blk	0020
7	a	Let	0061
6	s	Let	0073
6	d	Let	0064
6	f	Let	0066
3	,	Del	002C
3	v	Let	0076
3	3	Dig	0033
3	2	Dig	0032
2	7	Dig	0037
2	m	Let	006D
2	e	Let	0065
2	q	Let	0071
2	x	Let	0078
2	n	Let	006E
2	4	Dig	0034
2	z	Let	007A
2	-	Del	002D
2	5	Dig	0035
2	w	Let	0077
2	r	Let	0072
2	6	Dig	0036
2	c	Let	0063
1	j	Let	006A
1	1	Dig	0031
1	l	Let	006C
1	8	Dig	0038
1	b	Let	0062

Figura 2 Characters / Unicode

Characters informa la frecuencia, el tipo y el código correspondiente a cada carácter

8 different tokens	
Freq	Tokens
6	asdf
2	qwer
2	zxcv
1	ñlkj
1	n
1	m
1	mnbv
1	a

Figura 3 Tokens informa la frecuencia de unidades

4 reoccurring digrams / 8	
Freq	Digrams
2	asdf qwer
2	qwer zxcv
2	zxcv asdf
3	asdf asdf

Figura 4 Digrams reconoce secuencias de dos tokens que se reiteran en el texto

Además de estas funciones básicas, NooJ ejecuta otras que permiten al usuario incorporar diccionarios y gramáticas de diversas lenguas naturales. Para esto se tienen en cuenta propiedades universales del lenguaje o comunes a grandes grupos de lenguas:

- a pesar de las dificultades para definir las teóricamente, en todas las lenguas existen palabras,
- en muchísimas lenguas la representación gráfica de las palabras se hace mediante letras (de ahí que una de las funciones básicas de NooJ sea reconocer letras o secuencias de letras como palabras),
- las palabras se pueden agrupar por sus propiedades léxicas, morfológicas y sintácticas,
- en muchas lenguas las palabras tienen variaciones flexivas,
- muchas lenguas utilizan afijos derivativos para formar nuevas palabras a partir de las ya existentes,
- las variaciones morfológicas, muchas veces implican variaciones de la raíz (verbos irregulares, raíces supletivas, cambios gráficos y acentuales, etc.),
- todas las lenguas naturales son recursivas, es decir, las unidades se integran en niveles superiores (los morfemas forman palabras, las palabras forman sintagmas y los sintagmas se integran en sintagmas más complejos).

En síntesis, NooJ no «conoce» ninguna lengua, pero sí propiedades generales del lenguaje natural que permiten que los usuarios le «enseñen» distintas lenguas. En este sentido, podemos hablar de dos niveles de programación: la programación general, que son las funciones disponibles para

desarrollar diccionarios y gramáticas, y la programación que llevan a cabo los usuarios aportando diccionarios y gramáticas de lenguas específicas.

6- Diccionarios NooJ

NooJ reconoce como palabra cualquier letra o secuencias de letras delimitadas por caracteres no alfabéticos. Si queremos que identifique palabras reales de una lengua específica, debemos crear diccionarios que incluyan mínimamente información categorial. Cada usuario puede definir en un archivo –sin ninguna restricción– qué categorías resultan más adecuadas para sus necesidades: es posible utilizar tanto etiquetas convencionales (nombre, adjetivo, verbo, etc., con cualquier subclasificación que se estime conveniente) como etiquetas creadas para fines específicos (categoría1, categoría2, categoría3, etc.).

Un diccionario que tenga solo información categorial será capaz de reconocer únicamente las entradas declaradas, por ejemplo, si tiene registrada la palabra *libro*, no reconocerá *libros*; para que lo haga, será necesario bien incluir todas las variantes morfológicas de una palabra (lo cual es absolutamente antieconómico) o bien crear gramáticas con paradigmas flexivos. En este caso, además de la etiqueta categorial, en el diccionario se deberá declarar cuál es el paradigma flexivo de cada entrada del diccionario. En lo que respecta a las entradas o lemas, también queda a elección del usuario la alternativa utilizada por los diccionarios convencionales de ingresar los infinitivos de los verbos y la forma del masculino singular de los nombres y los adjetivos, o usar cualquier otra variante morfológica.

Además de información categorial y flexiva, se puede agregar todo tipo de información que sea relevante para calcular las propiedades gramaticales de una entrada, por ejemplo, si un nombre es animado o inanimado, si una palabra pertenece a alguna variedad lingüística específica, etc.

7- Gramáticas NooJ

En su versión actual, NooJ permite crear distintos tipos de gramáticas:

1. gramáticas morfológicas,
2. gramáticas de morfología productiva y
3. gramáticas sintácticas.

Las primeras, además de utilizarse para declarar los paradigmas flexivos de una palabra, permiten formalizar la morfología derivativa. En ambos casos se deben especificar todas las transformaciones que sufre una palabra para cambiar un afixo, por ejemplo, el cambio de sufijo de infinitivo por el sufijo de tercera persona del presente del subjuntivo del verbo *tragar* implica el borrado de los dos caracteres correspondientes a la vocal temática y el infinitivo (*ar*) y la adición de la letra *u* que no tiene motivación morfológica sino ortográfica; la incorporación del morfema apreciativo *-ito* en la palabra *chico*, además de la supresión de la *o* requiere el cambio de *c* por *qu*. De estos ejemplos se puede inferir que son gramáticas muy precisas, pero que requieren un trabajo muy meticuloso. El uso de estas gramáticas en los diccionarios creados manualmente permite que NooJ genere de forma automática diccionarios con todas las variaciones flexivas y derivativas previstas en la gramática, estos diccionarios son los que se aplican en el análisis de los textos.

Las gramáticas productivas permiten declarar, por ejemplo, que cualquier secuencia de letras seguida de la secuencia *-ito* es un diminutivo o que cualquier secuencia de letras seguidas de *-aba*, es un verbo en primera o tercera persona del singular del pretérito imperfecto del indicativo de la primera conjugación. Este tipo de gramáticas se aplica directamente en el análisis del texto y complementa el análisis producido por las anteriores.

Las gramáticas sintácticas son útiles para formalizar relaciones entre palabras que constituyen sintagmas. Además son recursivas, lo que implica que los sintagmas formalizados pueden ser constituyentes de otros que los incluyen. Pueden formalizarse tanto a partir de palabras como de etiquetas asignadas por el diccionario. Si optamos por la primera opción, obtendremos una gramática muy rudimentaria cuya única posibilidad será reconocer las secuencias de palabras especificadas; en el segundo caso, se pueden obtener gramáticas más compleja, capaces de reconocer secuencias de categorías.

Capítulo 2

EL CORPUS DE TRABAJO

1- El texto

Como punto inicial, se toma un corpus conformado por textos periodísticos para poder partir de un uso «cuidado» de la lengua, pero no de un registro literario. Se trata de textos de diferentes periódicos de la Argentina, la variedad de español es rioplatense. Se comprende que los textos presentan algunas desviaciones ortográficas, pero que se contemplan como posibles dentro de los textos que circulan. Reflejan un estado «real» de la lengua y no necesariamente el que encontramos en las gramáticas.

El texto completo se expone en el Anexo final. A continuación, se muestra solamente un fragmento.

Dorrego era amigo de «los de abajo», tenía arraigo popular y buenas relaciones con las provincias. Esto lo condenó. También las famosas cartas de Juan Cruz Varela y Salvador María Del Carril. ¿Por qué fueron famosas estas cartas, por qué trascendieron cuando tantas cosas se ocultan? Hace poco, en este diario, el sacerdote Domingo Bresci escribió sobre los archivos de la Iglesia durante los años de la dictadura. Pidió que se revelaran, que salieran a luz. Difícil, están bien guardados. La Iglesia participa del poder de la Argentina y fue cómplice de la dictadura. ¿Por qué ese poder –que escribe la historia a su antojo, el de los vencedores– no salvó a Del Carril y Varela, que eran figuras de la «civilización»? Para salvar al valiente y glorioso general Lavalle, que fue el «otro» San Martín, el del frente interno, el de las guerras civiles, el soldado de la ilustrada Buenos Aires. El que desenvainó su espada para derramar sangre de hermanos. La de Dorrego, en primer lugar y tal como se lo pidieron sus amigos unitarios. Juan Cruz Varela, consciente de la crueldad que le pedía a Lavalle, escribió: «Cartas como éstas, se rompen». Pero no se rompieron. El fusilamiento de Dorrego fue un crimen imperdonable, había que entregar a alguien para salvar al evidente ejecutor, Lavalle. Se decidió contar la historia publicando las cartas de un poeta extraviado y de un áspero, duro unitario.

El análisis del corpus conformado por estos textos permitirá establecer un parámetro sobre cómo debe ser un texto bien construido en español, en una variedad estándar de español rioplatense.

2- Análisis del corpus de textos periodísticos

La aplicación NooJ permite, entre otras cosas, formalizar descripciones de lenguajes naturales para analizar aspectos de la morfología flexiva y derivativa, la terminología, la sintaxis y la semántica en corpus lingüísticos. En este punto intentamos analizar un corpus de 3883 palabras creado a partir de artículos de distintos periódicos argentinos con la función Linguistic Analysis. Se pretende observar cuáles son las categorías lexicales que aparecen con mayor frecuencia.

Al realizar el análisis lingüístico del texto con NooJ, puede determinarse la presencia de 1479 *tokens*, según la información que aparece en el recuadro superior. Es que por medio de esta función, la aplicación proporciona información acerca de la cantidad de caracteres, *tokens* y palabras conocidas o desconocidas, entre otros datos, según se observa en:

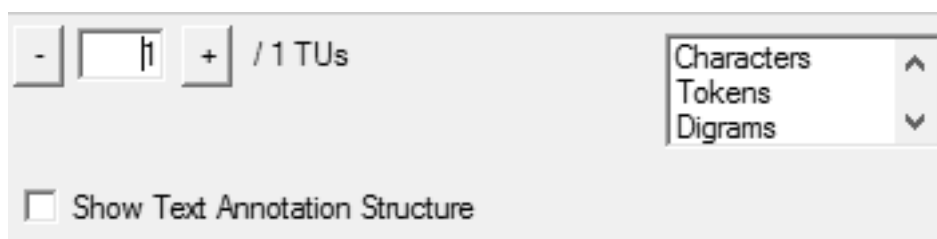


Figura 5 Información que brinda la plataforma NooJ en relación a un texto: C

Los *tokens* a diferencia de las palabras discriminan las palabras repetidas, por lo tanto su número es menor.

Frec	Tokens
1	debia
2	ahorro
1	matado
1	debia
3	hay
1	aplicar
1	imaginando
2	testigo
2	afirmo
1	discusar
1	lugares
1	atractiva
1	diarum
1	Movimiento
1	puerta
1	perceptible
2	han
1	indignas
4	relación
1	teléfono
1	título
1	aljaria
1	metropolitana
2	sino
1	frenar
1	Cámara
1	mil
1	asesinado
1	teléfono
1	podrían
1	gona
1	provocar

Figura 6 Tokens

Se procede a analizar de un modo muy general la frecuencia de palabras en el texto. Siempre con el texto abierto, y luego de clicar en TEXT Linguistic Analysis para que se apliquen los recursos del español, se solicita que se exporten las anotaciones como XML.

Dorrego era amigo de "los de abajo", más que de los de arriba. También las famosas cartas de Juan Cruz Varela y Salvador María Del Carril. ¿Por qué fueron famosas estas cartas, por qué trascendieron cuando tantas cosas se ocultan? Hace poco, en este diario, el sacerdote Domingo Bireci escribió sobre los archivos de la Iglesia durante los años de la dictadura. Pidió que se revelaran, que salieran a luz. Dónde, están bien guardados. La Iglesia participa del poder de la Argentina y fue cómplice de la dictadura. ¿Por qué ese poder —que escribe la historia a su antojo, el de los vencedores— no salvó a Del Carril y Varela, que eran figuras de la "civilización"? Para salvar al valiente y glorioso general Lavalle, que fue el "otro" San Martín, el del frente interno, el de las guerras civiles, el soldado de la Ilustrada Buenos Aires. El que desenfundó su espada para detramar sangre de hermanos. La de Dorrego, en primer lugar y tal como se lo pidieron sus amigos unitarios. Juan Cruz Varela, consciente de la crueldad que le pedía a Lavalle, escribió "Cartas como éstas, se rompen". Pero no se rompieron. El asesinato de Dorrego fue un crimen imperdonable, había que entregar a alguien para salvar al evidente ejecutor, Lavalle. Se decidió costar la historia publicando las cartas de un poeta estraviado y de un áspete, duro unitario.

Rosas usaba la muerte de Dorrego para animar su primer mandato. Anunciar para siempre esa muerte y establecer las leyes. Él, entonces, el Restaurador de las Leyes. Será, también, un estanción keynesiano. Se lleva bien con los gauchos, los indios y los negros. Tanto que los vende gauchos para poder comprarlos. Será, para San Martín, el más gauchito de los gauchos, el más bárbaro de los bárbaros. Con sólo mostrar el punto de equilibrio luego de la pampa sobre el suelo de maracaibo.

Después de huir a Dorrego, Lavalle usaba a los de sus caracoles a rastrear la frontera sur. Son Federico Rauch y Ramón Estomba. Entre gauchos a volver cartas de los señores. Los atan a los castillos y se forman hacer fuego. Estomba es coligacionista. Rauch es atrapado por los generales del jefe indio Arbolito. Le cortan la cabeza y la arrojan sobre Buenos Aires.

Rosas se presenta como "el vencedor de la anarquía". Los jóvenes del Salto Litrico —Alberdi, sobre todo— le ven como la etapa del Orden, que sigue a Mayo, la de la Espada, y concluye con la Intendencia, que representará Alberdi y sus amigos.

Rosas anuncia un segundo gobierno como el resurgido de Fernando Quiroga. Antes, en 1814, en unión, Encarnación Encarna, ama sus revoluciones, la de los Restauradores, para imponer a Rosas como jefe del país. Encarnación muere joven, como Eva Perón, quien, como ella, verá parte del movimiento del 17 de octubre, que gana el gobierno para Perón. La unión brinda que lucha por el poder de su hombre porfirista de Encarnación a Eva. A nuestra historia le gustan las simonías.

Con Rosas Regan al gobierno los sobrevivientes de la pampa húmeda. La burguesía mercantil de Buenos Aires se acepta porque quiere orden y su anarquía, que es perjudicial para los negocios. Las Provincias le ven como un federal, una representación de Dorrego. Rosas es el representante del poder público y la representación de las relaciones exteriores.

Alberdi (o, en menor grado, los jóvenes rosalinistas, Estaberría, Sastre, Gutiérrez etc) le ofrece su apoyo. Hay aquí una importante perennidad que tiene una necesidad que de estudio: la relación de los intelectuales con el poder. Moreno y Castelli temen al gobierno y ven sus propios intereses porque usan cultos. Luego, los señores de Lavalle-Agüero, Varela, Del Carril, Después Alberdi y Rosas, lo que no fue. Rosas creó en Perón de Aragón. Al fin del siglo XIX, la generación del 89 y Rosas. En el siglo XX, Lugones y Urbino, Forja y Perón. "Sin" y la Libertadiva. Hasta los movimientos y el tercer Perón. Todos estos proyectos se fueron. Durante los días que corren temen a Alejandro Rosalino y Macri, que vuelvan a ser una perennidad de la dialéctica entre Alberdi y Rosas. ¿Qué dijo que la historia no progresa?

Insisten provincias en que sigue sin encarrilarse plan ferroviario.

RECLAMAN OBRAS AL GOBIERNO NACIONAL Y QUE RAMALES NO CIERREN

Tras conocerse Presupuesto 2018, en el sur piden por el Tren del Valle y el Tren Patagónico. Chaco reclama suspensión del decreto de Macri que habilita a levantar vías. Tren a Vaca Muerta, sin retrasar.

Ampliación. En la Patagonia piden que se extienda el servicio en una serie de trayectos, y que se renueven vías en tramos donde las instalaciones no están en buenas condiciones.

La decisión del Ferrocarril Nacional y diversos asociados de la Casa Encana piden en alerta a provincias y municipios, que busquen obras para mejorar el transporte ferroviario y evitar ramales que podrían

Figura 7 Aplicar Linguistic Analysis a un texto

Un vez allí cliqueamos en la segunda opción² y allí escribimos: <N>, <V>, <ADJ>, <ADV>



Figura 8 Exportar anotaciones como XML

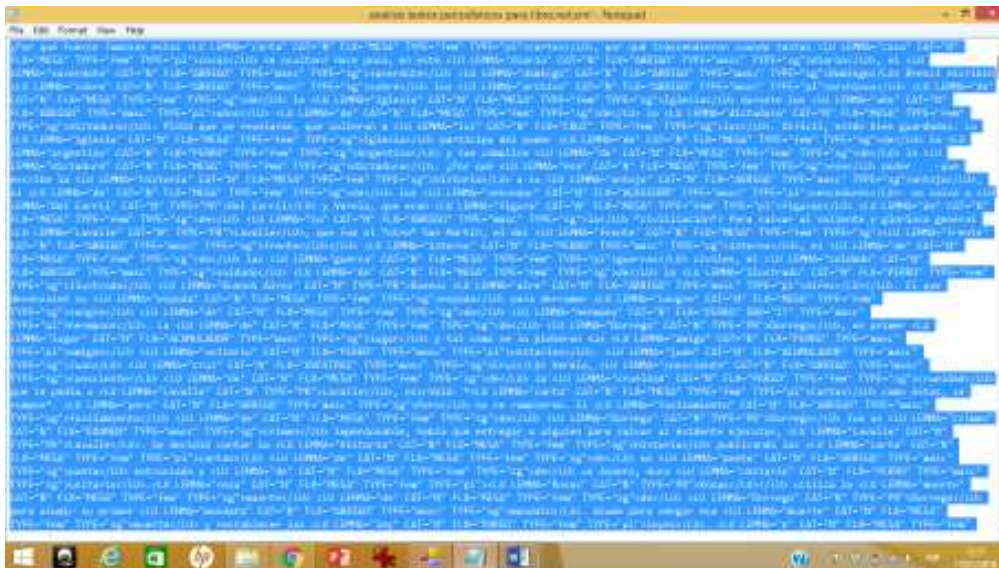


Figura 9 Texto anotado con formato XML

Ese texto anotado (en formato XML) constituye el insumo para un nuevo análisis. Entonces, creamos un nuevo texto, copiando y pegándolo. Luego aplicando nuevamente Linguistic Analysis y la función Locate es posible identificar en ese texto todas las ocurrencias de las categorías léxicas que necesitamos. Se las escribe en el cuadro de diálogo con mayúsculas, pero esta vez sin paréntesis angulares, en este caso N, V, ADJ, ADV. De este modo, se obtiene el siguiente resultado, con la función Statistical Analysis:

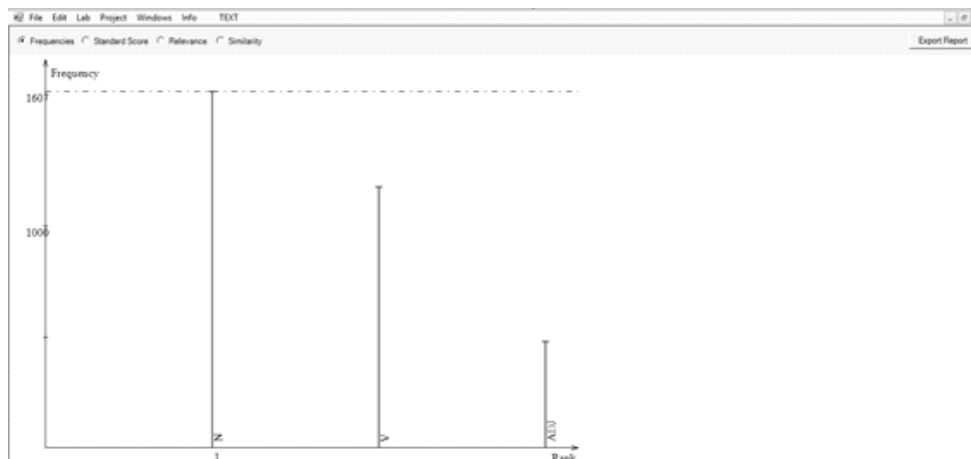


Figura 10 Frecuencia de Nombres, Verbos y Adjetivos

² Esto se explicará con más detenimiento en el punto correspondiente a la Aplicación didáctica.

Como una primera conclusión, se observa que sobre un total de 3883 palabras, existen 1518 nombres. En segundo lugar, se encuentran los verbos, 878 y finalmente los adjetivos, 467. La frecuencia de adverbios parece mucho menos significativa, apenas 150. Esto nos da una idea de hacia dónde es preciso apuntar las estrategias de enseñanza, es preciso reforzar la categoría de los adverbios y la categoría de adjetivos, puesto que demuestra una pobreza del input en este aspecto.

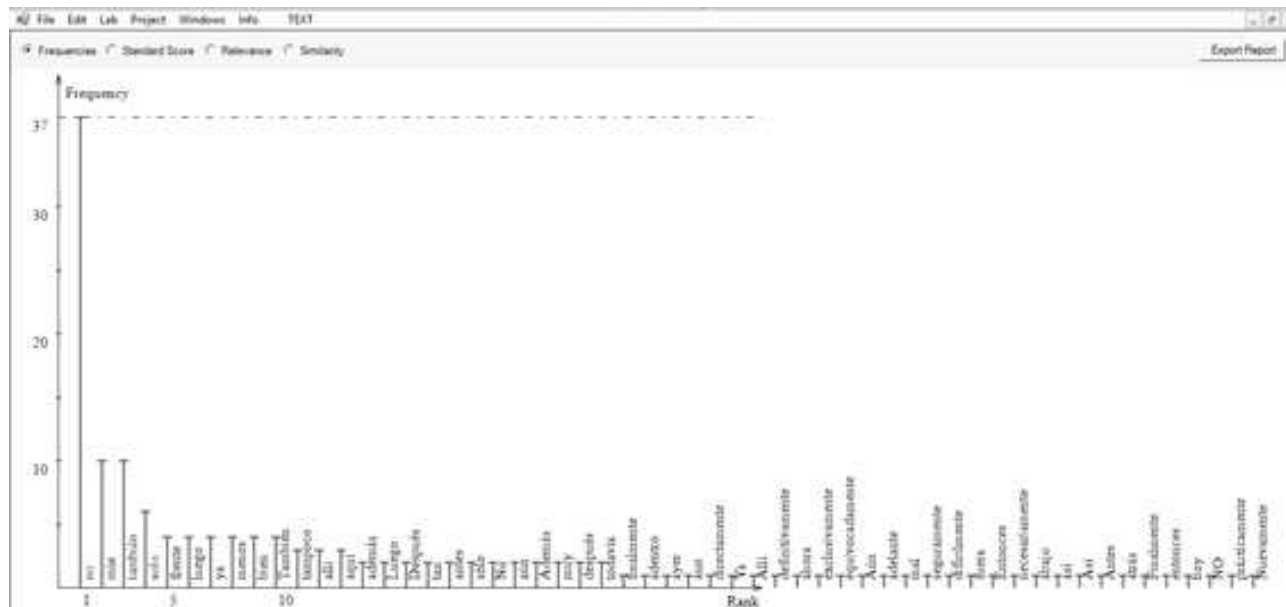


Figura 11 Frecuencia de adverbios en el texto periodístico

Este es el resultado general expresado como gráfico:

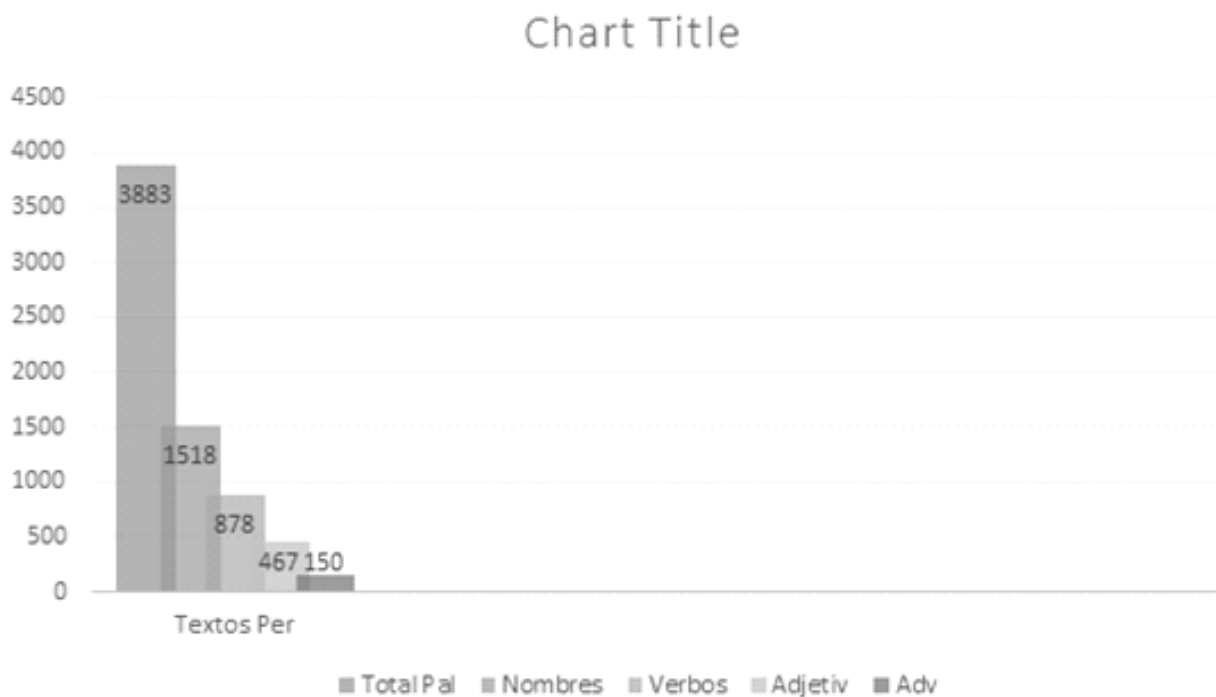


Figura 12 Gráfico de frecuencias de categorías en el texto periodístico

3- Contraste con las producciones de jóvenes, fortalezas y debilidades

Nuestro objetivo es trabajar con un corpus de textos escritos de manera digital. Para poder contar con este corpus creamos una página de Facebook, la que tratamos de promocionar entre amigos contando con la ayuda de colegas y alumnos: Aprendo con NooJ³. En esta página se insertó un documento de Google en el que luego de dar su consentimiento informado y completar algunos pocos datos, se realizó una pregunta abierta, *¿Qué pensás de la educación?* En función de las respuestas recibidas y de acuerdo a las siguientes variables:

Edad, 18 a 25 años

Lengua, español,

se llegó a construir un texto.

En este corpus se observan algunas palabras que no están escritas según la norma ortográfica lo que puede deberse a cuestiones de tipeo o al corrector que utilizan algunos programas incluso habría que considerar que se utilice el corrector del celular. A continuación, se expone un fragmento, el corpus completo se encuentra en el Anexo:

la educacion es fundamental para el desarrollo de la persona

Para mí la educación es un derecho individual y un deber del estado hacerla llegar a toda su población. La educación en Argentina la considero muy buena, ya que le da la posibilidad a personas con bajos recursos interiorizarse, tener una opinión, una postura y una visión sobre el entorno que nos rodea. No debemos desperdiciar la oportunidad que nos da nuestro país de poder estudiar gratuitamente. Me parece repugnante que vengan extranjeros a usurpar cupos en universidades Argentinas, si su país no les brinda una educación pública gratuita por qué nosotros si?

Datos que brinda la página Aprendo con NooJ

Sobre la edad de los jóvenes que dejaron sus respuestas en el documento de Google:

Edad (entre 18 y 25 años)

62 respuestas

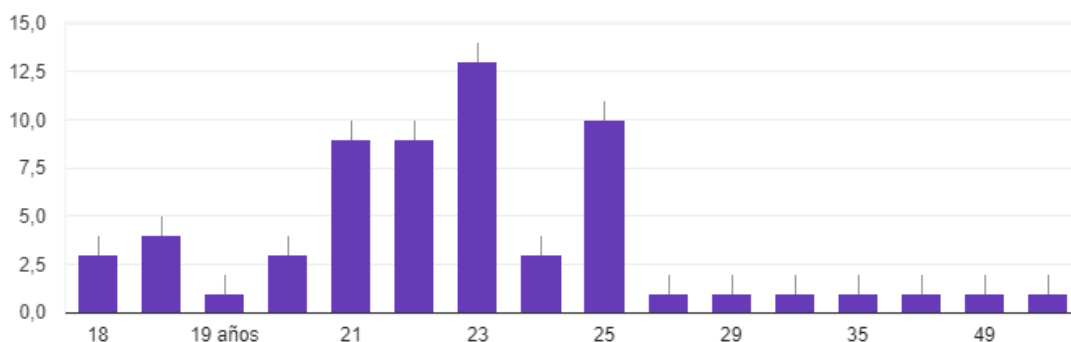


Figura 13 Promedio de edades de las respuestas de los jóvenes

Se observa que el mayor porcentaje se sitúa entre los 21 y 25 años.

3 Enlace a la página Aprendo con NooJ: https://docs.google.com/forms/d/1maxe0DMzNcEPzPYL7ARw_sLes25xN6V9Z1gYH6EyXM/edit

4- Análisis de los textos producidos por los jóvenes

Del análisis de los textos producidos por los jóvenes, en función de la frecuencia de palabras se observa que predominan Verbos y Nombres en contraposición a los Adverbios:

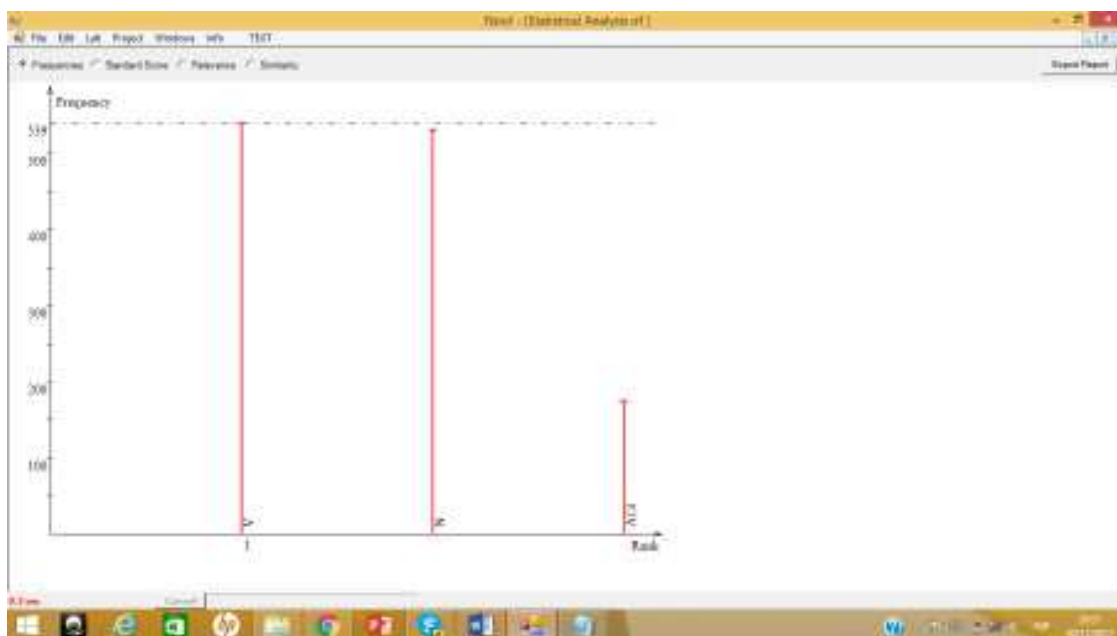


Figura 14 Frecuencia de palabras en textos de los jóvenes

Este análisis surge de la aplicación de un mismo diccionario al análisis de los textos periodísticos como a los textos de los jóvenes. Sin embargo, nuestra intención aquí es no considerar como error las palabras cuya ortografía no coincide con la «norma», por lo tanto es preciso ingresar algunas palabras nuevas las que van a aparecer con una etiqueta [+jov]. Se trata de las siguientes palabras:

educacion/N+jov
interiorizarce/V+jov
trasmite/V+jov
autocriticas/ADJ+jov
concientes/ADJ+jov
esta/V+jov

Lo que se registra por lo común son errores de concordancia que tienen más que ver con aspectos sintácticos, ejemplo:

Creo que debería haber más exigencia hacia los educando y, también al docente.

También se observa el desconocimiento de la interrogativa indirecta:

Gracias a la educación podemos decidir que hacer con nuestras vidas y expresarnos de distintas maneras.

5- Comparación entre el corpus de textos periodísticos y el corpus de textos producidos por los jóvenes

En ambos casos se trata de textos de opinión, es decir, que la base de comparación se apoya en el tipo textual.

Pueden observarse tanto similitudes como diferencias:

- la frecuencia de palabras es similar
- existen mayores desviaciones ortográficas en el texto producido por los jóvenes

- la concordancia sintáctica se revela como un aspecto que es preciso reforzar
Si bien en la Segunda parte no se abordarán estas cuestiones quedan pendientes para ser retrabajadas en futuras investigaciones del equipo.

6- Opiniones de los alumnos

A continuación se exponen algunas de las impresiones de alumnos que han participado de diferentes formaciones a cargo de docentes del equipo y como parte del proyecto de investigación que estamos realizando. Creemos que es útil compartirlas aquí.

♦ Experiencia en el «Taller de lingüística aplicada: ‘La herramienta NooJ y su aplicación en el estudio de la morfología del español» (2017)

«Mi acercamiento al curso y a la herramienta NooJ estuvo motivado principalmente por la curiosidad. Suelo visitar la página web del Instituto «Olga Cossettini» en búsqueda de capacitaciones y, en una de esas visitas, encontré una invitación que particularmente me pareció novedosa. Nunca había escuchado hablar de la herramienta NooJ previo a mi acercamiento al curso. No obstante, la propuesta sonaba prometedora, especialmente la posibilidad de crear nuestros propios diccionarios. Me anoté sin saber exactamente con qué me encontraría. Siendo traductora novel, mi mayor expectativa era que la herramienta me ayudara al momento de traducir. Como anticipo, cabe destacar que esto no sucedió, aunque sí me llevé otras gratas sorpresas de la experiencia. Por temor a que el resto de las personas asistentes conocieran la herramienta y yo fuera la única ignorante al respecto (lo cual no sucedió), luego de anotarme visité la página web de NooJ para esclarecer mis ideas iniciales. Mi primera impresión fue extraña. Encontré una página un tanto caótica y con vocabulario informático que yo no manejaba. Puse en duda si había tomado la decisión correcta al inscribirme en el curso. Al día siguiente, luego de mi pequeño pánico inicial, le di otra oportunidad a la herramienta. Me tomé mi tiempo para navegar por las diversas pestañas de la página y encontré algunos tutoriales y explicaciones. Observé algunos tutoriales en español (tiempo después iba a descubrir que la mujer de los videos era una de las docentes del taller, Andrea Rodrigo) y me relajé. No estaba equivocada al inscribirme. Finalmente llegó el día de asistir al primer encuentro del «Taller de lingüística aplicada: ‘La herramienta NooJ y su aplicación en el estudio de la morfología del español’». El ambiente amigable que encontré allí distó mucho de la solemnidad que el título del curso presentaba. Vi pocas caras conocidas pero muchas sonrisas. Comenzó una ronda de mates mientras descargábamos la herramienta y ahí pude descubrir que todas las personas presentes teníamos el mismo nivel de inexperiencia y de curiosidad. Esto fue un alivio; me permitió sentirme cómoda con el grupo y acercarme a la herramienta de otro modo. Resalto la importancia de utilizar NooJ en actividades grupales y no en aprendizaje solitario. El profesor Rodolfo Bonino es una persona con mucho conocimiento informático y lingüístico. Sin embargo, esto no significa que se haya presentado ante el curso como una figura de autoridad o de manera lejana. Tanto él como Paula Alonso y Andrea Rodrigo explicaron los conceptos todas las veces que les fueron requeridas y acompañaron a cada cual con su proceso de descubrimiento de la herramienta».

«Con respecto a la teoría dada en el curso, fueron muy esclarecedoras las nociones generales sobre el tratamiento informático de las lenguas naturales. Además, se brindó una pequeña historización que nos ayudó a situarnos en la actualidad del fenómeno. A su vez, recibimos explicaciones básicas sobre la lingüística aplicada y las diversas corrientes de pensamiento sobre las herramientas computacionales. Con respecto a la herramienta NooJ en sí, tuve sentimientos encontrados. Al principio, debo admitir que me sentí un tanto decepcionada ya que me di cuenta de que no iba a cumplir mis expectativas iniciales de poder agilizar mis traducciones. De hecho, luego de la primera clase, regresé a mi casa pensando qué tanto podía aportarme esta herramienta a mi trabajo. Como traductora novel, tengo algunos conocimientos básicos sobre herramientas informáticas y ninguna de ellas funciona como recurso para aprender gramática. Al principio del curso, debido a nuestro

desconocimiento, los ejercicios eran más bien simples y repetitivos, a fin de comprender el funcionamiento de la herramienta. A pesar de su sencillez, las actividades presentaban un desafío. En esos momentos iniciales de pequeñas frustraciones frente a las dificultades, surgían las ventajas de trabajar en modalidad de taller. Cuando alguien tenía un problema, compartía lo ocurrido y podíamos ayudarnos a mejorar de manera grupal, aunque siempre con la guía del docente. En algunas oportunidades, aprovechábamos los errores de otras personas para no cometerlos y, en otras, celebrábamos los aciertos y descubrimientos ajenos para mejorar. Me llevó cerca de dos clases descubrir qué tenía NooJ para ofrecerme. A diferencia de otras herramientas, NooJ permite controlar los procesos intermedios que se realizan y no solo la entrada y salida de información. Esto posibilita la colocación de formalizaciones particulares del lenguaje, es decir, le proporciona a cada cual la libertad de formular sus propias gramáticas y diccionarios y de crear categorías propias para organizar las unidades léxicas».

«Una cuestión interesante que me permitió descubrir la herramienta NooJ es la noción de recursividad. Durante mi formación académica en el Instituto «Olga Cossettini», he estudiado la gramática generativa a partir de los textos de Noam Chomsky. En consecuencia, he aprendido las propiedades de dicha gramática, entre ellas la recursividad, entendida como la capacidad de una gramática de generar una infinita cantidad de construcciones sintácticas a partir de un número limitado de reglas y unidades abstractas. Esta noción siempre me había parecido llamativa pero me era difícil observarla en ejemplos concretos o poderla construir con ejemplos propios. Cuando comenzamos a cargar datos en la herramienta NooJ, pude entender este concepto mediante casos claros. Por ejemplo, en nuestras gramáticas agregamos explicaciones generales bajo el nombre de 'GÉNERO 1' o 'NÚMERO 3' y luego etiquetamos a las palabras a partir de esas categorías. A raíz de esto me percaté de que existe un número limitado de patrones para crear el género o el número y que todas las palabras se engloban en estos grupos. Evidentemente, a partir de un conjunto de elementos finitos se pueden generar e interpretar un número infinito de construcciones»

«Punto a destacar sobre mi aprendizaje con la herramienta NooJ refiere a las irregularidades sistemáticas. Al tener que recurrir a diversos comandos para explicarle al sistema cómo crear nuevas palabras a partir de una base, me vi obligada a re-explicarme a mí misma cómo funciona la morfología del español. Tal vez suena extraño pero, como hablante nativa, normalmente no me cuestionaba los tipos de irregularidades de la lengua. A partir de los ejercicios realizados en el taller, pude darme cuenta de los diversos tipos de irregularidades verbales y distinguir entre cambios gráficos, diptongaciones, combinación de irregularidades, entre otras. Por ejemplo, usamos un comando especial («<U>») para la diptongación, como en el caso de «acertar '! acierto». Esta posibilidad que brinda la herramienta me parece sumamente útil para aquellas personas que ejercen la docencia de idiomas extranjeros o de español como segunda lengua ya que, al tener que analizar con tanto detalle una lengua, se pueden entender los errores comunes que cometen quienes la están aprendiendo».

♦ **Experiencia en el Seminario orientado en Lingüística y Lenguas «Aportes de la lingüística informática en la enseñanza de español como L1 y L2» a cargo de las profesoras Silvia Reyes y Andrea Rodrigo (2017)**

«El punto de partida del trabajo fue la lectura del Tutorial en Español de NooJ, diseñado por Rodolfo Bonino y Andrea Rodrigo, en el cual se ofrecen al lector y aprendiz una serie de indicaciones y ejercicios para introducirlo en el funcionamiento de NooJ. El Tutorial constituye la introducción básica indispensable a la herramienta NooJ para aquellos inexpertos en programas informáticos de procesamiento de lenguaje natural (PLN). Para el español, puede encontrarse no solo el tutorial escrito, el cual se obtiene una vez descargada la aplicación NooJ, de libre disponibilidad en la página www.nooj-association.org, sino también una serie de videos en el canal de YouTube. En el recorrido

presentado por Andrea Rodrigo a lo largo de los once videos que componen el tutorial, el estudiante se encuentra con la información necesaria para el uso de NooJ, partiendo de su instalación hacia actividades cada vez más complejas».

«Lo que resultó de particular interés fue la propuesta en torno a la creación de una gramática derivacional o flexiva regular, en la clase 2, y la presentación de los comandos que se pueden utilizar para la creación de dicha gramática. En el Tutorial, uno de los ejemplos que ilustra el uso de los comandos –y en especial la posibilidad de combinación de comandos– es en relación al verbo lucir: LUCIR=<B2><L>z<R>o/pte+1era+sg, produce «luzco» con la etiqueta correspondiente a sus rasgos flexivos. La observación de esta secuencia condujo a plantear la posibilidad de una manera diferente de representar las conjugaciones del paradigma verbal, apuntando a una fórmula más económica, es decir, que requiera de una cantidad mínima de comandos para expresar las diversas etiquetas de la flexión verbal».

«NooJ se destaca por ser una herramienta con numerosas actividades de uso sencillo, de acceso libre y gratuito y que es compatible con diversos sistemas operativos. Considero que se trata de un nuevo enfoque para el estudio de la lengua que opera ya no con la memorización de reglas gramaticales sino con ensayos, con pruebas y tareas de carácter lúdico, que permiten al estudiante aprender la lengua en tanto la manipula, crea textos y los analiza. Por lo tanto, estimo que interviene positivamente, y en clave renovadora, en el aprendizaje del español».

«Mi primer acercamiento a la herramienta NooJ giró en torno a poder abordar la categoría verbal desde otro enfoque, determinando cuánto de regular y de irregular tiene un ítem léxico verbal, cómo puede identificarse esta regularidad e irregularidad, qué conocimientos se desglosan del tratamiento del verbo en NooJ y cómo podrían aplicarse para seguir trabajando y ampliando el modelo de flexión verbal. En mi experiencia con NooJ, destaco fundamentalmente la operatividad de la función Lab Morphology, ya que es una función simple, que el usuario comprende rápidamente cómo se utiliza y, además, le permite operar con diversas formas de la lengua –en este caso, con la totalidad del paradigma de conjugación verbal– que, a los fines educativos, registra cuáles son los formantes de un ítem léxico, cómo varían, cuáles son sus regularidades e irregularidades; en resumidas cuentas, una función de NooJ que permite «jugar» con el ítem léxico para aprender».

Capítulo 3

APLICACIÓN DIDÁCTICA DE NOOJ

1- Cómo trabajar con NooJ en la enseñanza del Español

El trabajo que se expone a continuación muestra cómo es posible a lo largo de tres clases interactuar con la herramienta NooJ con datos del español rioplatense de modo de poder realizar el análisis automático de un texto o bien poder trabajar con un aspecto en particular de la gramática del español como la flexión de verbos o de nombres. No es preciso para realizar estos ejercicios disponer de ningún recurso lingüístico, pero sí un mínimo conocimiento de español o bien contar con un referente docente que pueda ir proporcionando las nociones básicas de esta lengua. Al final de cada etapa, se propone una ejercitación suplementaria que permite aplicar los conceptos aprendidos. Cada clase está presentada con las tareas propias de dicha clase y luego de una manera muy gráfica se van dando mínimas explicaciones y se muestran las copias de pantalla para ir guiando el trabajo del aprendiente.

En función de esta aplicación didáctica, se plantean los siguientes objetivos específicos:

- ♦ Conocer la interfaz del programa
- ♦ Cargar textos
- ♦ Interpretar la información que proporciona NooJ antes y después de aplicar diccionarios
- ♦ Declarar propiedades
- ♦ Elaborar diccionarios
- ♦ Elaborar gramáticas morfológicas flexionales y derivacionales
- ♦ Elaborar gramáticas morfológicas productivas
- ♦ Elaborar gramáticas sintácticas

CLASE 1

En esta clase aprenderemos a:

- ♦ Conocer la interfaz de NooJ
- ♦ Cargar un texto
- ♦ Interpretar la información que brinda en análisis automático de NooJ antes y después de aplicar diccionarios
- ♦ Definir propiedades
- ♦ Elaborar un diccionario sencillo

1. EL APLICATIVO NOOJ¹

1.1. Abrir NooJ

Para comenzar, debemos abrir el programa ejecutable Nooj.exe, que se encuentra dentro de la carpeta _App y está representado por el ícono:

Si pensamos trabajar habitualmente con el programa, conviene crear un acceso directo en el escritorio.

Haciendo doble clic el ícono, se visualiza la siguiente pantalla:

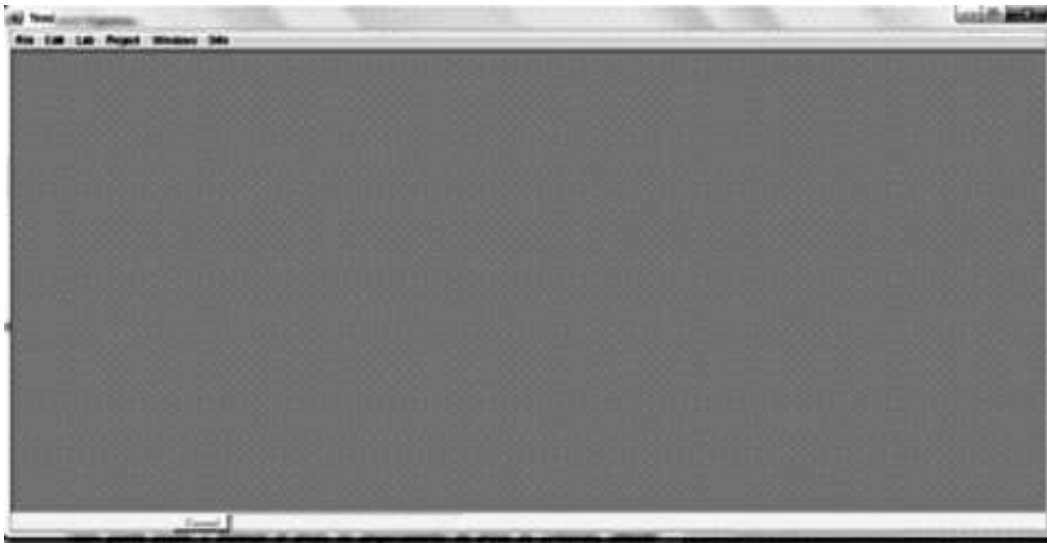


Figura 15 Pantalla de inicio del aplicativo NooJ

1.2. Seleccionar preferencias

Haciendo clic en *Info > Preferences* aparece una nueva ventana con tres solapas *General*, *Lexical Analysis* y *Syntactic Analysis*.

En *General*, inmediatamente debajo de las solapa, tenemos una pequeña ventana desplegable que nos permite seleccionar el idioma con el que vamos a trabajar, en nuestro caso *sp* (español) y hacemos clic en *Apply*.

La versión de NooJ que bajamos de Internet puede tener seleccionados por defecto algunos diccionarios o gramáticas, que presenta como modelos. Con la finalidad de empezar el trabajo desde

¹ Para completar lo que se propone esta sección se sugiere utilizar los videos tutoriales para el uso de NooJ en español disponibles en la plataforma NooJ, dichos videos están también en el siguiente link: https://www.youtube.com/watch?v=5tejAsm1ti8&list=PL3Uihe6glBR9-vyp7a0PV2LWB_gdFlkXw

cero, en la pestaña *Lexical Analysis*, en *Info>Preferences* hacemos clic sobre *Clear* o bien *Remove Resource*, ambos botones se encuentran debajo de cada uno de los paneles en los que se divide la parte superior de la ventana.

1.3. Cargar texto

En primer lugar, cargaremos un pequeño texto, que nos permitirá verificar en forma manual el análisis automático producido por NooJ. Para ello, hacemos clic en *File > New > Text*, verificamos que en la nueva ventana se encuentre seleccionada la opción *sp*, no modificamos los demás valores y hacemos clic sobre *OK* (en la parte inferior derecha de la ventana).

En la siguiente ventana podemos escribir el texto que vamos a utilizar:

*Los chicos comen galletitas en el jardín.
Un chico comía una galletita de chocolate.
En los jardines florecen los jacarandás y los jazmines.
El jacarandá del jardín de los chicos todavía no floreció.
Los chicos comerán bajo el jacarandá.
Las hormigas comieron el jazmín.*

Luego hacemos clic en *File > Save*, le asignamos el nombre tutorial (el programa agrega automáticamente la extensión *.not*) y lo guardamos tocando el botón *Guardar*.

1.4. Analizar un texto sin aplicar diccionarios ni gramáticas

Una vez que hayamos guardado el texto, veremos que en la barra de menús aparece el comando *TEXT*; si hacemos clic en *TEXT > Linguistic Analysis*, obtenemos la siguiente ventana:

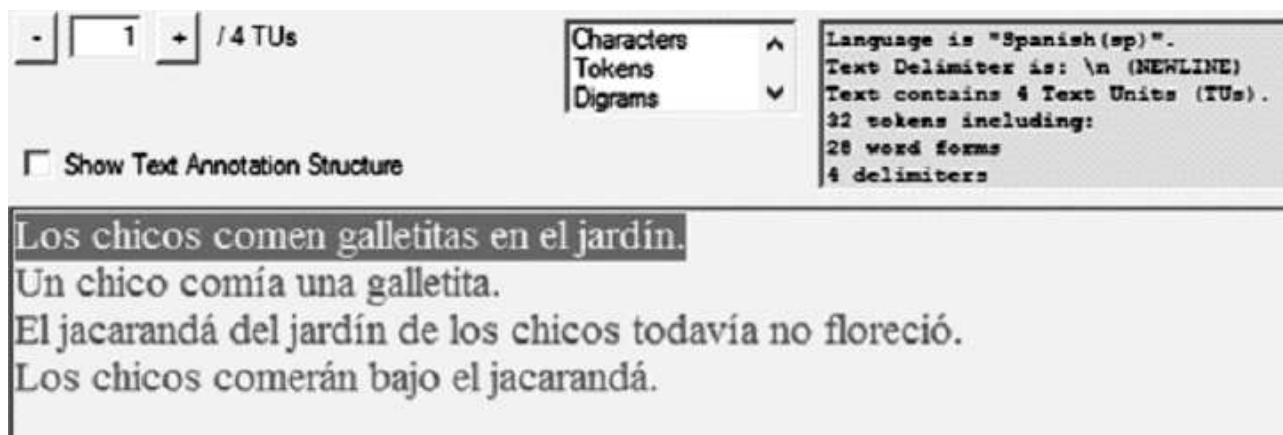


Figura 16 Análisis de oraciones con el comando *TEXT*

A continuación, podemos observar el análisis automático que produce NooJ:

Inmediatamente debajo de la barra de menús y por encima del texto (atención siempre debe estar abierto), se puede ver - de izquierda a derecha- la siguiente información:

1. Un recuadro con un número (1, en nuestro caso) precedido por el signo - y seguido por *+6TUs*:
 - a) *6TUs* indica el número de unidades textuales del archivo; en este caso se refiere a las seis párrafos que escribimos (en nuestro caso, los párrafos coinciden con oraciones, los párrafos con más de una oración cuentan como una sola unidad textual), si modificáramos el texto agregando o quitando oraciones, ese número se modificaría.
 - b) El número indica cuál es la unidad textual que se está analizando en un momento determinado. Como el texto no tiene anotaciones, por ahora, esta información es irrelevante.

2. Debajo de ese recuadro, se lee *Show Text Annotation Structure* precedido de un espacio para seleccionar esa opción. Cuando agreguemos diccionarios y gramáticas, nos permitirá ver el resultado de su aplicación en el texto.

3. En la parte central vemos una pequeña ventana con una barra de desplazamiento. Cada una de las líneas de esa ventana, que se pueden seleccionar haciendo doble clic sobre ellas, da alguna información sobre el texto:

a) *Characters*: indica la frecuencia con la que aparece cada carácter en el texto, el tipo de carácter -los caracteres pueden ser letras, espacios en blanco o delimitadores (signos de puntuación)- y el código numérico que le corresponde a cada uno de ellos en el sistema Unicode. Haciendo clic en la barra superior se pueden seleccionar distintos criterios para ordenarlos; por ejemplo, si hacemos clic en *Freq* los ordenará de mayor a menor por la frecuencia con que aparecen en el texto, en nuestro caso el carácter que aparece con mayor frecuencia es el espacio en blanco, que aparece veintiocho veces.

b) *Tokens*: indica la frecuencia con la que aparecen conjuntos de caracteres delimitados por espacios en blanco. Como se puede observar, las variaciones morfológicas de cada palabra constituyen distintos *tokens*.

c) *Digrams*: marca las agrupaciones de dos *tokens* que se repiten en el texto. En nuestro caso, la única secuencia que se repite dos veces es los *chicos*.

d) *Annotations*: abre automáticamente un diccionario en blanco para que se incorporen anotaciones al texto.

e) *Unknowns*: marca las palabras que no aparecen en ningún diccionario. En nuestro caso, las selecciona a todas porque no hay diccionarios.

f) *Ambiguities*: cuando hay diccionarios, indica cuáles son las palabras que tienen más de una entrada en los diccionarios. Por ejemplo, canto puede estar etiquetado como sustantivo y como verbo.

g) *Unambiguous Words*: marca las palabras que reciben una sola etiqueta en los diccionarios.

Haciendo clic en *TEXT > Locate* aparece la siguiente pantalla:



Figura 17 Utilización de Locate

Si escribimos cualquier token o secuencia de tokens del texto y tocamos cualquiera de los botones que se encuentran en la parte inferior de la pantalla, NooJ nos muestra todas las veces que aparece en el texto; por ejemplo, si buscamos la palabra chicos, obtenemos el siguiente resultado.

Text	Before	Seq.	After
	Los	chicos	comen galletitas en el jardín
jacarandá del jardín de los	chicos	chicos	todavía no floreció. Los chicos
chicos todavía no floreció. Los	chicos	chicos	comerán bajo el jacarandá.

Figura 18 Locate: chicos

El color de los botones determina el color de los caracteres de salida (rojo, azul, verde o negro), en la parte superior de la pantalla se pueden seleccionar la cantidad de palabras o caracteres del contexto anterior o posterior que queremos ver. En la parte inferior derecha 3/3 (no se ve en la imagen de fig 18) indica que en la pantalla se muestran tres coincidencias con la búsqueda, que es el total de las coincidencias.

1.5. Definir propiedades

Las propiedades son los rasgos que se asignan a las entradas del diccionario. Los más comunes son los rasgos categoriales (N, V, Adj, Adv, Prep, etc.) y los morfológicos (persona, número, género, tiempo, etc.).

El rasgo principal de una entrada es el categorial (V, N, ADV, ADJ, etc.) al que se suman todos los demás, que tienen distintos valores, por ejemplo, al rasgo categorial V (verbos) se le pueden sumar los rasgos de persona con los valores *1era*, *2da* y *3era*; de número con los valores *sg* y *pl*, y cualquier otro rasgo que sea relevante para el tema que se está estudiando. Los rasgos se pueden ver como columnas del diccionario. Todos los diccionarios tienen columnas predeterminadas para el rasgo categorial y para el modelo de flexión (que se verá en la próxima clase), el usuario puede definir los demás; para hacerlo debe declarar los rasgos que quiere utilizar y asociarlos a las distintas categorías, el título de la columna es el nombre del rasgo y en cada línea se muestra el valor de rasgo que corresponde a cada entrada.

Todos los diccionarios toman los rasgos declarados en un archivo que lleva el nombre de *_properties.def* y se guarda en la carpeta *Lexical Analysis* del idioma seleccionado. Para crearlo haremos clic en *File > New > Properties definition*, seleccionaremos *sp* y veremos la siguiente pantalla:

```
# NooJ V5
# Dictionary properties' definition
#
# Language is: sp
#
# Special Characters: ' _ ' = ' | ' ; ' ! # '
#
# Starting with V3, there is only ONE Dictionary Properties' Definition File per language;
# it should always be named _properties.def and be saved in the "Lexical Analysis" folder
#
# List categories and properties associated with features
# Example: N_Number = singular | plural;
# Special keyword: INFLECTION lists all inflectional features (used by variables $x$ALLF)
# Example: INFLECTION = masc | fem | Present | Futur;
```

Figura 19 Archivo Propertie's Definition

En nuestro diccionario solo incluiremos verbos (V) y sustantivos (N), de modo que a continuación de la última línea de la pantalla predeterminada escribimos:

Categorías = V | N;

El archivo de propiedades que utilizará nuestro primer diccionario será:

```
# NooJ V5
# Dictionary properties' definition
#
# Language is: sp
#
# Special Characters: '_' '=' '|' ';' '#'
#
# Starting with V3, there is only ONE Dictionary Properties' Definition File per language;
# it should always be named _properties.def and be saved in the "Lexical Analysis" folder
#
# List categories and properties associated with features
# Example: N_Number = singular | plural;
# Special keyword: INFLECTION lists all inflectional features (used by variables $x$ALLF)
# Example: INFLECTION = masc | fem | Present | Futur;

Categorías V|N;

FLEXIÓN = pres | pret | fut |1a | 2a | 3a | sg | pl| masc | fem;
V_persona = 1a | 2a | 3a;
V_UNIP = UNIP;
V_número = sg | pl;

N_género = masc | fem | g_inv | neutro;
N_número = sg | pl;
```

Figura 20 Properties Definition (Nombre y Verbo en español)

Las etiquetas no debe ser necesariamente la que utilizamos en este ejemplo; en lugar de V, podríamos utilizar VB y en lugar de N, NM. Solo debemos tener cuidado de utilizar siempre la misma etiqueta con el mismo valor.

El carácter "|" es un operador del programa que indica disyunción; dicho en lenguaje no formal, la sentencia "*Categorías = V | N;*" significa "*las categorías que utilizamos son V o N*".

En la línea siguiente declaramos los rasgos morfológicos que se van utilizar en el diccionario y, a continuación, asociamos las categorías con los rasgos flexivos que les corresponden. Al finalizar hacemos clic en *File > Save as*, le damos el nombre *_properties.def* y lo guardamos (NooJ suele tener un archivos con el nombre *_properties.def*, por lo tanto nos pedirá indicación de reemplazarlo; sería conveniente haber guardado previamente el archivo original en otra carpeta, pero, si no se efectuó esta operación, el archivo reemplazado se puede recupera bajando nuevamente el programa). Después de esto se verá un nuevo mensaje:

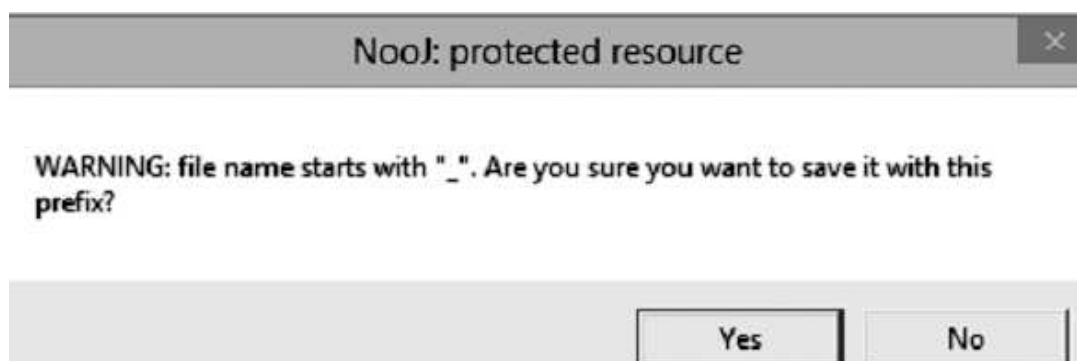


Figura 21 Mensaje de advertencia por guardar un archivo precedido por guión bajo

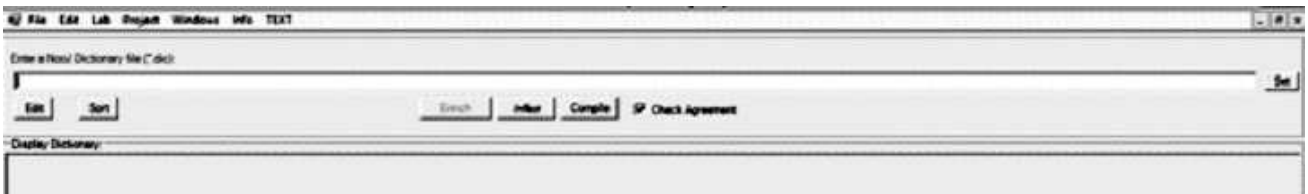


Figura 24 Compilar el diccionario

Haciendo clic en *Set*, podemos seleccionar los diccionarios disponibles para compilar, hacemos clic sobre el nombre del diccionario y luego en el botón *abrir*; de este modo, el nombre del diccionario queda en el espacio en blanco que precede a *Set*. Luego hacemos clic en *Compile*, aceptamos la advertencia de que el diccionario no está asociado con ninguna gramática flexional y las dos advertencias siguientes:

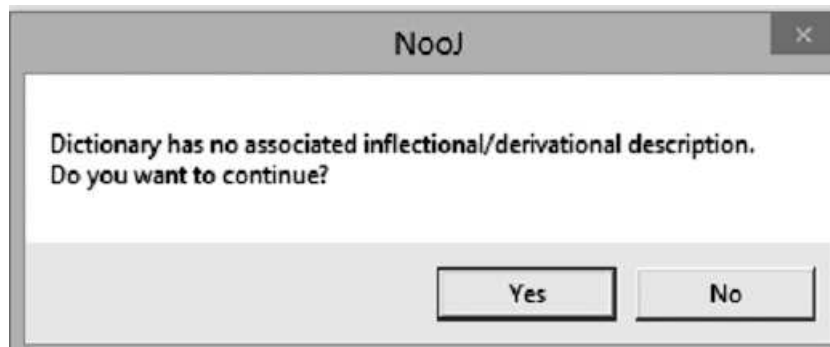


Figura 25 Mensaje de advertencia por no asociar información flexional y derivacional

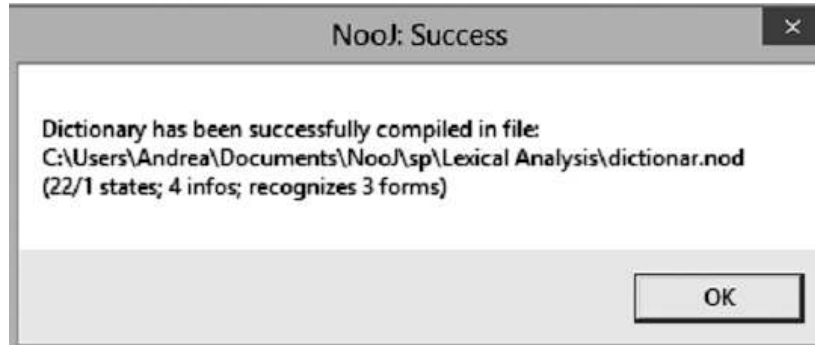


Figura 26 Mensaje que nos informa que la compilación ha sido satisfactoria

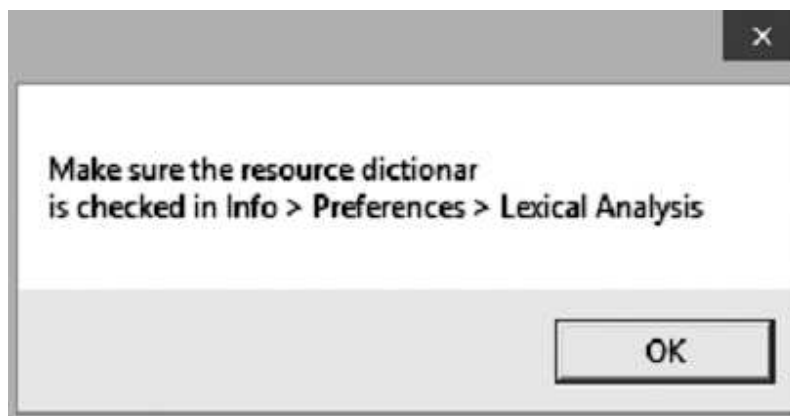


Figura 27 Mensaje que nos indica cómo volver operativo el diccionario para analizar el texto

1.7. Aplicar el diccionario al análisis del texto

Para aplicar el diccionario al análisis del texto, debemos abrirlo haciendo clic en *File > Open > Text* (el texto se puede modificar haciendo clic en *TEXT > Modify Text*). Luego hacemos clic en *Info > Preferences*; en la pestaña General, verificamos que esté seleccionado el idioma sp; en la pestaña *Lexical Analysis*, tocamos el botón *Add resources* para que aparezca el diccionario que hemos compilado y tildamos el casillero que se encuentra a su izquierda; finalmente, verificamos que en la ventana de la derecha de la pestaña *Syntactic Analysis* no haya ningún archivo seleccionado; si hubiera alguno hacemos clic sobre el nombre y lo eliminamos con *Clear* o *Remove Resource* para evitar que aplique gramáticas que no hemos hecho nosotros. Una vez que hayamos seleccionado todos los archivos que queremos aplicar al análisis, hacemos clic en *Apply*.

Para efectuar el análisis, hacemos clic en *TEXT > Linguistic Analysis* y verificamos los resultados.

a) Tildando el casillero que está a la izquierda de *Show Text Annotation Structure*, vemos que en la parte inferior de la ventana aparece un análisis de la primera oración, que muestra los sustantivos y verbos que incluimos en el diccionario con los rasgos categoriales y morfológicos que les asignamos.

b) Cada vez que hacemos clic en "+" pasamos a la siguiente oración y vemos el análisis de cada una de ellas. Como se puede ver, solo se muestran salidas de análisis en las oraciones 1, 4 y 5, que son las únicas donde se repiten algunos *tokens* de la primera oración. Nótese que no se analizan las variantes flexionales porque, si bien se trata de la misma palabra, son distintos *tokens*.

c) En la ventana central de la parte superior, se puede obtener información que no estaba disponible antes de aplicar el diccionario: haciendo doble clic en *Annotations* se abre automáticamente el diccionario y en *Unknowns* se ven las palabras que no están incluidas en el diccionario.

Ahora, con la opción *TEXT > Locate*, además de *tokens* o secuencias de *tokens*, podemos buscar categorías solas o con determinados rasgos (siempre que hayan sido incluidos en el diccionario), en este caso el objeto de la búsqueda debe aparecer entre <>; por ejemplo, si queremos buscar sustantivos, pondremos <N>; si queremos buscar sustantivos en singular, pondremos <N+sg>, si queremos buscar sustantivos masculinos en singular, pondremos <N+mas+sg>. No es posible buscar rasgos que no estén asociados a alguna categoría; por lo tanto, si ponemos <+mas+sg> el resultado de la búsqueda es *Concordance is empty*. En la siguiente pantalla se muestra el resultado de la búsqueda <N> en el contexto de las cinco palabras anteriores y las cinco posteriores:

The screenshot shows a software window with a menu bar (File, Edit, Lab, Project, Windows, Info, TEXT, CONCORDANCE) and a toolbar with options like 'Reset', 'Display', 'characters before, and 5 after', 'Display: Matches', and 'Outputs'. Below the toolbar is a table with three columns: 'Text', 'Before', 'Seq.', and 'After'. The table displays the search results for the word 'chicos' and 'jacarandá' in a text passage.

Text	Before	Seq.	After
	Los	chicos	comen galletitas en el jardín
chico comía una galletita. El		jacarandá	del jardín de los chicos
jacarandá del jardín de los		chicos	todavía no floreció. Los chicos
chicos todavía no floreció. Los		chicos	comerán bajo el jacarandá.
Los chicos comerán bajo el		jacarandá	.

Figura 28 Búsqueda con Locate de: chicos y jacarandá

EJERCITACIÓN

- 1) Modifique el texto y agregue las oraciones: *El camino del jardín es de grava* y *Camino por el jardín*.
- 2) Agregue al diccionario los sustantivos y los verbos de estas dos oraciones (verifique que no estén ya incluidos y recuerde guardar los cambios)
- 3) Compile el diccionario.
- 4) Analice el texto modificado.
- 5) Indique la frecuencia de caracteres *a, d, c, ". "* y espacios en blanco.
- 6) ¿Cuáles son los *tokens* que aparecen con mayor frecuencia?
- 7) ¿Cuáles son los *digrams* que hay en el texto y con qué frecuencia aparecen?
- 8) ¿Qué función utilizaría para saber qué palabras del texto no están incluidas en el diccionario?
- 9) ¿Qué ambigüedades encuentra NooJ y cómo las analiza?
- 10) ¿Cómo efectuaría la búsqueda de todos los verbos del texto?, ¿qué verbos obtiene con la aplicación del diccionario modificado?

CLASE 2

En esta clase aprenderemos a:

- ♦ Crear gramáticas regulares y gráficas, morfológicas y derivacionales
- ♦ Conocer los operadores que se pueden utilizar
- ♦ Elaborar diccionarios que utilicen estas gramáticas
- ♦ Evaluar los diccionarios creados y los resultados de distintas consultas

1. Gramáticas morfológicas

Los diccionarios no electrónicos del español tienen como entrada los infinitivos de los verbos y la forma masculina singular de los adjetivos y sustantivos. Antes de efectuar una búsqueda, el usuario aplica intuitivamente su conocimiento gramatical para asociar las variantes morfológicas de una palabra con la forma que proporciona el diccionario. Esta operación es elemental para la inteligencia humana, pero no para la inteligencia artificial. Por lo tanto, para lograr el análisis automático de un texto, es necesario lograr que el sistema informático reconozca todas las variaciones que puede tener una palabra.

Si pretendiéramos analizar automáticamente un texto con diccionarios diseñados con el criterio que aplicamos en la clase anterior, tendríamos que incluir como entrada cada una de las variantes de cada palabra. Sin embargo, dado que las variaciones morfológicas tienen patrones que presentan cierta regularidad, resulta mucho más sencillo y económico crear gramáticas que contengan modelos que se puedan asociar con distintas palabras, así, por ejemplo el modelo del verbo *cantar*, se puede aplicar a todos los verbos regulares de la primera conjugación y el modelo de *perro* se puede aplicar a todos los sustantivos y adjetivos que tengan masculino terminado en *o*, femenino en *a* y plural en *s*. Estas gramáticas permiten que un autómatas efectúe una operación similar a la que aplica un usuario humano al consultar un diccionario tradicional.

1.1. Editor de gramáticas regulares

NooJ ofrece una aplicación para crear gramáticas. En primer lugar crearemos una gramática flexiva regular. Si hacemos clic en *File > New > Grammar*, aparece una ventana dividida en dos paneles superiores -(1) *Select Languages* y (2) *Optional Information*- y uno inferior -*Select Grammar Type*-. En el superior izquierdo (1) seleccionamos el idioma de entrada y de salida; en nuestro caso ambos son *sp* (español); el panel superior derecho permite consignar el autor y la institución que desarrolla la gramática, seleccionar opciones del tamaño del editor gráfico (*Large graph*), y bloquear acceso de la gramática mediante claves (estas funciones son opcionales y no las utilizaremos); en el cuadro inferior seleccionamos el tipo de gramática que vamos a desarrollar. Como indica la pantalla, las gramáticas morfológicas (*Inflection & Derivation* y *Productive Morphology*) reconocen secuencias de letras y las gramáticas sintácticas (*Syntax*) secuencias de palabras.

Dado que lo que queremos crear es una gramática flexiva regular, haremos clic en *rule editor* y luego en *Inflection & Derivation* y aparecerá la siguiente pantalla:

```

# Generic Commands:
# <B>: keyboard Backspace
# <C>: change Case
# <D>: Duplicate current char
# <E>: Empty string
# <L>: keyboard Left arrow
# <N>: go to end of Next word form
# <P>: go to end of Previous word form
# <R>: keyboard Right arrow
# <S>: delete/Suppress current char
# Arguments for commands <B>, <L>, <N>, <P>, <R>, <S>:
# xx number: repeat xx times
# W: whole word
# Examples
# <R3>: go right 3 times
# <LW>: go to beg. of word
#
# Language-Specific Commands:
# <A>: remove accute accent from last letter aeiou
# <Á>: add acute accent to last letter aeiou
# <À>: add grave accent to last letter aeiou
# <Â>: add circumflexe to last letter aeiou
# <Ä>: add dieresis/trema/umlaut to last letter aeiou
# <U>: replace last 'e' with 'ie' or last 'o' with 'ue' or last 'u' with 'ue' or last 'i' with 'ie'
# <V>: replace last 'g' with 'gu' or last 'z' with 'c' or last 'c' with 'qu' or last 'u' with 'ü'
# <X>: replace last 'g' with 'j' or last 'c' with 'z' or last 'qu' with 'c' or last 'gu' with 'g'
# <Y>: replace last 'e' with 'i' or last 'o' with 'u'
#
# Special Characters: '=' '<' '>' '\' '""' ':' ']' '+' '-' '/' '$' '_' '?' '#'

```

Figura 29 Creación de una gramática flexiva regular con los comandos necesarios para el español (Language-Specific Comands)

En la figura 29, se ven los comandos que se pueden utilizar; para mayor claridad los repetimos traducidos con una breve explicación:

 Borra caracteres de izquierda a derecha. Por ejemplo; PERRO = a/fem+sg, indica que debe borrar la "o" y agregar la "a" con la etiqueta femenino singular.

<C> Cambia minúsculas por mayúsculas y viceversa.

<D> Duplica un carácter.

<E> Agrega un elemento vacío. Se utiliza cuando queremos poner una etiqueta a un elemento que no está representado por ningún carácter. Por ejemplo, CANTAR = <E>/inf, indica que a la cadena se le debe agregar la etiqueta sin agregar ni quitar ningún carácter. En cambio, CANTAR = é/fut+1era+sg, indica que debe agregar el carácter "é" y agregar la etiqueta mencionada.

<L> Mueve el cursor a la izquierda.

<N> Mueve el cursor al final de la siguiente palabra.

<P> Mueve el cursor al final de la palabra anterior.

<R> Mueve el cursor a la derecha.

Los comandos se escriben en mayúscula y entre <> y las etiquetas se escriben precedidas de /; si se omiten estos operadores, NooJ lo interpreta como un carácter y se limita a agregarlos a la cadena anterior; por ejemplo, si escribimos PERRO = Ba fem+sg, obtendremos la cadena "perrBa fem+sg" como si fuera una palabra.

Si se agregan números arábigos después del comando, la operación se repite la cantidad de veces que indica el número. Por ejemplo, si ponemos <B3>, borra tres caracteres de izquierda a derecha; si ponemos <L2> mueve el cursor dos caracteres a la izquierda. Es posible combinar varios comandos; por ejemplo, LUCIR = <B2><L>z<R>o/pte+1era+sg, produce "luzco" con la etiqueta correspondiente.

El comando <W> indica la palabra completa; por ejemplo <LW> mueve el cursor hacia la izquierda, al principio de la palabra.

<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.
 <À> Agrega tilde grave a la primera vocal que encuentra de izquierda a derecha.
 <Â> Agrega tilde circunfleja a la primera vocal que encuentra de izquierda a derecha.
 <U> Reemplaza la última 'e' por 'ie' o la última 'o' por 'ue' o la última 'u' por 'ue' o la última 'i' por 'ie'.
 <V> Reemplaza la última 'g' por 'gu' o la última 'z' por 'c' o la última 'c' por 'qu' o la última 'u' por 'ü'.
 <X> Reemplaza la última 'g' por 'j' o la última 'c' por 'z' o la última 'qu' por 'c' o la última 'gu' por 'g'.
 <Y> Reemplaza la última que se indican en la pantalla 'e' por 'i' o la última 'o' por 'u'.

A continuación de esta pantalla comenzaremos a desarrollar una pequeña gramática que permita reconocer las variaciones morfológicas de las palabras que aparecen en el texto que analizamos en la clase anterior.

En primer lugar, se indica el nombre del modelo, que es la referencia que se utiliza en el diccionario para aplicar las operaciones que se establecen en la gramática. La única restricción para el nombre del modelo es que debe estar escrito en mayúscula; para facilitar su identificación tomamos una palabra; por ejemplo la palabra "CHICO" puede servir de nombre para el modelo de sustantivos que en femenino reemplazan el último carácter por "a" y agregan "s" al plural, y "COMER" como nombre del modelo de los verbos regulares de la segunda conjugación.

En el texto *chicos* aparece solo en masculino singular y plural; si además de estas variaciones, quisiéramos agregar los femeninos tendríamos que hacer alguna de estas reglas:

CHICO = <E>/masc+sg | s/masc+pl | a/fem+sg | as/fem+pl;
 CHICO = (<E>/masc | a/fem) s/pl;

Ambas reglas producen el mismo resultado. La diferencia entre ellas es que la primera es más clara y sencilla para el usuario porque efectúa todas las operaciones a partir de la palabra completa, pero es menos económica desde el punto de vista del procesamiento porque agrega dos veces cada etiqueta; mientras la segunda, por el contrario, es más económica para el procesamiento, pero menos clara para el usuario. De cualquier modo, para enunciar una regla se utilizan los siguientes elementos:

1. La barra hacia la derecha "/" indica que lo que sigue es la etiqueta.
2. Cuando a una misma cadena de caracteres se agrega más de una etiqueta, a partir de la segunda, se utiliza el signo "+".
3. La barra recta "|" indica alternativa: en la primera regla decimos que agregue la cadena vacía "<E>" seguida de las etiquetas "masc" y "sing" o que agregue "s" seguida de las etiquetas masc y pl o etc.
4. El espacio en blanco indica concatenación: en la segunda regla decimos que agregue la cadena vacía seguida de la etiqueta "masc" o borre el último carácter y agregue "a" seguida de la etiqueta "fem" y que luego concatene cualquiera de ambos resultados con una "s" seguida de la etiqueta "pl".
5. Los paréntesis agrupan elementos: la ausencia de paréntesis en la segunda regla daría como resultado que "s/pl" se agregara únicamente a la operación "a/fem" y, consecuentemente, no obtendríamos el masculino plural.
6. Cada regla finaliza en punto y coma.

La siguiente gramática permite generar y analizar todas las variantes morfológicas de los sustantivos y adjetivos que aparecen en el texto:

1.3. Resultados del análisis y las consultas

Antes de aplicar el nuevo diccionario al análisis del texto, debemos compilarlo como se explicó en la clase anterior. Una vez que lo tengamos compilado, abrimos el texto tutorial.not (File > Open > Text); en la barra de menú, tocamos Info > Preferences; en la pestaña General, seleccionamos Lenguaje Name sp; en Lexical Analysis, desmarcamos el diccionario Tutorial.dic, marcamos Tutorial1.dic y hacemos clic sobre Apply; y, por último, hacemos clic en TEXT>Linguistic Analysis.

Si marcamos el casillero Show Text Annotation Structure, podemos ver que NooJ reconoce todos los sustantivos y verbos e identifica el lema, que es la entrada que se le dio en el diccionario (sustantivo masculino y verbo infinitivo) y los rasgos morfológicos que lo caracterizan:

File Edit Lab Project Windows Info TEXT

1 / 4 TUs

Characters
Tokens
Digrams

Language is "Spanish(sp)".
Text Delimiter is: \n (NEWLINE)
Text contains 4 Text Units (TUs).
92 tokens including:
28 word forms
4 delimiters

Show Text Annotation Structure

Los chicos comen galletitas en el jardín.
Un chico comía una galletita.
El jacarandá del jardín de los chicos todavía no floreció.
Los chicos comerán bajo el jacarandá.

0	11	17	34
chico, N+masc+pl	comer, V+pres+3era+pl	galletita, N+fem+pl	jardín, N+masc+sg

Figura 32 Linguistic Analysis del texto Tutorial.not

Ahora, haciendo clic en TEXT > Locate, además de buscar una categoría con un rasgo morfológico determinado, se pueden buscar todas las variantes morfológicas de un lema, para hacerlo basta escribir en minúsculas cualquier variante del lema entre <>. El resultado de la búsqueda <comer> es:

File Edit Lab Project Windows Info TEXT CONCORDANCE

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

Text	Before	Seq.	After
	Los chicos	comen	galletitas en el jardín. Un
	en el jardín. Un chico	comía	una galletita. El jacarandá del
	todavía no floreció. Los chicos	comerán	bajo el jacarandá.

Figura 33 Búsqueda con Locate de <comer>

Si ponemos *comer* sin paréntesis angulares, no obtenemos ningún resultado porque, en este caso, NooJ busca la palabra exacta, que no se encuentra en el texto; si ponemos *comen* sin paréntesis angulares obtendremos una sola salida, en cambio, si ponemos <comen> obtendremos todas las variantes. Esta función podría ser de utilidad para personas que estudian el español como segunda lengua porque les permitiría identificar todas las variantes morfológicas que aparecen en un texto.

1.4. Editor gráfico de gramáticas

Las gramáticas gráficas utilizan los mismos operadores y producen los mismos resultados que las gramáticas regulares. Pueden resultar más sencillas, porque evitan el uso de paréntesis y punto y coma.

Para desarrollar con el sistema de gráficos la misma gramática morfológica que hicimos con el sistema de reglas, hacemos clic en File > New > Grammar, verificamos que se encuentre seleccionado sp como idioma de entrada y de salida, marcamos el casillero circular **graphical editor**, luego, tocamos Infection & Derivation y aparece la siguiente pantalla:

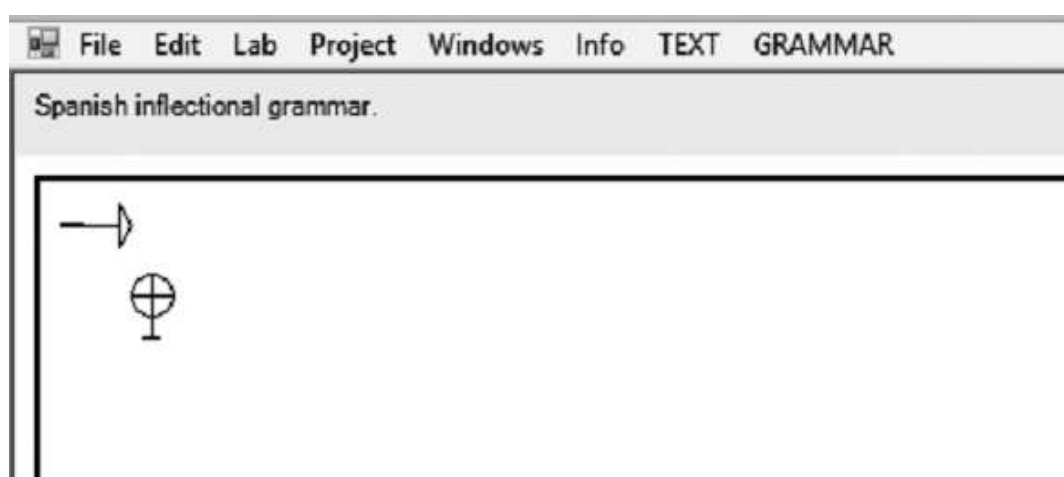


Figura 34 Creación de una gramática flexiva con Graphical editor

La flecha representa el estado inicial y el círculo con la cruz el estado final. Entre ambos, debemos crear los estados intermedios; para ello, arrastramos con el mouse el círculo hacia abajo a la derecha y la flecha hacia abajo; luego, mientras mantenemos presionada la tecla ctrl hacemos clic y se crea un nuevo nodo vacío, representado por <E>.



Figura 35 Se abre un nodo con Control+Clic

Para crear el gráfico que corresponde a CHICO, en el primer nodo dejamos la categoría vacía y agregamos la etiqueta "/masc" y, mientras mantenemos presionada la tecla ctrl presionamos enter; luego, hacemos clic en cualquier lugar de la pantalla para deseleccionar ese nodo.



Figura 36 Se valida con Control+Enter, rasgo masculino

Al lado de este nodo, creamos otro, donde a la categoría vacía le agregamos la etiqueta "/sg"; para el femenino, borramos la categoría vacía y escribimos "a/fem"; para el plural, borramos la categoría vacía y agregamos "s/pl" (recuerde que los operadores del programa van entre <> y los caracteres (letras) van sin paréntesis angulares). Para poder asociar la entrada del diccionario con el modelo de flexión, debemos asignarle un nombre al modelo; a tal efecto, hacemos clic en GRAMMAR > Show Structure y a la izquierda se despliega un nuevo panel con la palabra Main, hacemos doble clic sobre ella y escribimos CHICO y guardamos la gramática con el nombre *tutorialgraf*.



Figura 37 Gramática flexiva para CHICO, creación de nodos

A continuación, es necesario marcar el camino que debe seguir el proceso de NooJ uniendo los nodos. En este caso, el proceso es de izquierda a derecha, pero también es posible hacerlo en sentido inverso o unir un nodo sobre sí mismo para repetirlo. Por esta razón, suele suceder que se establezcan conexiones inesperadas; para borrarlas, es suficiente repetir el procedimiento que se utilizó para crearlas.

Para vincular los nodos, seleccionamos el nodo inicial (la flecha) haciendo clic sobre él (aparecen cuatro círculos de colores en cada uno de sus vértices) y lo unimos con el de arriba (<E>/maso) con otro clic, después repetimos el procedimiento para unir el nodo inicial con el de abajo (a/fem), luego unimos cada uno de ellos con <E>/sg y con s/pl y, a estos con el estado final (el círculo con la cruz). El resultado es el siguiente:

CHICO

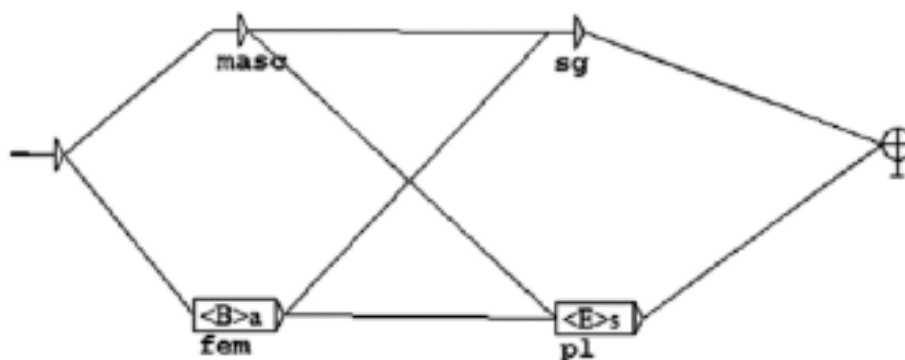


Figura 38 Se unen los nodos para formar la flexión de CHICO

Haciendo clic en GRAMMAR>New Graph, en el panel izquierdo aparece una nueva ramificación al mismo nivel de CHICO con el nombre Untitled 0; haciendo clic sobre Untitled 0, cambiamos el nombre por el del modelo que vamos a crear (en nuestro caso, CHOCOLATE) y, en el panel derecho creamos, el gráfico que corresponde. A continuación, se muestran los gráficos de cada uno de los modelos utilizados en el diccionario, cada uno de ellos se crea haciendo clic en GRAMMAR>New Graph y cambiando el nombre Untitled 0 por el nombre del modelo:

CHOCOLATE

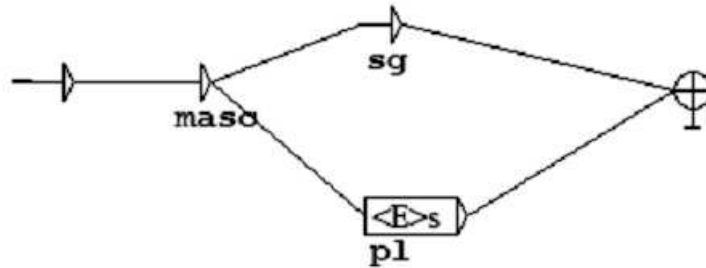


Figura 39 Gramática para la flexión de CHOCOLATE

COMER

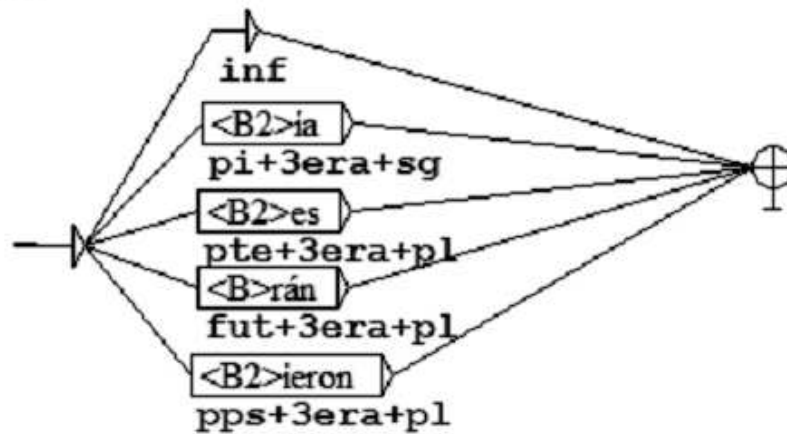


Figura 40 Gramática para la flexión de COMER

GALLETITA

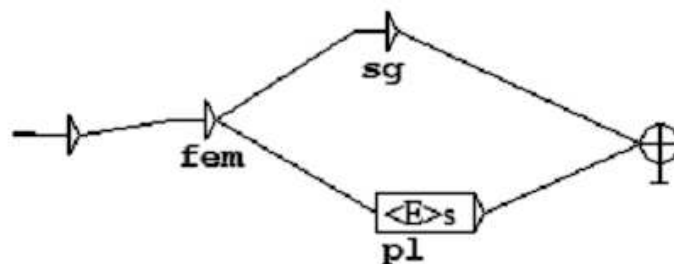


Figura 41 Gramática para la flexión de GALLETITA

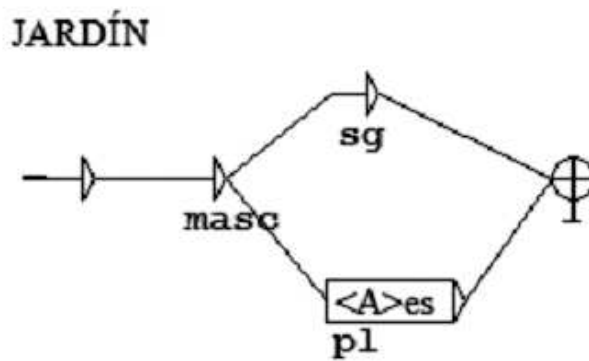


Figura 42 Gramática para la flexión de JARDÍN

Una vez que hayamos incluido todos los modelos; guardamos la gramática, reabrimos el diccionario tutorial1.dic; en la primera línea, reemplazamos tutorial.nof por tutorialgraf.nof y compilamos nuevamente el diccionario. De este modo, podemos comprobar que la gramática gráfica funciona exactamente igual que la gramática regular. Consecuentemente, se obtendrá el mismo análisis del texto y las consultas arrojarán los mismos resultados.

2. Gramáticas derivacionales

Las gramáticas derivacionales utilizan los mismos operadores que las gramáticas morfológicas y, también permiten optar entre utilizar sistemas de reglas o gráficos. En las gramáticas y el diccionario que elaboramos, *galletita* se trató como una palabra simple; pero, en realidad, se trata de un diminutivo, es decir, de una palabra derivada. Si queremos tratarla como tal, deberemos reemplazar en las gramáticas morfológicas (tutorial.nof o tutorialgraf.nof) el modelo GALLETITA por GALLETA y en el diccionario (tutorial.dic) modificaremos la entrada galletita,N+FLX=GALLETITA por galleta,N+FLX=GALLETA (se deben cambiar todas las palabras que siguen el modelo GALLETITA), luego crearemos una gramática derivacional. Tenga en cuenta que el diccionario utiliza solo la gramática que declaramos en la primera línea, no obstante, sugerimos cambiar las dos para conservar la equivalencia)

A continuación, vamos a elaborar una gramática derivacional que genere y reconozca los diminutivos de los sustantivos del texto. Hacemos clic en File>New > Grammar, seleccionamos *sp*, *rule editor* y hacemos clic en Inflection & Derivation. En esta nueva gramática vamos a crear los modelos para los diminutivos, que guardaremos con el nombre derivación (NooJ agrega la extensión .nof):

IT = <L>it/N+dim; (galletita: va un carácter hacia atrás, agrega it y, luego, la etiqueta categorial N con la especificación +dim)

ITO = ito/N+dim; (chocolatito: borra la e final, agrega ito y, luego, la etiqueta categorial N con la especificación +dim)

QUIT = <L>quit/N+dim; (chiquito: va un carácter hacia atrás, borra la c, agrega quit y, luego, la etiqueta categorial N con la especificación +dim)

GUIT = <L>uit/N+dim; (hormiguita: va un carácter hacia atrás, agrega guit y, luego, la etiqueta categorial N con la especificación +dim)

CITO = <A>cito/N+dim; (jardincito, jazmincito y jacarandacito: quita el acento de la vocal, agrega cito y, luego, la etiqueta categorial N con la especificación +dim).

Para que la etiqueta del tipo de derivación *dim* se pueda ver como una columna del diccionario, deberemos modificar el archivo de propiedades (_properties.def) y agregar debajo de la línea FLEXIÓN una línea DERIVACIÓN = dim; y, debajo de las propiedades de N, N_ derivación = dim; (no omita guardar)

Si se utilizan otros rasgos para caracterizar otros tipos derivación (aumentativa, despectiva, etc.), en la misma línea, antes del punto y coma, agregaremos precedido de la barra recta "|" en la lista de rasgos derivacionales y, luego, como una propiedad derivacional de la categoría que corresponda.

jacarandás,jacarandá,N+FLX=CHOCOLATE+DRV=CITO+masc+pl
 jacarandacito,jacarandá,N+dim+FLX=CHOCOLATE+DRV=CITO+masc+sg
 jacarandacitos,jacarandá,N+dim+FLX=CHOCOLATE+DRV=CITO+masc+pl
 jazmín,jazmín,N+FLX=JARDÍN+DRV=CITO+masc+sg
 jazmines,jazmín,N+FLX=JARDÍN+DRV=CITO+masc+pl
 jazmincito,jazmín,N+dim+FLX=JARDÍN+DRV=CITO+masc+sg
 * jazmincitoes,jazmín,N+dim+FLX=JARDÍN+DRV=CITO+masc+pl
 hormiga,hormiga,N+FLX=GALLETA+DRV=GIT+fem+sg
 hormigas,hormiga,N+FLX=GALLETA+DRV=GIT+fem+pl
 hormiguita,hormiga,N+dim+FLX=GALLETA+DRV=GIT+fem+sg
 hormiguitas,hormiga,N+dim+FLX=GALLETA+DRV=GIT+fem+pl
 florecer,florecer,V+FLX=COMER+inf
 florecerán,florecer,V+FLX=COMER+fut+3era+pl
 florecen,florecer,V+FLX=COMER+pte+3era+sg
 florecía,florecer,V+FLX=COMER+pps+3era+sg
 florecieron,florecer,V+FLX=COMER+pps+3era+pl

Como puede verse en las líneas resaltadas, el plural del diminutivo de *jardín* y *jazmín* es incorrecto. Esto sucede a causa de que, salvo que haya alguna otra indicación, NooJ asigna al derivado el mismo modelo flexivo que a la palabra simple. En los otros casos no se produce ningún error porque la palabra derivada tiene el mismo modelo flexivo que la simple, por ejemplo, en *chico* el masculino singular coincide con la entrada, es decir, no se agrega ni se quita ningún carácter para generar esta forma flexiva, el diminutivo es idéntico: *chiquito* es masculino singular sin agregar ni quitar ningún carácter; si borramos la *o* y agregamos *a*, generamos el singular tanto en la variación chico / chica como en *chiquito/chiquita* y, si agregamos *s*, transformamos las formas del singular y del plural en ambos casos (*chicos/chicas, chiquitos/chiquitas*). En cambio, *jardín* y *jazmín* tienen plural en -es mientras *jazmincito* y *jardincito* tienen plural en -s.

Cuando el modelo flexivo del derivado no coincide con el modelo de la palabra simple de la que deriva es posible asignarle otro modelo de flexión agregando ":" y el nombre del nuevo modelo. En el caso de *jardincito* y *jazmincito*, son palabras masculinas invariables con plural en -s, es decir, siguen el modelo flexivo de chocolate. Por lo tanto, para salvar el error, en el diccionario, al modelo derivacional debemos agregar el nuevo modelo flexivo:

jazmín,N+FLX=JARDÍN+DRV=CITO:CHOCOLATE
 jardín,N+FLX=JARDÍN+DRV=CITO:CHOCOLATE

Si hacemos esta modificación en nuestro diccionario y volvemos a compilarlo, podremos observar que los errores desaparecen.

Además de modificar el modelo flexivo, es posible modificar la etiqueta categorial del derivado. Por ejemplo *comedor* es un sustantivo derivado del verbo *comer*; en un diccionario podemos optar por tratar estas nominalizaciones como palabras simples y darles entrada como sustantivo, o bien, tratarlas como derivaciones. Para ello, debemos crear en la gramática un modelo como el siguiente:

DOR= dor/N+nom;

Si a cualquier verbo de un diccionario agregamos la marca +DRV=DOR:JARDÍN, por indicación de la gramática derivacional, el sistema borra un carácter, agrega *dor*, lo etiqueta como N seguida por el rasgo nom (nominalización) (no olvidemos agregar el rasgo en el archivo `_properties.def`) y, por indicación de la gramática flexiva, lo flexiona como JARDÍN. El inconveniente de no introducir *comedor* directamente en el diccionario es que, por tratarse de una palabra derivada, no es posible agregar una nueva derivación, como sería la de diminutivo.

Nótese que el modelo flexivo JARDÍN tiene el operador <A>, que indica que quite el acento de la vocal; como el derivado *comedor* no tiene ninguna vocal acentuada, el operador no encuentra su argumento y se neutraliza, consecuentemente, no genera error.

EJERCITACIÓN

1) Modifique el texto y agregue las oraciones:

Las hormiguitas hicieron un caminito del jacarandá al jazmín.

El comedor da a un jardincito.

Las hormigas comieron el jazmincito y el jacarandacito del jardín.

2) Analice el texto utilizando el diccionario tutorial2.dic

3) Consulte en TEXT > Locate cuántos diminutivos y cuántas nominalizaciones aparecen en el texto y anote los resultados.

4) Elabore una gramática derivacional gráfica con los mismos modelos que la gramática regular, denomínela *derivacióngraf* y aplíquela al diccionario *tutorial.dic* (debe reemplazar la gramática *derivación.nof* por esta nueva gramática en la segunda línea del diccionario).

5) Compile el diccionario, vuelva a analizar el texto y efectúe la consulta para verificar que se localicen la misma cantidad de diminutivos y nominalizaciones que en la consulta anterior.

6) Agregue la oración: *La nueva ley prevé que el consumo de marihuana sea legal, pero el tráfico seguirá siendo ilegal.*

7) En la gramática flexiva, agregue los modelos de flexión faltantes para las palabras simples (*ver, ser, legal, traficar, seguir*)

8) En una de las gramáticas derivacionales (regular o gráfica) agregue los modelos de las palabras derivadas (*prever, tráfico, ilegal*). Recuerde que el operador <LW> se utiliza para llevar el cursor al principio de la palabra.

9) En `_properties.def`, agregue los rasgos flexivos y derivacionales que faltan.

10) Agregue al diccionario las nuevas palabras con su etiqueta categorial y el modelo flexivo y derivacional que les corresponda (asegúrese de que el diccionario utilice las gramáticas modificadas).

11) Compile el diccionario, genere nuevamente *tutorial-flex.dic* y verifique que todas las entradas sean correctas.

12) Analice el texto y efectúe consultas.

13) Agregue al diccionario palabras que sigan los modelos flexivos y derivacionales existentes, compílelo y verifique que todas las formas generadas sean correctas.

CLASE 3

En esta clase aprenderemos a:

- ♦ Crear gramáticas productivas
- ♦ Realizar consultas aplicando los diccionarios y gramáticas productivas
- ♦ Comparar la eficiencia de ambos
- ♦ Crear gramáticas sintácticas gráficas simples
- ♦ Utilizar gráficos subyacentes
- ♦ Aplicar las gramáticas sintácticas al análisis del texto

1. Gramáticas productivas

Las gramáticas productivas se utilizan para reconocer palabras que no están incluidas en los diccionarios, pero que tienen determinados rasgos formales comunes. Por ejemplo, sería prácticamente imposible crear un diccionario con todos los nombres propios que pueden aparecer en un gran corpus de textos periodísticos; en ese caso, resultaría de cierta utilidad crear una gramática productiva que etiquetara las palabras que comienzan en mayúscula como nombres propios.

Haciendo clic en File > New > Grammar; seleccionamos sp como lenguaje de entrada y salida, graphical editor; y hacemos clic en Productive Morphology; se despliega una pantalla idéntica a la del editor gráfico de gramáticas flexivas y derivacionales.

El módulo morfológico de NooJ usa los siguientes símbolos especiales:

- ♦ <L> cualquier letra
- ♦ <U> cualquier letra mayúscula
- ♦ <W> cualquier letra minúscula
- ♦ <A> cualquier letra acentuada
- ♦ <N> cualquier letra sin acento

Para crear una gramática que etiquete cualquier palabra en mayúscula como nombre propio, ubicamos el estado inicial y el estado final, y creamos un nuevo nodo haciendo ctrl clic, reemplazamos el nodo vacío <E>, que aparece automáticamente cuando creamos un nodo, por <U> (cualquier letra mayúscula); hacemos clic en cualquier parte de la pantalla para deseleccionar ese nodo; creamos un nuevo nodo <W> (cualquier letra minúscula), volvemos a hacer clic en cualquier parte de la pantalla para deseleccionarlo y creamos un nuevo nodo donde a la cadena vacía <E> le agregamos, precedida de barra, la etiqueta N+PR. Luego unimos los nodos de izquierda a derecha, hacemos doble clic sobre el nodo <W> para que se una sobre sí mismo, lo que indica un número indeterminado de letras minúsculas, y guardamos con el nombre npropio, al que se le agrega automáticamente la extensión .nom. La gramática resultante expresa que cualquier secuencia que empieza con una letra mayúscula y sigue con un número indeterminado de minúsculas se debe etiquetar como N+Pr:

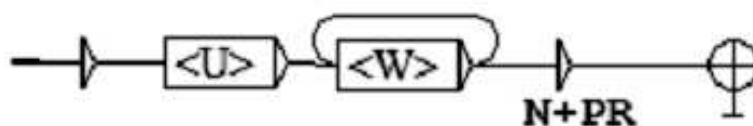


Figura 44 Gramática morfológica para el nombre propio en español

Si hacemos clic en la barra de menús GRAMMAR > Debug... y bajamos la barra de desplazamiento de la ventana principal, vemos un panel donde podemos introducir distintas cadenas para verificar cuáles son las cadena incluidas y excluidas por esta gramática. Dado que la gramática es inespecífica, reconocerá como nombre propio una cadena como Exmnopx, en cambio no reconocerá un nombre propio escrito en minúscula:

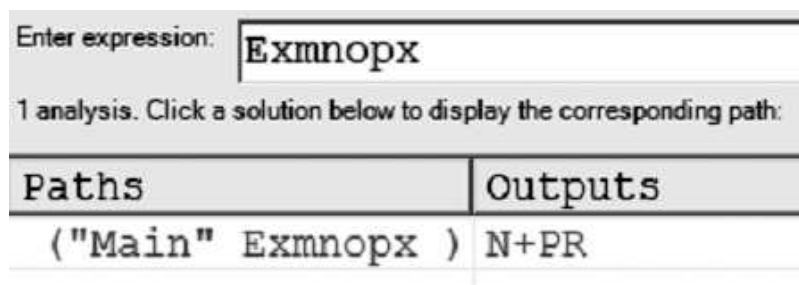


Figura 45 Aplicación de Show Debug en la gramática nombre propio

Antes de comprobar el funcionamiento de esta gramática en el texto, en el archivo de propiedades agregamos el rasgo

N_tipo = PR;

Luego abrimos el texto, en Info > Preferences, en el panel superior derecho de la pestaña Lexical Analysis, marcamos la gramática npropio.nom y, luego, Apply.



Figura 46 Selección de la gramática npropio en Info>Preferences

Si hacemos un nuevo análisis del texto, veremos que todas las palabras mayúsculas de comienzo de oración resultan etiquetadas como N+PR, lo cual es incorrecto.

Con la finalidad de seguir analizando el funcionamiento del sistema, vamos a agregar en el archivo de propiedades las etiquetas correspondientes a estas palabras (ART y PREP) y los rasgos que determinan al artículo:

ART_tipo = det | indet;

ART_género = masc | fem;

ART_número = sg | pl;

Luego, incluimos en el diccionario tutorial.dic los artículos con sus rasgos de género y número y las preposiciones que aparecen en el texto, y compilamos.

```

los, ART+masc+pl
el, ART+masc+sg
la, ART+masc+sg
las, ART+masc+pl
en, PREP

```

Figura 47 Entradas que se agregan al diccionario

En un nuevo análisis del texto, vemos que cuando estas palabras aparecen en mayúscula Nool aplica una doble (ART y N+Pr). Si en TEXT > Locate hacemos la búsqueda de la expresión regular <N> (sustantivo), aparecerán erróneamente todos los artículos y las preposiciones que inician oración, como se muestra en la siguiente pantalla.

Text	Before	Seq.	After
Los chicos comen galletitas en el jardín. Un chico comía una galletita. El jacarandá del jardín de los chicos todavía no floreció. Los chicos comerán bajo el jacarandá.	Los	Los	chicos comen galletitas en el comen galletitas en el jardín en el jardín. Un chico . Un chico comía una galletita chico comía una galletita. El comía una galletita. El jacarandá . El jacarandá del jardín de jacarandá del jardín de los del jardín de los chicos de los chicos todavía no todavía no floreció. Los chicos chicos comerán bajo el jacarandá .
Los chicos comen galletitas en el jardín. Un chico comía una galletita. El jacarandá del jardín de los chicos todavía no floreció. Los chicos comerán bajo el jacarandá.	Los	Los	chicos comen galletitas en el comen galletitas en el jardín en el jardín. Un chico . Un chico comía una galletita chico comía una galletita. El comía una galletita. El jacarandá . El jacarandá del jardín de jacarandá del jardín de los del jardín de los chicos de los chicos todavía no todavía no floreció. Los chicos chicos comerán bajo el jacarandá .
Los chicos comen galletitas en el jardín. Un chico comía una galletita. El jacarandá del jardín de los chicos todavía no floreció. Los chicos comerán bajo el jacarandá.	Un	Un	chicos comen galletitas en el comen galletitas en el jardín en el jardín. Un chico . Un chico comía una galletita chico comía una galletita. El comía una galletita. El jacarandá . El jacarandá del jardín de jacarandá del jardín de los del jardín de los chicos de los chicos todavía no todavía no floreció. Los chicos chicos comerán bajo el jacarandá .
Los chicos comen galletitas en el jardín. Un chico comía una galletita. El jacarandá del jardín de los chicos todavía no floreció. Los chicos comerán bajo el jacarandá.	Un	Un	chicos comen galletitas en el comen galletitas en el jardín en el jardín. Un chico . Un chico comía una galletita chico comía una galletita. El comía una galletita. El jacarandá . El jacarandá del jardín de jacarandá del jardín de los del jardín de los chicos de los chicos todavía no todavía no floreció. Los chicos chicos comerán bajo el jacarandá .
Los chicos comen galletitas en el jardín. Un chico comía una galletita. El jacarandá del jardín de los chicos todavía no floreció. Los chicos comerán bajo el jacarandá.	El	El	chicos comen galletitas en el comen galletitas en el jardín en el jardín. Un chico . Un chico comía una galletita chico comía una galletita. El comía una galletita. El jacarandá . El jacarandá del jardín de jacarandá del jardín de los del jardín de los chicos de los chicos todavía no todavía no floreció. Los chicos chicos comerán bajo el jacarandá .
Los chicos comen galletitas en el jardín. Un chico comía una galletita. El jacarandá del jardín de los chicos todavía no floreció. Los chicos comerán bajo el jacarandá.	El	El	chicos comen galletitas en el comen galletitas en el jardín en el jardín. Un chico . Un chico comía una galletita chico comía una galletita. El comía una galletita. El jacarandá . El jacarandá del jardín de jacarandá del jardín de los del jardín de los chicos de los chicos todavía no todavía no floreció. Los chicos chicos comerán bajo el jacarandá .
Los chicos comen galletitas en el jardín. Un chico comía una galletita. El jacarandá del jardín de los chicos todavía no floreció. Los chicos comerán bajo el jacarandá.	Los	Los	chicos comen galletitas en el comen galletitas en el jardín en el jardín. Un chico . Un chico comía una galletita chico comía una galletita. El comía una galletita. El jacarandá . El jacarandá del jardín de jacarandá del jardín de los del jardín de los chicos de los chicos todavía no todavía no floreció. Los chicos chicos comerán bajo el jacarandá .
Los chicos comen galletitas en el jardín. Un chico comía una galletita. El jacarandá del jardín de los chicos todavía no floreció. Los chicos comerán bajo el jacarandá.	Los	Los	chicos comen galletitas en el comen galletitas en el jardín en el jardín. Un chico . Un chico comía una galletita chico comía una galletita. El comía una galletita. El jacarandá . El jacarandá del jardín de jacarandá del jardín de los del jardín de los chicos de los chicos todavía no todavía no floreció. Los chicos chicos comerán bajo el jacarandá .

Figura 48 Aplicación de la gramática de nombre propio con resultados incorrectos

Para eludir el doble análisis y el etiquetado erróneo, debemos hacer clic en Info > Preferences; en la parte superior derecha de la pestaña, vemos el título "Before y After". Seleccionando con un clic las gramáticas productivas, mediante el uso de estos botones podemos asignarle distintos niveles de prioridad; si asignamos bajo nivel (Before) se aplicará una gramática antes que otra.

En síntesis, las gramáticas productivas son de utilidad si los diccionarios que utilizamos son completos; de lo contrario, arrojarán demasiados resultados erróneos, que habrá que revisar manualmente.

2. Gramáticas sintácticas

Mientras las unidades mínimas de las gramáticas flexivas y derivativas, y de las productivas son los caracteres, la unidad mínima de las gramáticas sintácticas son las palabras. Es posible crear gramáticas sintácticas utilizando palabras, por ejemplo, *chico*; lemas, por ejemplo <chico>; categorías léxicas, por ejemplo <V>; o categorías acotadas con determinados rasgos, por ejemplo <V+1era>, que incluye cualquier verbo en primera persona. Para poder utilizar lemas o categorías, es necesario que estos estén declarados en los diccionarios.

2.1. Gramáticas sintácticas con palabras

A continuación, desarrollaremos una gramática sintáctica que reconozca los sintagmas nominales formados por un determinante y un sustantivo. Hacemos clic en File > New > Grammar, verificamos que el lenguaje de entrada y el de salida sea sp, marcamos *graphical editor* y hacemos clic sobre el botón Syntax. La pantalla que aparece a continuación es idéntica a la de las gramáticas gráficas anteriores, la diferencia es que en las gramáticas sintácticas separan con espacios en blanco las unidades de cada nodo. Los procedimientos de creación y unión de nodos son iguales a las anteriores, por lo tanto, no nos detendremos a explicarlos.

La siguiente, es una gramática del SN desarrollada a partir de palabras, la llamamos SNPAL (Nool agrega automáticamente la extensión .nog):

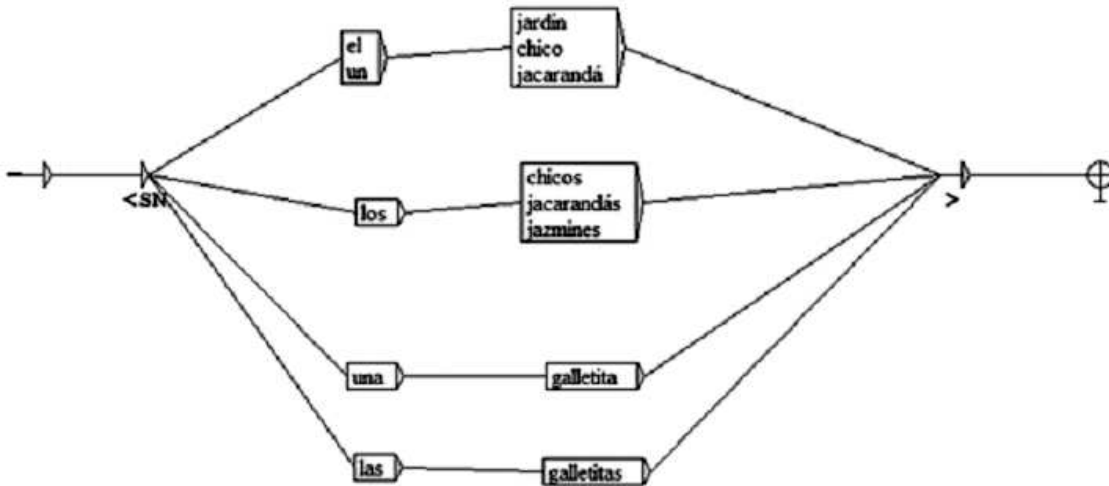


Figura 49 Gramática del SN a partir de palabras (SNPAL)

El nodo que sigue al inicial es un nodo vacío con la etiqueta del sintagma: <E>/<SN. La etiqueta va precedida de un paréntesis angular de apertura que cierra en el nodo que precede al terminal: <E>/> esto nodos se incluyen para indicar el comienzo y el final del sintagma. Los artículos y los sustantivos se agrupan en los siguientes nodos por género y número; cuando un nodo tiene más de un elemento después de cada uno de ellos se inserta un *enter*.

Haciendo clic en GRAMMAR > Debug y bajando la barra de desplazamiento de la ventana principal, podemos verificar distintas cadenas que la gramática reconoce como válidas o inválidas:

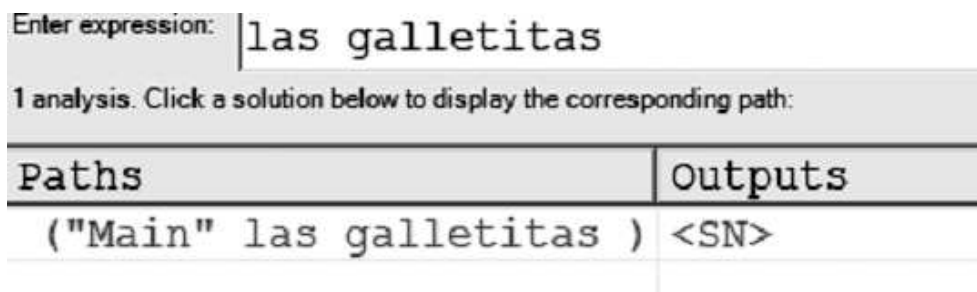


Figura 50 Aplicación de SHOW DEBUG sobre la gramática SNPAL

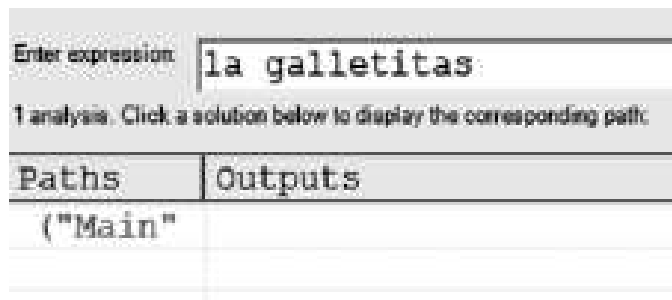


Figura 51. Busca de la secuencia que la gramática SNPAL reconoce como SN. La etiqueta *es incorrecta*

Para analizar el texto con esta gramática; lo abrimos, en la pestaña de Syntactic Analysis (Info > Preferences), hacemos clic en el nombre de la gramática, que se encuentra en panel de la izquierda, lo pasamos al panel de la derecha con el paréntesis angular rojo > que hay entre ambos paneles y hacemos clic en Apply. Luego, hacemos clic en TEXT > Linguistic Analysis y en Show Text Annotation Structure. Así podremos ver que en cada oración los SN se marcan con una línea verde.

En TEXT>Locate podemos seleccionar a NooJ grammar, buscar la gramática SNPAL.nog y ver todas las secuencias del texto que la gramática reconoce como SN haciendo clic en cualquiera de las teclas coloreadas de la parte inferior del panel. Seleccionando el casillero Outputs (a la derecha, debajo de la barra de menús) se ve la etiqueta, deseleccionándolo se oculta.

Text	Before	Seq.	After
		Los chicos/<SN>	comen galletitas en el jardín
Los chicos comen galletitas en el jardín.		el jardín/<SN>	. Un chico comía una galletita
comen galletitas en el jardín.		Un chico/<SN>	comía una galletita. El jacarandá
el jardín. Un chico comía una galletita.		una galletita/<SN>	. El jacarandá del jardín de
Un chico comía una galletita.		El jacarandá/<SN>	del jardín de los chicos
El jacarandá del jardín de los chicos todavía no floreció.		los chicos/<SN>	todavía no floreció. Los chicos
Los chicos comerán bajo el jacarandá.		Los chicos/<SN>	comerán bajo el jacarandá.
		el jacarandá/<SN>	.

Figura 52 Aplicación de SNPAL con el texto abierto desde TEXT>LOCATE

2.2. Gramáticas sintácticas con categorías léxicas

Una gramática que utilice categorías en lugar de palabras sería mucho más eficiente; para desarrollarla, abrimos una nueva gramática, a la que denominaremos SNCAT, y reemplazamos las palabras por etiquetas categoriales con sus rasgos de género y número:

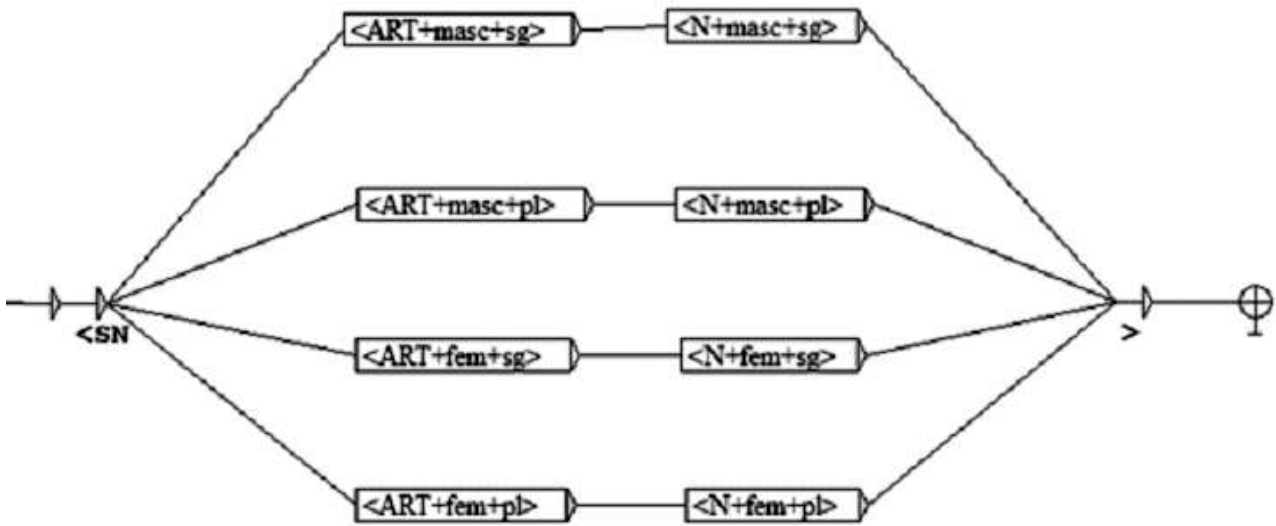


Figura 53 Creación de una gramática de categorías: SNCAT

Para analizar el texto con esta nueva gramática, en Syntactic Analysis (Info > Preferences), hacemos seleccionamos la SNPAL.nog , la eliminamos con REMOVE RESOURCE y hacemos clic en Apply.

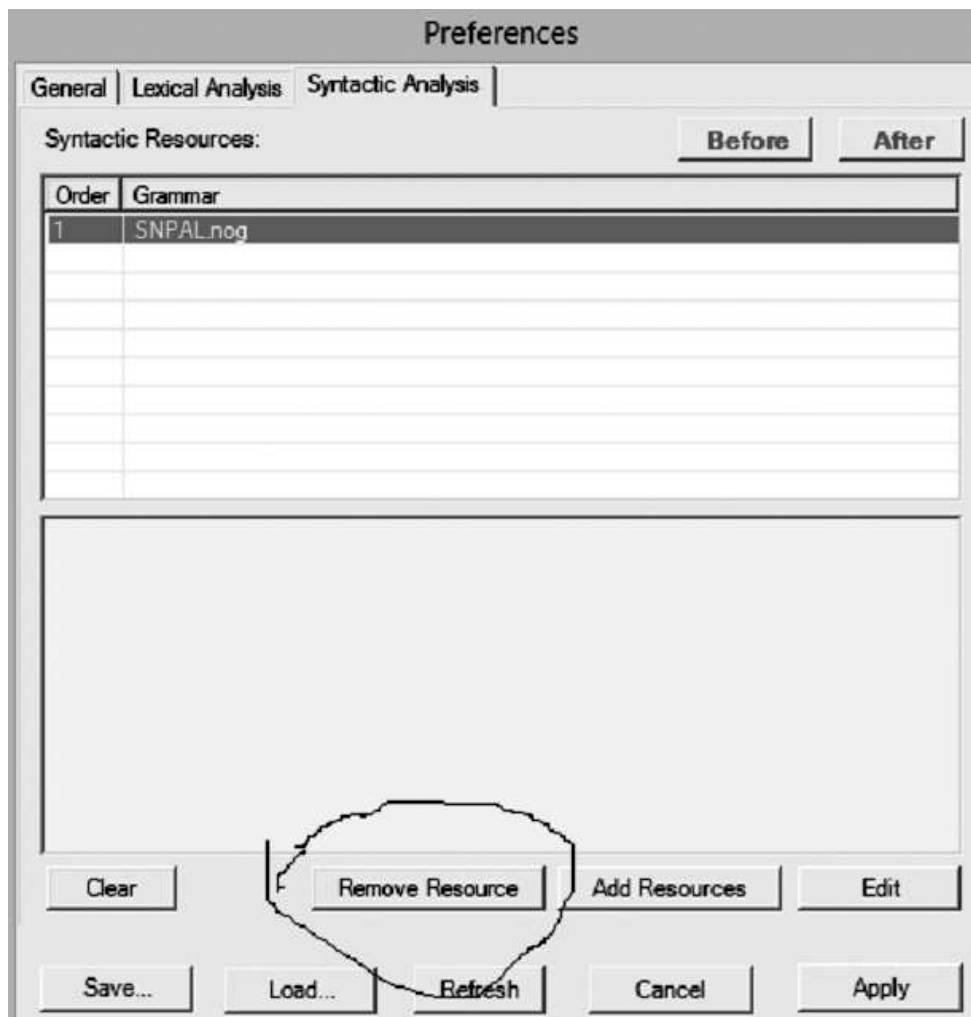


Figura 54 Se remueve la gramática SNPAL en Info>Preferences

Luego con ADD RESOURCES, seleccionamos SNCAT.nog, (Open)



Figura 55 Selección de la gramática SNCAT en Info>Preferences

Luego, hacemos clic en TEXT > Linguistic Analysis y en Show Text Annotation Structure. Ahora podemos ver que la nueva gramática es mucho más eficiente que la anterior porque reconoce todos los SN sin necesidad de incluir cada una de las palabras.

Si la seleccionamos en TEXT > Locate > a NooJ grammar, aparecen las siguientes salidas:

Text	Before	Seq.	After
Los chicos comen galletitas en el jardín.		Los chicos/<SN>	comen galletitas en el jardín
Un chico comía una galletita.		el jardín/<SN>	. Un chico comía una galletita
El jacarandá del jardín de los chicos todavía no floreció.		El jacarandá/<SN>	del jardín de los chicos
floreció. Los chicos comerán bajo el jacarandá.		los chicos/<SN>	todavía no floreció. Los chicos
		Los chicos/<SN>	comerán bajo el jacarandá.
		el jacarandá/<SN>	.

Figura 56 Resultados de SNCAT sobre el texto, con OUTPUTS se marca SN

2.3. Gramáticas sintácticas con categorías sintácticas

Además de palabras y categorías léxicas, las gramáticas sintácticas pueden tomar como entradas unidades definidas por otras gramáticas sintácticas. Una gramática sencilla de este tipo es la del sintagma preposicional, formado por una preposición (categoría léxica definida en el diccionario) y un SN (categoría sintáctica definida en una gramática).

Antes de crear una gramática que los reconozca los SP, tenemos que agregar en el diccionario las preposiciones y las contracciones que utilizamos en el texto.

Las preposiciones se agregan directamente con la etiqueta categorial:

- en,PREP
- a,PREP
- de,PREP
- bajo,PREP

Las contracciones deben ser reconocidas como preposición más artículo:

- al,<a,PREP><el,ART+masc+sg>
- del,<de,PREP><el,ART+masc+sg>

Luego, creamos la siguiente gramática sintáctica, que denominamos SP.nog:

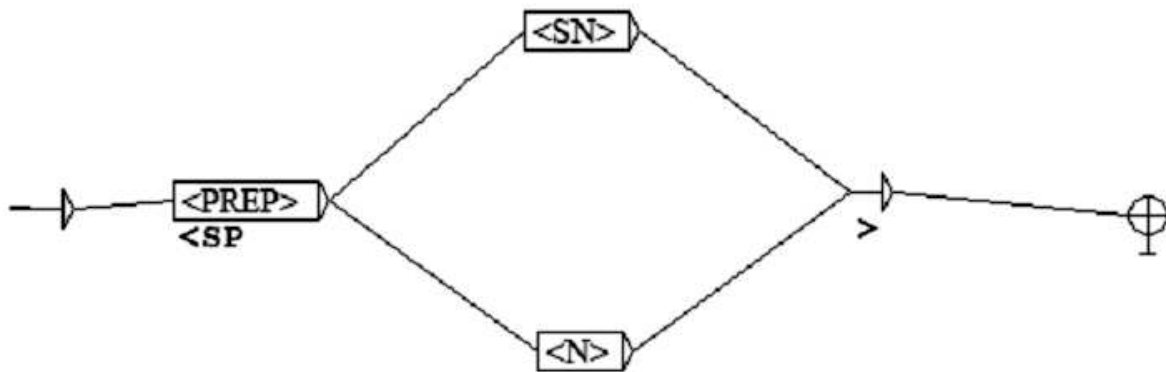


Figura 57 Creación de la gramática sintáctica SP.nog

Esta gramática reconoce como SP cualquier secuencia que contenga una preposición y un sustantivo (<N> en la rama inferior del gráfico) o un sintagma nominal (<SN> en la superior); esta última rama presupone la aplicación previa de una gramática sintáctica que haya reconocido el SN.

Para aplicar la gramática al análisis del texto, en *Info>Preferences, Syntactic Analysis*, mediante *Add Resources*, seleccionamos la gramática SP.nog:

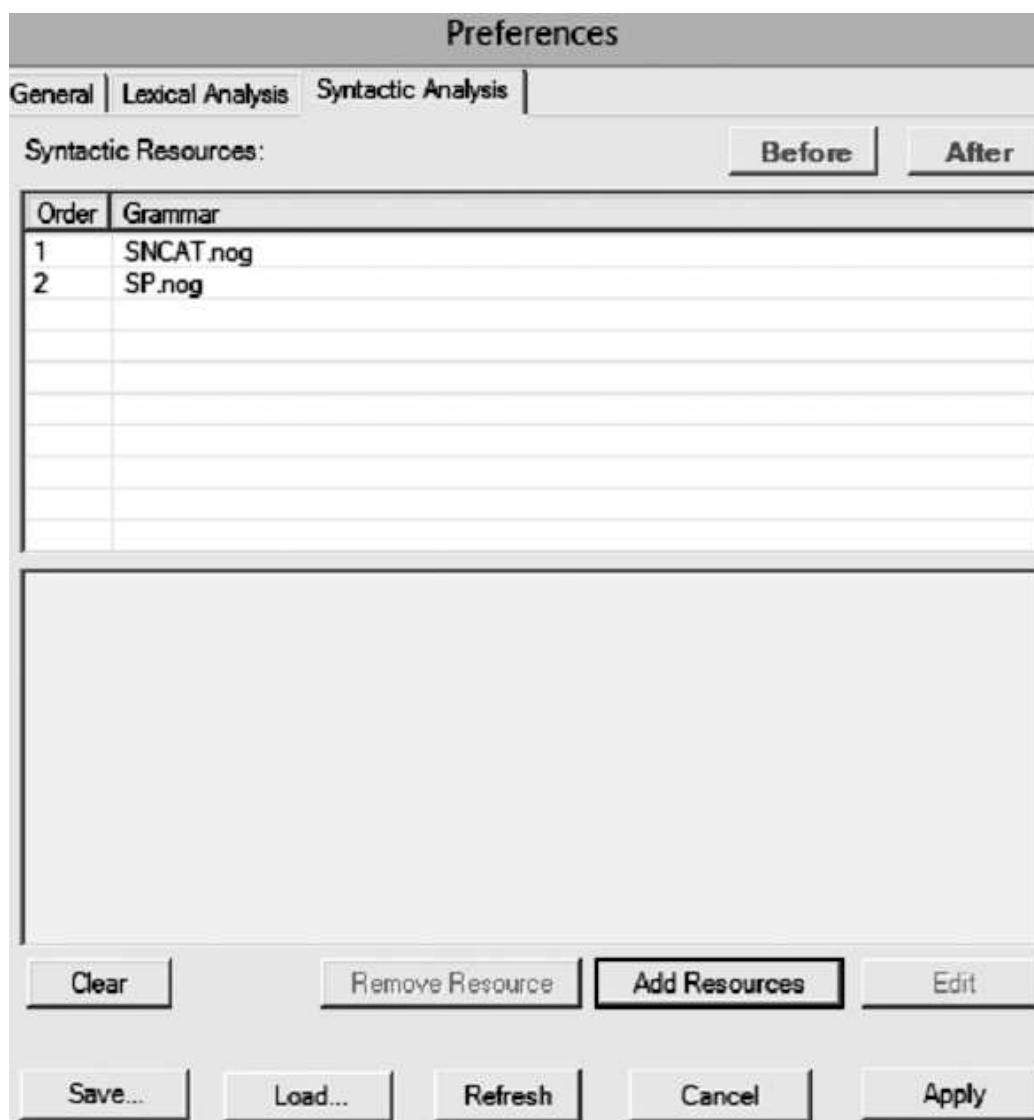


Figura 58 Seleccionar la gramática SP.nog mediante Add Resources

El orden en el que aparecen en el panel derecho es relevante: SNCAT.not debe estar antes que SP.nog porque la salida del primero es una entrada del segundo y, si se invierte el orden, no se produce el análisis.

El resultado del análisis es el siguiente:

The screenshot shows a software interface for linguistic analysis. At the top, there are controls for 'Characters', 'Tokens', and 'Digrams', and a status bar indicating 'Language is "Spanish(sp)", Text Delimiter is: \n (NEWLINE), Text contains 4 Text Units (TUs), 32 tokens including: 28 word forms, 4 delimiters'. Below this, the text to be analyzed is displayed: 'Los chicos comen galletitas en el jardín. Un chico comía una galletita. El jacarandá del jardín de los chicos todavía no floreció. Los chicos comerán bajo el jacarandá.' The bottom part of the image shows a grid representing the syntactic structure. The first row shows 'V+pres+3era+pl' pointing to 'galletita,N+fem+pl' (token 17), which is part of an 'SP' (Sentence Phrase) starting at token 28. The second row shows 'en,PREP' pointing to 'SN' (Noun Phrase) starting at token 31. The third row shows 'el,ART+masc+sg' pointing to 'jardín,N+masc+sg' (token 34).

Figura 59 Aplicar SP.nog después de SNCAT.nog

Como se puede ver, el análisis tiene dos niveles: el de SN (ART + N) y el de SP (ART +SN). Si seleccionamos SP.nog en TEXT>Locate >a NooJ grammar, aparecen las siguientes salidas:

The screenshot shows the 'LOCATE' tool interface. The menu bar includes 'File', 'Edit', 'Lab', 'Project', 'Windows', 'Info', 'TEXT', and 'CONCORDANCE'. Below the menu, there are search parameters: 'Reset', 'Display: 5 characters before, and 5 after', 'word forms', 'Matches', and 'Outputs'. The main display area shows the text 'Los chicos comen galletitas en el jardín/<SP> . Un chico comía una galletita' with the search results highlighted.

Figura 60 Análisis del texto con LOCATE, se selecciona SP.nog

2.4. Gramáticas sintácticas con gráficos subyacentes

En la gramática anterior, el SN es una unidad lingüística independiente, por eso no aparece como un elemento sintáctico del SP en la salida de la consulta TEXT > Locate > a NooJ grammar. Si queremos que en esta consulta se visualice que SN es una unidad sintáctica, debemos introducir el SN como un gráfico subyacente.

Para lograrlo, hacemos clic en File > New > Grammar, seleccionamos sp en el idioma de entrada y de salida, graphical editor, Syntax, ubicamos el estado inicial y final, creamos los nodos <PREP>/<SP, :SN y <E>/>, y los unimos de izquierda a derecha.

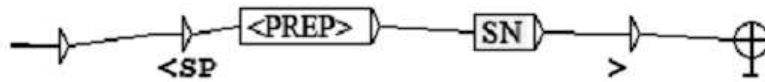


Figura 61 Sintagma Preposicional incluyendo el SN

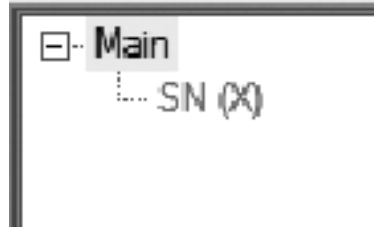


Figura 62 Creación de un nuevo grafo

Al unirlos, :SN se resalta automáticamente dando la posibilidad de crear un gráfico subyacente. Hacemos clic en Show Structure y a la izquierda aparece un nuevo panel con una ramificación SN (X), hacemos clic sobre esa ramificación y aparecerá superpuesto el siguiente cartel:



Figura 63 Para crear con Show Structure

Hacemos clic en Sí y se abre la pantalla de un nuevo gráfico, que recibe automáticamente la denominación SN. Allí ubicamos el estado inicial y el estado final, creamos los nodos correspondientes al SN, los unimos de izquierda a derecha y guardamos la gramática con el nombre SP1.

Luego abrimos el texto, en el panel derecho de la pestaña *Syntactic Analysis*, seleccionamos las gramáticas *SNCAT.nog* y *SP.nog*, las eliminamos utilizando el botón *Clear*, con *Add Resources* seleccionamos *SP1.nog* con y volvemos a analizar el texto (*TEXT>Linguistic Analysis*).

Como podemos comprobar, el análisis que produce *SP1.nog* en la parte inferior de la pantalla es igual al que produce *SP.nog*. Sin embargo, en la consulta *TEXT > Locate > a NooJ grammar* los resultados son diferentes:

Text	Before	Seq.	After
Los chicos comen galletitas	en el jardín/	<SP<SN>>	. Un chico comía una galletita

Figura 64 Análisis con TEXT>LOCATE aplicando SP1.nog

Como se puede ver, acá, el SN es un formante del SP. Esto sucede, porque en el primer caso SN no está incluido en *SP.nog*, por lo tanto, para esta gramática es equivalente a una categoría léxica.

EJERCITACIÓN

- 1) Elabore una gramática productiva para abreviaturas como aprox., avda., sr., Bs. As., Cía., EE. UU., FF. AA., R.I.P. (recuerde que una gramática puede tener más de una ramificación).
- 2) Elabore una gramática para siglas como ONU, FMI, UNAM, ONG
- 3) Verifique las gramáticas introduciendo ejemplos positivos y negativos en GRAMMAR> Debug...
- 4) Elabore una gramática sintáctica para oraciones transitivas e intransitivas (*Las hormigas hicieron un caminito. Los chicos juegan*)
- 5) Agregue al texto diferentes oraciones con estas estructuras.
- 6) Agregue al diccionario los nuevos lemas utilizados (no olvide compilarlo).
- 7) Analice el texto con estas gramáticas.
- 8) Agregue a la gramática un camino que incluya a las oraciones transitivas que tienen una oración subordinada como objeto directo (*Juan dijo que Pedro no podía venir*)
- 9) Agregue al texto oraciones que contengan estas estructuras.
- 10) Actualice el diccionario.
- 11) Vuelva a analizar el texto.

CONCLUSIONES

A lo largo de este libro, se ha tomado un supuesto fundamental entendiendo que al utilizar una herramienta informática es posible generar un conocimiento explícito de la lengua. Desde este supuesto y en función del uso de la plataforma NooJ creemos que es posible realizar hipótesis en torno al lenguaje, en este caso el español, ya que al corroborar en máquina las diferentes hipótesis, se produce dicho conocimiento.

Se intentó partir de elementos de observación que permitieran pensar en si existe o no una gramática de jóvenes. Para ello, nos valimos de la página de Facebook "Aprendo con NooJ" . Se transformó en un recurso eficaz, ya que insertando en ella un documento de Google logramos que jóvenes de 18 a 25 años contestaran una serie de preguntas que nos permitieron elaborar un diagnóstico sobre sus producciones y establecer algunas primeras conclusiones de trabajo.

En la Segunda parte, presentamos ejercicios concretos para la utilización de NooJ en español, en dichas clases se dan simultáneamente dos aprendizajes: cómo trabajar con la herramienta informática y a la vez se va dando la reflexión lingüística, en este caso sobre los datos del español. En efecto, se vuelve sobre lo que ya se conoce, por ejemplo, en torno al género, en torno a los modelos verbales, en torno a diferentes construcciones sintácticas.

De ningún modo se espera agotar esta temática, simplemente constituye apenas un primer acercamiento. Sustenta la hipótesis de que al partir de una gramática de jóvenes, la misma debe ser dilucidada como puntapié inicial de la clase de lengua. El "error" en este caso se convierte en un lugar de análisis, sobre el que es preciso pensar juntos. Sin duda estamos en proceso de cambio permanente, inclusive habría que registrar cómo se va introduciendo paulatinamente la iconicidad en la comunicación entre jóvenes, en función del chat, del emoticón, del famoso esquere como expresión que irrumpe desde el rap. La aceptación del otro es parte de aceptar cómo habla y cómo escribe.

En este contexto, nos ofrece un terreno ideal la neutralidad de una herramienta informática vacía de contenido y donde vamos cargando los datos del español, así como la posibilidad de ir cotejando lo que pasa en otras lenguas desde un formato común y por tanto avizorar la adquisición de una proficiencia "multilingüística". Esto va delineando un nuevo camino que nos enfrenta como docentes investigadores ante elementos dignos de ser atendidos en su complejidad creciente. Una búsqueda permanente por acentuar el carácter proactivo de la enseñanza, que no puede concebirse sin dar una participación real y efectiva a los estudiantes.

BIBLIOGRAFÍA

- Beltrán, Celina, Solana, Zulema y Tramallino, Carolina, Análisis automático morfológico con las herramientas SMORPH y NOOJ, *Revista de Epistemología y Ciencias Humanas*, Agosto de 2013, ISSN N° 1852-625X, disponible en: <http://www.revistaepistemologi.com.ar/biblioteca/12%20solana%20tramallino%20beltran.pdf> última consulta: mayo 2018
- Blanco, Xavier & Lajmi, Douha. (2004). Dictionnaire électronique français-espagnol-catalan-arabe des noms des professions et des métiers.
- Blanco, Xavier. (2002) Regroupements sémantiques dans un dictionnaire d'adverbes composés en Espagnol. Linda Barone, Mario Monteleone and Max Silberztein. 2016. Automatic Processing of Natural-Language Electronic Texts with NooJ, NooJ 2016. DOI: 10.1007/978-3-319-55002-2. Communications in Computer
- Blanco, Xavier (2015) "A Hierarchy of Semantic Labels for Spanish Dictionaries", in Okrut, T, Hetsevich, Y, Silberztein, M & Stanislavenka, H. (eds): Automatic Processing of Natural-Language Electronic Texts with NooJ. Communications in Computer and Information Science, volume 607: 66-73.
- Barone, Linda, Monteleone, Mario and Silberztein, Max. (2016) Automatic Processing of Natural-Language Electronic Texts with NooJ, NooJ 2016. DOI:10.1007/978-3-319-55002-2. Communications in Computer and Information Science, volume 667: 1-258.
- Bès, Gabriel G. (1999). La phrase verbal noyau en français. Recherches sur le français parlé, volume 15: 273-358. Université de Provence.
- Bès, Gabriel G. y Solana, Zulema (2007) Sintagma verbal núcleo flexionado en español, [http://infosurrevista.com.ar/biblioteca/INFOSUR-Nro1-2007-BesSolana\(1\).pdf](http://infosurrevista.com.ar/biblioteca/INFOSUR-Nro1-2007-BesSolana(1).pdf), última consulta: mayo 2018
- Blanco, Xavier (2015) "A Hierarchy of Semantic Labels for Spanish Dictionaries", en Okrut, T, Hetsevich, Y, Silberztein, M & Stanislavenka, H. (eds): Automatic Processing of Natural Language Electronic Texts with NooJ. Communications in Computer and Information Science, volume 607: 66-73. En: Formalizing Natural Languages with NooJ 2014, Johanna Monti, Max Silberztein, Mario Monteleone and Maria Pia di Buono (Eds.), Cambridge Scholars Publishing London, 186-197.
- Bonino, Rodolfo y Rodrigo, Andrea (2012) Una propuesta para la extracción automática del Sintagma Adverbial, *Revista Infosur* N° 6, ISSN 1851-1956, Diciembre de 2012, disponible en: <http://www.infosurrevista.com.ar/biblioteca/INFOSUR-Nro6-2012-Rodrigo-Bonino.pdf>, última consulta: mayo 2018
- Bonino, Rodolfo (2011) Una propuesta para la implantación en máquina de la morfología verbal en NooJ, *Revista Infosur* N° 6, ISSN 1851-1956, Octubre de 2011, disponible en: <http://www.infosurrevista.com.ar/biblioteca/INFOSUR-Nro5-2011-Bonino.pdf>, última consulta: mayo 2018
- Duckworth, Michael (1995) *Grammar and Practice*. Oxford University Press. Oxford
- Frigière, Julia y Fuentes, Sandrine (2015) Pedagogical Use of NooJ dealing with French as a Foreign Language. En: Formalizing Natural Languages with NooJ 2014, Johanna Monti, Max Silberztein, Mario Monteleone and Di Buono, María Pía (Eds.), Cambridge Scholars Publishing, London, 186-197.
- Monteleone, Mario (2016). Local Grammars and Formal Semantics: Past Participles vs. Adjectives in Italian. En: Automatic Processing of Natural-Language Electronic Texts with NooJ- 9th International Conference, NooJ 2015, Minsk, Belarus, June 11-13, 2015, Revised Selected Papers- ISBN: 978-3-319-42471-2. 1-227.
- Real Academia Española. Diccionario de la Lengua Española. Edición electrónica. <http://www.rae.es>.
- Real Academia Española. 2010. Nueva gramática de la lengua española. Manual, Madrid, Espasa Libros.

Real Academia Española. Diccionario de la lengua Española. Edición electrónica. <http://www.rae.es>

Rodrigo, Andrea (2011) Tratamiento automático de textos: el sintagma adverbial núcleo. Tesis doctoral. Facultad de Humanidades y Artes. UNR. Ediciones Juglaría, Rosario.

Rodrigo, Andrea (2013) Análisis automático de textos: el sintagma nominal núcleo. Tesis de maestría. Facultad de Humanidades y Artes. UNR. Ediciones Juglaría, Rosario.

Rodrigo, Andrea (2017), Spanish NooJ Tutorial, disponible en. http://www.nooj-association.org/index.php?option=com_k2&view=item&id=116:spanish-video-tutorial&Itemid=616, última consulta: mayo 2018

Rodrigo, Andrea, Reyes, Silvia y Bonino, Rodolfo (2018) Some Aspects Concerning the Automatic Treatment of Adjectives and Adverbs in Spanish: A Pedagogical Application of the NooJ Platform. En: Mbarki S., Mourchid M., Silberztein M. (eds) Formalizing Natural Languages with NooJ and Its Natural Language Processing Applications. NooJ 2017. Springer, Cham. https://doi.org/10.1007/978-3-319-73420-0_11. Communications in Computer and Information Science, volumen 811: 130-140.

Rodrigo, Andrea, Monteleone, Mario y Reyes, Silvia (2018) A Pedagogical Application of NooJ in Language Teaching: the Adjective in Spanish and Italian. En: Proceedings of the First Workshop on Linguistic Resources for Natural Language Processing, págs 47-56 Santa Fe. New Mexico, USA, August 20, 2018. Disponible en: <http://aclweb.org/anthology/W18-3807>.

Silberztein, Max (2015) La formalization des langues, l'approche de NooJ. Iste Ediciones, London.

Silberztein, Max (2016) Formalizing natural languages: The NooJ approach. Iste Ediciones. London.

Solana, Zulema; Méndez, Bárbara y Rodrigo, Andrea (2011) Extracción del Sintagma Nominal, Revista Infosur N° 5, ISSN 1851-1956, Octubre de 2011, disponible en: <http://www.infosurrevista.com.ar/biblioteca/INFOSUR-Nro5-2011-Solana-Rodrigo-Mendez.pdf>, última consulta: mayo 2018

Trappe, Tonya y Tullis, Graham (2005) Intelligent Business. Coursebook Intermediate. Pearson Education Limited. Harlow.

Plataforma NooJ disponible en: <http://nooj-association.org>

Manual NooJ disponible en: <http://nooj-association.org>

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

Módulo Español (Barcelona) disponible en:
http://www.nooj-association.org/index.php?option=com_k2&view=item&id=5:spanish-module-uab&Itemid=611

Rodrigo, Andrea. Tutorial para el uso de NooJ en español 2016. Disponible en:
https://www.youtube.com/watch?v=5tejAsm1ti8&list=PL3Uihe6glBR9vyp7a0PV2LWB_gFlkXw

Software NooJ available at: <http://nooj-association.org>

Manual NooJ available at: <http://nooj-association.org>

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

Aprendo con NooJ: Facebook page disponible en: <https://www.facebook.com/indagacioneslinguisticas>

Learning with NooJ: Facebook page disponible en:
https://www.facebook.com/computationallinguisticresearch/?ref=py_c

Anexo I TEXTOS PERIODÍSTICOS

Autor: José Pablo Feinmann

Diario: Página 12

Fecha: 17 de septiembre de 2017

Dorrego era amigo de "los de abajo", tenía arraigo popular y buenas relaciones con las provincias. Esto lo condenó. También las famosas cartas de Juan Cruz Varela y Salvador María Del Carril. ¿Por qué fueron famosas estas cartas, por qué trascendieron cuando tantas cosas se ocultan? Hace poco, en este diario, el sacerdote Domingo Bresci escribió sobre los archivos de la Iglesia durante los años de la dictadura. Pidió que se revelaran, que salieran a luz. Difícil, están bien guardados. La Iglesia participa del poder de la Argentina y fue cómplice de la dictadura. ¿Por qué ese poder -que escribe la historia a su antojo, el de los vencedores- no salvó a Del Carril y Varela, que eran figuras de la "civilización"? Para salvar al valiente y glorioso general Lavalle, que fue el "otro" San Martín, el del frente interno, el de las guerras civiles, el soldado de la ilustrada Buenos Aires. El que desenvainó su espada para derramar sangre de hermanos. La de Dorrego, en primer lugar y tal como se lo pidieron sus amigos unitarios. Juan Cruz Varela, consciente de la crueldad que le pedía a Lavalle, escribió: "Cartas como éstas, se rompen". Pero no se rompieron. El fusilamiento de Dorrego fue un crimen imperdonable, había que entregar a alguien para salvar al evidente ejecutor, Lavalle. Se decidió contar la historia publicando las cartas de un poeta extraviado y de un áspero, duro unitario.

Rosas utiliza la muerte de Dorrego para asumir su primer mandato. Asume para vengar esa muerte y restablecer las leyes. Es, entonces, el Restaurador de las Leyes. Será, también, un estanciero keynesiano. Se lleva bien con los gauchos, los indios y los negros. Tuvo que hacerse gaucho para poder convocarlos. Será, para Sarmiento, el más gaucho de los gauchos, el más bárbaro de los bárbaros. Con sólo masticar el pasto de cualquier lugar de la pampa sabrá dónde se encuentra.

Después de fusilar a Dorrego, Lavalle envía a dos de sus coroneles a rastrillar la frontera sur. Son Federico Rauch y Ramón Estomba. Entre gauchos e indios matan dos mil hombres. Los atan a los cañones y ordenan hacer fuego. Estomba va enloqueciendo. Rauch es atrapado por los guerreros del jefe indio Arbolito. Le cortan la cabeza y la arrojan sobre Buenos Aires.

Rosas se presenta como "el vencedor de la anarquía". Los jóvenes del Salón Literario -Alberdi, sobre todo- lo ven como la etapa del Orden, que sigue a Mayo, la de la Espada, y concluye con la Inteligencia, que representan Alberdi y sus amigos.

Rosas asume su segundo gobierno como el vengador de Facundo Quiroga. Antes, en 1834, su mujer, Encarnación Ezcurra, arma una revolución, la de los Restauradores, para imponer a Rosas como jefe del país. Encarnación muere joven, como Eva Perón, quien, como ella, será parte del movimiento del 17 de octubre, que gana el gobierno para Perón. La mujer bravía que lucha por el poder de su hombre persiste de Encarnación a Eva. A nuestra historia le gustan las simetrías.

Con Rosas llegan al gobierno los saladeristas de la pampa húmeda. La burguesía mercantil de Buenos Aires lo acepta porque quiere orden y no anarquía, que es perjudicial para los negocios. Las Provincias lo ven como un federal, una persistencia de Dorrego. Rosas exige la suma del poder público y la representación de las relaciones exteriores.

Alberdi (y, en menor grado, los jóvenes románticos, Echeverría, Sastre, Gutiérrez etc) le ofrece su apoyo. Hay aquí una importante persistencia que trama una necesaria guía de estudio: la relación de los intelectuales con el poder. Moreno y Castelli tenían el gobierno y eran sus propios asesores porque eran cultos. Luego, los asesores de Lavalle: Agüero, Varela, Del Carril. Después Alberdi y Rosas, lo que no fue. Rosas confió en Pedro de Angelis. Al fin del siglo XIX, la generación del 80 y Roca. En el siglo XX, Lugones y Uriburu. Forja y Perón. "Sur" y la Libertadora. Hasta los montoneros y el tercer Perón.

Todos estos proyectos no funcionaron. Durante los días que corren tenemos a Alejandro Rozitchner y Macri, que vendrían a ser una persistencia de la dialéctica entre Alberdi y Rosas. ¿Quién dijo que la historia no progresa?

Autor: Ariel Basile

Diario: Ámbito Financiero

Fecha: 19 de septiembre de 2017

Insisten provincias en que sigue sin encarrilarse plan ferroviario

RECLAMAN OBRAS AL GOBIERNO NACIONAL Y QUE RAMALES NO CIERREN

Tras conocerse Presupuesto 2018, en el sur piden por el Tren del Valle y el Tren Patagónico. Chaco reclamó suspensión de decreto de Macri que habilita a levantar vías. Tren a Vaca Muerta, sin arrancar.

Ampliación. En la Patagonia piden que se extienda el servicio en una serie de trayectos, y que se renueven vías en tramos donde las instalaciones no están en buenas condiciones.

La discusión del Presupuesto Nacional y diversos anuncios de la Casa Rosada pusieron en alerta a provincias y municipios, que buscan obras para mejorar el transporte ferroviario y salvar ramales que podrían discontinuarse.

La luz amarilla se encendió primero en agosto, cuando el presidente Mauricio Macri delegó por decreto en su ministro de Transporte, Guillermo Dietrich, la facultad de clausurar ramales ferroviarios en forma definitiva y proceder al levantamiento de las vías y demás instalaciones ferroviarias. Se trató, en principio, de una orden burocrática en relación a instalaciones obsoletas.

En las provincias temen que ese escenario lleve a que se frenen servicios que no se consideran prioritarios para el gobierno nacional. Y una reacción se vio ayer en Chaco, donde el defensor del Pueblo de la provincia, Gustavo Corregido, solicitó a la Justicia Federal la inconstitucionalidad del decreto de Macri. Corregido menciona la Ley 27.132 que declara al transporte en ferrocarril, tanto de pasajeros como de carga, como de interés público nacional. Además, pide con una cautelar que no se levanten vías en Chaco.

También hubo preocupación en la Patagonia, con un abrazo simbólico al Tren del Valle en diferentes estaciones de Neuquén y de Río Negro. El pedido fue doble, por un lado prevenir la suspensión del servicio en cualquiera instancia del recorrido.

Pero, además, en ambas provincias patagónicas reclaman la ampliación del trayecto. Esas obras estaban contempladas en el Presupuesto Nacional de este año con una partida de unos \$1.600 millones. En abril el jefe de Gabinete, Marcos Peña, sostuvo que la inversión "no era estratégica". En una misiva, subrayó que la prioridad estaba puesta en "los flujos masivos de pasajeros, como en el Área Metropolitana de Buenos Aires, y a los flujos masivos de cargas de más de 500 kilómetros".

El Presupuesto 2018 que ya está en el Congreso mantiene esa lógica. Allí se menciona el Plan Ferroviario de Cargas, donde "se encuentran en renovación 1.500 km de la línea Belgrano Cargas" que va del centro al norte del país. Y, en cuanto a transporte urbano, se pone el foco en aumentar la cantidad de metrobuses y en "la Red de Expresos Regionales y el Soterramiento del FFCC Sarmiento", obras orientadas al área metropolitana de Buenos Aires.

En Río Negro, también tienen pedidos para la Casa Rosada. En primer lugar, "rescatar el tramo que va de Bahía Blanca hasta Carmen de Patagones", como indicaron desde la provincia a este medio. A su vez, el Tren Patagónico, propiedad del Estado provincial, que une Viedma y San Carlos de Bariloche con una frecuencia semanal, también requiere mejoras y hay un pedido de Río Negro a Nación de hace dos años para renovar vías. La solicitud llegó a la Balcarce 50 sobre el final del mandato de Cristina de Kirchner y hasta ahora no hubo respuesta. En despachos rionegrinos destacan que el servicio se sustenta en fondos exclusivamente provinciales.

Otra de las obras que espera ejecutarse es el tren que unirá Vaca Muerta con el puerto de Bahía Blanca. La inversión para llegar hasta la localidad de Añelo, cabeza del megayacimiento de hidrocarburos no convencionales, es de unos u\$s500 millones y el trazado comenzaría a ejecutarse

en 2018, aunque no figura en el Presupuesto presentado en el Congreso, si bien la financiación sería a través de un esquema de Participación Público Privada (PPP).

También en la bonaerense Tandil, el intendente Miguel Lunghi (Cambiamos) pidió al Ministerio de Transporte la reactivación del tren que unía esa localidad con la terminal porteña de Constitución. La cartera de Dietrich tomó la responsabilidad de Ferrobaires, empresa que hacía ese trayecto, suspendido en rigor en junio de 2016 tras un accidente bajo la promesa de un diagnóstico que permitiera mejorar la calidad del servicio en el corto plazo. Quince meses después, sigue el suspenso.

Autor: Editorial

Diario: La Nación

Fecha: 19 de septiembre de 2017

Gravar la renta financiera: una medida contraproducente

UN IMPUESTO SOBRE LOS INTERESES DE PLAZOS FIJOS, BONOS Y OTROS ACTIVOS FINANCIEROS DE PERSONAS FÍSICAS ALEJARÍA AHORRISTAS Y ENCARECERÍA EL CRÉDITO

Aunque no hay todavía detalles sobre la iniciativa oficial, ha trascendido que el proyecto de reforma tributaria que enviaría el Poder Ejecutivo al Congreso incluiría un impuesto a la renta financiera de las personas físicas. Las empresas ya están alcanzadas por este impuesto. Se trata de una iniciativa que, por su impacto, podría ser finalmente perjudicial para el propio fisco, aunque su principal inconveniente debe observarse en su efecto sobre el ahorro y la inversión. La incorporación de este nuevo alcance del impuesto a las ganancias generaría mayor desconfianza en todo inversor.

Un gravamen sobre los intereses de los depósitos a plazo fijo, los bonos y otros activos financieros, en lugar de provocar un efecto redistributivo, necesariamente afectará resortes de la economía que son esenciales para el financiamiento del sector productivo y del Estado.

Las declamaciones populistas suelen equivocadamente identificar al sector financiero con la imagen de millonarios sentados sobre una montaña de dinero que han acumulado sin esfuerzo. A partir de allí afirman que un impuesto sobre la renta financiera recaerá justicieramente sobre ricos en beneficio del resto de la población. Se trata, sin embargo, de un grosero error. Lo concreto es que el tan vilipendiado sistema financiero desempeña un papel esencial en cualquier economía moderna. Es el instrumento para canalizar los ahorros de las personas y las empresas hacia la inversión productiva. Sin esos ahorros muy difícilmente podría el sistema financiero, formado por los bancos públicos y privados, prestarles dinero a otros particulares para comprar viviendas, automóviles o artefactos electrodomésticos. Y tampoco podría financiar proyectos comerciales o industriales de las empresas.

De incorporarse un impuesto sobre los intereses que reciba un particular por un depósito a plazo fijo, es altamente probable que ese ahorrista exija una tasa de interés más elevada que compense el nuevo tributo que deberá abonar. Y es obvio que los bancos que reciban esos depósitos trasladarán la mayor tasa a los créditos que otorguen. Así, habrá menor demanda de crédito, menos oportunidades para inversiones productivas y, desde luego, menor creación de empleo.

En la Argentina de hoy, todavía sometida a inflación, los efectos del proyectado impuesto serían incluso más negativos, por cuanto las tasas de interés en pesos contienen un componente que compensa la desvalorización de la moneda y que no es propiamente renta. El Estado, al cobrar un impuesto sobre los intereses nominales, se quedaría directamente con una parte del capital y no sólo de la ganancia.

Algo similar ocurriría en caso de gravarse la diferencia entre el precio de compra y el de venta de otros activos financieros si hubieran sido negociados en pesos sujetos a inflación. Volvemos, de este modo, sobre un problema por demás conocido, como el de muchas sociedades que pagan el impuesto a las ganancias sobre rentas ficticias, al no poder aplicar ningún ajuste por inflación.

Respecto de la imposición de la renta obtenida con los títulos públicos, cabe remitirse al criterio sentado, ya en 1974, por el especialista Alberto Tomás López, quien expresó: "Cuando se sostiene que la renta de títulos públicos debe ser gravada igual que la renta de cualquier papel privado, puede

descubrirse que, sin reajuste del capital, sin alta tasa de interés y muchas veces sin exención de impuestos para el capital y la renta, las emisiones públicas resultan de difícil colocación. Esto, desde luego, es bien perceptible en un régimen de alta inflación".

Por citar un ejemplo, si para suscribir un bono del Estado sujeto al impuesto a la renta los inversores no encontrasen atractiva una tasa de interés del 15% y quisieran el 20%, el emisor del título tendría dos alternativas. En la primera, podría dar el 20% sin exención del impuesto, con lo cual el Estado pagaría el 20% y cobraría -imaginando un impuesto a la renta del 30%- el 6%, quedando una tasa o desembolso neto del 14%. La segunda opción sería fijar el 15% libre de impuestos, con lo cual el Estado desembolsaría prácticamente lo mismo y se ahorraría la tarea de cobrar el impuesto, que le costará seguramente más de un 1% sobre el capital servido.

En conclusión, un gravamen a la renta financiera sobre personas físicas alejaría ahorristas y encarecería el crédito. Son dos razones para archivar definitivamente este proyecto.

Autor: Elizabeth Gómez Alcorta

Diario: Página 12

Fecha: 19 de septiembre de 2017

LA DISPUTA POR LA VERDAD

Cuarenta y nueve días de la desaparición forzada de Santiago Maldonado representan cuarenta y nueve días de disputar la verdad de los hechos.

Los cambios de posición y de discursos esgrimidos por el gobierno, que se manifiestan a través de diferentes funcionarios y son amplificados por los principales medios de comunicación, muestran que la verdad está en un claro terreno de disputa.

Podemos hacer un breve repaso que revele esta afirmación: i) Santiago Maldonado no estaba en la lof en Resistencia Depto. de Cushmanen; ii) habría estado comprando algo en una estación de servicio en Entre Ríos, y fue visto por un camionero; iii) habría estado en Mendoza; iv) habría muerto por lo menos una semana antes del 1 de agosto, en virtud de un cuchillazo que le dio el puestero; v) -que se deriva de la anterior- Santiago habría sido parte de la RAM; vi) habría estado en Tierra del Fuego; vii) el operativo represivo contó con la orden del juez Otranto; viii) el operativo ingresó al territorio comunitario en razón de las piedras que recibieron los gendarmes en la Ruta 40; ix) ingresaron solo siete gendarmes; x) los gendarmes no estaban armados; xi) los gendarmes no se acercaron al río; xii) no hubo ninguna orden ni comunicación con responsables políticos en aquel momento; xiii) El jefe de gabinete del ministerio de Seguridad, Pablo Noceti, pasaba de casualidad por la zona y solo paró a saludar a los gendarmes.

Con el correr de los días, las afirmaciones hechas ante los medios, la Cámara de Senadores o los organismos de derechos humanos se fueron cayendo una a una. Y tras cada caída, se fue levantando otra perversa enunciación.

Santiago no estaba en Entre Ríos, a pesar de que exista un pueblo en el que todos se le parecen; y tampoco estaba en ninguno de los otros lugares en los que se los quiso situar; estaba donde se dijo desde el primer día. Resultó difícil que los miembros de la comunidad prestaran declaración testimonial frente a la amenaza pública -que además tuvo su palmaria evidencia de cumplimiento- de Noceti, relativa a que se estaba investigando la identidad de todos los miembros de la comunidad en pos de meterlos presos (averiguar qué juez y en qué causa se los está investigando es materia pendiente aún); sin embargo, ello se logró.

Por otro lado, una importante cantidad de periodistas comprometidos tomaron contacto con personas que habían estado con Santiago los días previos a su desaparición o que lo vieron el día de los hechos en la comunidad. También aparecieron videos e imágenes que lo ubican en el lugar de la represión. Finalmente, el ADN del cuchillo del puestero dio negativo y, así, cayó la coartada de que el joven no estaba en la lof el día del operativo represivo.

Ya sabemos que se lavaron las camionetas, que eso dice -aparentemente- el reglamento y que no hay que tirar por la ventana a ningún gendarme. El Ministerio de Seguridad entregó las listas con la nómina de los supuestos agentes que estuvieron presentes aquel día, el 8 y el 15 de agosto; pero, a partir de las declaraciones testimoniales de los gendarmes, parece que en ellas no están todos aquellos que participaron. Luego, la versión de que solo ingresaron siete de ellos, también se derrumbó y se transformaron en 15 o 20, los que se habrían acercado al río, arrojado piedras y disparado balas. Incluso, según la declaración de una enfermera que atendió al agente herido en la cara por una piedra, quien afirmó que lo atendió adentro de la comunidad y no en la ruta, debemos declinar también la razón que habilitó la actuación por flagrancia -ya que en el camino también se había derrumbado la versión de la existencia de la orden judicial-.

Entonces, Santiago estaba donde se dijo que había estado, corrió hacia el río, y, en principio pudo haber sido herido por gendarmes que ingresaron al territorio comunitario sin orden judicial y sin razón legal que lo autorizara. Hasta aquí, los datos que se han ido obteniendo.

Fue necesario un urgente cambio de estrategia discursiva frente al conjunto de revelaciones, sumado a que el gobierno contó con antelación con el resultado de las pericias sobre los setenta teléfonos secuestrados de los agentes de la fuerza de seguridad y pudo conocer algunos intercambios que referían a que una piedra fue lanzada por un gendarme hacia una persona en el operativo; lo que devino en la construcción de la versión del exceso del gendarme solitario. El ministro de Justicia comunicó que los gendarmes pudieron haber mentido.

No obstante, la presencia de Noceti en el lugar no fue casual, como lo intentó justificar Bullrich desde el primer momento. El abogado, que reivindica la dictadura cívico-militar, había viajado a la Patagonia a fin de coordinar las fuerzas represivas provinciales de Río Negro, Chubut y las federales, para llevar adelante operativos contra comunidades mapuches, en particular contra la Lof de Cushamen. En 2016, había tomado contacto con funcionarios judiciales y del ministerio público fiscal, federal y provinciales, para que calificasen los hechos que tuvieran como protagonistas a mapuches como federales y les indicó que se debía utilizar las leyes antiterroristas. Noceti dijo que no habló previo al operativo represivo del 1 de agosto con los responsables de gendarmería que estaban en el procedimiento. Mintió. Nuevamente nos encontramos con prueba que afirma que Noceti habló por teléfono previo al ingreso a la comunidad, durante y también luego, tanto con el Jefe del Escuadrón 35 de El Bolsón, como con el subjefe del Escuadrón 36 de Esquel. Aún más, el 3 de agosto, a 48 horas de la desaparición de Santiago, Noceti informó al responsable de gendarmería de Esquel que se iban a realizar pericias a los vehículos utilizados en el operativo. Esto, claro está, antes de que se realicen las pesquisas sobre los vehículos que fueron lavados.

Gendarmería ingresó a la lof sin orden judicial, sin que exista hecho alguno que habilite la utilización del procedimiento de flagrancia y después de que sus responsables hablaran telefónicamente con Pablo Noceti. Ingresaron por lo menos 20 agentes, demoraron a mujeres y niños, prendieron fuego objetos, secuestraron herramientas de trabajo comunitarias y persiguieron a hombres que se dirigían hacia el río. Les dispararon. Santiago estaba allí. Ingresó un Unimog hasta la zona cercana al río: el Unimog no ha sido peritado aún, pero fue corroborado su tránsito en el primer rastillaje realizado por las huellas que dejó.

Hasta aquí solo presento un relato y un contra-relato a base de pruebas que se sostienen y otras que se desploman.

La disputa por la verdad hace que frente a lo evidente se eche mano a la más antigua de las herramientas del poder político: la construcción del enemigo. La RAM concentra el epicentro del enemigo político. El Gobierno habla de esa organización, relata que los están investigando, que saben quienes la integran, aseguran que los testigos forman parte de ella, incluso se refieren a la comunidad de la RAM y no a la comunidad en Resistencia Depto. Cushamen -invisibilizando su nombre y sus reivindicaciones territoriales-.

Santiago, militante de la causa indígena y en particular de la comunidad Lof en Resistencia Cushamen y de su longko, fue convertido en enemigo.

El discurso y el mensaje que esto transmite es contundente: se niega al ser humano y se elimina toda posibilidad de empatía, es decir, de colocarse en el lugar y la perspectiva del otro y comprenderlo.

Las hipótesis de que Santiago está vivo y escondido o que fue asesinado por la comunidad mapuche y que se mantiene en secreto por la oscura finalidad de desestabilizar al gobierno -únicas que explicarían que no lo haya matado y desaparecido la Gendarmería- son tan burdas que ni siquiera el peor de los mercenarios llamados periodistas se ha animado a esgrimir.

A pesar de ello, en la época de la posverdad lo que importa es mediar en la emocionalidad de la gente, aun al costo de banalizar y pervertir la realidad y también el mal que ella puede encarnar -capaz de desaparecer personas-.

De ese modo, el hilo que se intenta presentar es que el kirchnerismo está atrás del caso de Santiago. Por lo tanto, todo aquel que esté del otro lado de la grieta debe dudar de la desaparición forzada de él. No importa que el joven fuese anarquista o que la agenda de los pueblos indígenas no entrara en los 12 años de la gestión del anterior gobierno -Facundo Jones Huala fue investigado de modo ilegal por más de un año por un agente de la AFI en 2014 y 2015-. Alcanza con que el reclamo masivo por la aparición con vida lo hagan los organismos de derechos humanos y que dirigentes políticos de diferentes fuerzas -incluida las del FPV- se sumen a esa demanda.

Por último, quiero hacer una breve mención a mi caso personal, que ilustra con claridad el tipo de maniobra predilecta por el gobierno de turno.

En el mundo binario de la grieta, si defendes a Milagro Sala sos kirchnerista, no defensora de derechos humanos. De allí que estos días, algún diario afirme que soy candidata a diputada nacional por el kirchnerismo, lo que no solo es falso porque soy militante de Patria Grande, sino que es un error muy fácil de chequear -suponiendo que pudiera serlo-. Además, se afirmó que soy abogada de Quebracho; también es falso. Se aseveró que tomé declaraciones a los miembros de la comunidad encapuchados y que me ofrecí como vínculo entre los testigos y el juzgado por el caso de Santiago, también todo ello es falaz. Aunque fuera kirchnerista y defensora de todos sus referentes y dirigentes nada cambiaría la verdad de los hechos en el caso de Santiago; pero, como tampoco importa que le pasó a él, menos va a importar una mentira más en relación a todo lo que lo rodea.

La desaparición forzada es un delito que se configura por medio de dos acciones diferentes que deben estar en cabeza de un funcionario público -o de una persona que actúa con la autorización o el apoyo del Estado-: a) debe privar de la libertad a una persona y b) luego de ello, se debe omitir dar la información o negarse a reconocer dicha privación de libertad o de informar sobre su paradero.

Sobre la segunda de las acciones habla este artículo. Sobre una serie de discursos esgrimidos por el Gobierno que no solo no guardan relación con la verdad de lo sucedido, sino que entorpecen su averiguación, mientras la desaparición forzada de Santiago continúa cometiéndose.

Anexo II TEXTOS DE LOS JÓVENES

La educación es fundamental para el desarrollo de la persona

Para mí la educación es un derecho individual y un deber del estado hacerla llegar a toda su población. La educación en Argentina la considero muy buena, ya que le da la posibilidad a personas con bajos recursos interiorizarse, tener una opinión, una postura y una visión sobre el entorno que nos rodea. No debemos desperdiciar la oportunidad que nos da nuestro país de poder estudiar gratuitamente. Me parece repugnante que vengan extranjeros a usurpar cupos en universidades Argentinas, si su país no les brinda una educación pública gratuita por qué nosotros sí?

Bueno, educación es un término muy amplio. Yo creo que la educación es el pilar para construir una sociedad sana y autosustentable. En Argentina tenemos la posibilidad de educación pública, a la que considero entre las mejores instituciones de educación aunque realmente se encuentran obsoletas en varias áreas y con métodos de evaluación en el mismo estado.

Considero que la educación es un arma de liberación y transformación no solo personal sino también social. Por eso, creo que es importante que no responda a la lógica del mercado. Con respecto a la relación entre docente y estudiante, esta no debe ser vertical. La educación es un intercambio de experiencias y saberes entre iguales. El diálogo y el respeto son esenciales para que la educación se convierta en una herramienta revolucionaria. Actualmente, ante los intentos del neoliberalismo por desprestigiar la escuela pública, se vuelve esencial defender la educación como una trinchera.

Creo que educación es la base de toda sociedad. Más allá del contenido intelectual es el medio a través del cual se trasmite la cultura, valores y formas de actuar.

Creo que el sistema está bastante atrasado actualmente. Un chico sentado durante 6 horas en un banco se aburre. La idea principal no debe ser solo pasar el conocimiento del maestro al alumno, tenemos que pensar en las personas como individuos, con todo lo que eso acarrea. Por otro lado, al menos en Argentina, creo que hay muchos maestros a los cuales no les interesan los alumnos como personas, sólo "hacer su trabajo".

Pienso que la educación es la herramienta que nos permite ser mejores personas día a día. Nos forma como seres humanos. Gracias a la educación podemos decidir que hacer con nuestras vidas y expresarnos de distintas maneras. Nos permite relacionarnos con el mundo.

Bien utilizada, es la mejor herramienta para cambiar este mundo

La educación consiste en una serie de herramientas cuya finalidad es la libertad del humano, y que además debe estar al alcance de todos.

Es necesario mejorarla, sobre todo el nivel primario. Faltan escuelas, talleres, escuelas con comedores, ya que hay chicos que no llegan a dar el contenido básico por problemas familiares o económicos, y muchos quedan afuera por falta de cupos

La educación pasa por una nueva crisis, producto de estar inmersa en una sociedad enferma, lastimada, insensible y egoísta, la educación es el cajón para salir de esta situación, pero necesitamos

más personas comprometidas, no solo docentes con vocación sino padres comprometidos con la vida, con la libertad, con la sociedad, herramientas tecnológicas son indispensables para lograr captar a jóvenes estimulados desde su nacimiento

La educación facilita el aprendizaje, es indispensable en cualquier sociedad. Es la transmisión de conocimientos, cultura y vivencias de una persona a otra.

La educación es la base de toda sociedad. Debería prestarse mucha más atención a las cosas que se enseñan y cómo.

La educación es la base de una sociedad, junto con salud y trabajo. Es de los más importante para que los jóvenes puedan formarse de la mejor manera. En algunos ámbitos es muy buena en nuestro país, pero no es de lo mejor.

La educación es un camino que se recorre propone una búsqueda. Caminan con otro y colabora con tus sueños y lo de otros.

Un bien necesario. Útil en muchos sentido. Muy pocos poseen educación.

Pienso que la educación es un proceso en los individuos

La educación tiene que ser más exigida y tener un seguimiento sobre la persona ya que muchos no terminan el secundario y menos hablemos de un terciario.

La educación cambia a las personas que van a cambiar el mundo como dice el gran pedagogo Freire. Es una herramienta estructural de la sociedad para poder mejorar nuestra calidad de Vida tanto individual como de manera colectiva por eso creo trascendental el apoyo a las docentes en su pedido de materiales para su utilización pedagógica.

Pienso que debería haber alguna entidad que regule la enseñanza de las materias, para que el todas las instituciones públicas se den de manera responsable. Si esa entidad ya existe debería comenzar a cumplir el rol primeramente de respaldar el trabajo docente y segundo respetar el derecho de que todos tengamos una educación equitativa y completa.

Pienso que la educación es la mayor herramienta, tanto para el progreso individual como social. Desde la educación se forman personas con ideales, capaces de defender lo que creen correcto, profesionales, etc.

Está mal enseñada, y hay muchas de las cosas que hoy en día nadie sabe cómo contestarla o que son, para que sirven, etc. Me gustaría que la educación de ahora haga que todos puedan aprender como son las cosas y que sean con la verdad y no con la MENTIRA. Porque la MENTIRA tiene patas cortas.

Creo que falta mucho avance aún, que se necesitan docentes que puedan evolucionar y pensar en el alumno no como una persona a la que hay que llenar se información, sino que se debe llenar de conocimiento. Pero para que pueda ser incorporado realmente, creo que el docente también debe comenzar a buscar otras alternativas en las estrategias dentro del aula. Más palabras de aliento hacia el alumno y acompañarlo en el recorrido. Creo que la educación es la base fundamental de una sociedad y que a partir de ella podemos lograr grandes cosas.

Pienso que la educación es la base fundamental de una persona, pero que lamentablemente está muy "manoseada" siendo que tendría que ser el eje principal de cualquier sociedad.

La educación es una de las bases de nuestra cultura para que podamos ser personas pensantes reflexivas autocríticas.

La educación es la base con la que deberíamos contar todos, para una mejor relación y sociedad

La educación es la base fundamental por la que tendría que atravesar toda persona.

La educación es o debería ser esa herramienta que nos permita soñar siempre. Y cumplir esos sueños!!!

Desde mi punto de vista creo que la educación es fundamental para la formación de un individuo. Aunque tenga algunas falencias es muy importante seguir apostando a la educación y mejorarla cada día un poco más.

Pienso que hay una caracterización particular dependiendo de factores como época, contexto socio cultural y por supuesto económico también que influyen directamente e indirectamente a "tirar" hacia determinado lado. Personalmente trabajo en una escuela primaria donde concurre gente completamente marginada en cuanto a lo económico mayormente, y es diferente el concepto educación que puede llegar a encontrarse allí ya que en su gran mayoría, se hace un trabajo más asistencial que educativo por las necesidades sociales que tienen los chicos. Por otro lado estudiando en un terciario también puedo dar cuenta de la educación (refiriéndome a enseñanza) recibida allí que en algunos casos es excelente y en otros buena o regular. Educación es un término amplio que según en que contexto se use, podemos caracterizarla de determinada manera.

Es muy importante para el desarrollo intelectual y psicológico de las personas. La educación nos da acceso a ciertos conocimientos. A Ser personas de bien. La educación nos saca de la ignorancia, pobreza, miseria, corrupción, injusticia en que se vive hoy en día.

Que hay Que hacer varios cambio atendiendo a las necesidades socio económica a de esta época

Es necesario reformular la curricula y proveer de recursos para que culminen la modalidad sabiendo. Incluir no es solo lograr retenerlos en las aulas es darles las herramientas para abrirse camino en la vida.

Yo creo que la educación es una de las herramientas más importante en la vida de cualquier persona. O por lo menos para mí lo es. Leer y estudiar abrió mi cabeza, me permitió formular mis propios argumentos y gracias a la educación que recibí hoy soy capaz de defender lo que pienso fundamentando, argumentando y explicando.

Pienso que su naturaleza es la de lo móvil, es decir, pienso que la educación es movimiento.

La educación en nuestras instituciones forma parte de un sistema en donde se prepara al niño u adolescente y adulto a competir en el mundo. Se enseña que uno siempre es peor que el otro y que otro es mejor que uno. Los contenidos no importan, importa aprobar o tener una calificación alta. No se enseña a pensar y a tener un pensamiento crítico, en vez de eso se enseña a repetir y a estudiar de memoria, logrando que los contenidos no se afiancen y que se olviden después de un tiempo.

Si el modelo de educación sigue siendo el tradicional hoy en día resulta arcaico y los aprendizajes no se vuelven significativos para vida cotidiana en la sociedad actual.

Pienso que educar es enseñar instrumentos, facilitar el uso de herramientas para que personas puedan expresarse de una mejor manera, sin poner barreras o juicios de valor que puedan limitar tal aprendizaje

Hoy en día se puede reconocer al conocimiento como una mercancía más. Esto es posible debido a la forma en la cual se construye, se hace circular y se consume, donde cada una de éstas etapas se considera como un posible negocio.

Considero a la educación como un pilar fundamental para el desarrollo de cada país, así como también para la formación de cada individuo. La educación nos brinda herramientas para lograr un mejor entendimiento del mundo y así desarrollarnos activamente como miembros responsables de una sociedad. También nos permite desarrollar el pensamiento crítico, lo cual nos hace libres, nos da la posibilidad de evaluar situaciones y elegir qué caminos nos conviene seguir. Lamentablemente, veo muchas carencias en la educación de numerosos países del mundo y estoy convencida de que esa es una de las principales razones de los problemas que tienen las sociedades.

La educación permite expandir el conocimiento de las personas, y a su vez, con una buena base, permite hacer a las personas más concientes de lo que pasa. Da el pie para la reflexión, que es necesaria para un pensamiento crítico.

Me parece la herramienta básica para que los integrantes de una sociedad podamos tener un futuro digno, es la verdadera revolución social

DESACTUALIZADA. la educación actual no está a la altura de las nuevas necesidades del ámbito laboral y social. Creo que se deben reasignar las prioridades de cada asignatura y especializar más los planes de estudio

Creo que los planes de estudio están bien encarados, que muchas veces los contenidos no llegan al alumno en tiempo y forma por diferentes razones (paros, licencias interminables etc.)

Creo que la educación hoy en día esta con problema en la adaptación con nuevas tecnologías y nuevos métodos de enseñanza, en el cual se nota una clara falta de adaptación docente, siguiendo prácticas muy tradicionales en un tiempo en el cual es cada vez más difícil llevar a los chicos a los colegios y captar su atención en un ámbito en el cual las preocupaciones económicas son muy destacables en la deserción escolar. Saliendo de este planteo opino, que la educación es el instrumento más poderoso que posee la humanidad para un porvenir más brillante.

Pienso que la educación, se fue deteriorando a lo largo de los años por falta de un plan propio de los distintos gobiernos argentinos que fueron copiando planes de estudio y modalidades que en otros países ya habían fallado

En el ámbito público hay una decadencia, desde la tendencia de los padres a cuestionar la autoridad del docente, en el ámbito privado hay una mejor situación no solo por la mayor exigencia sino por un mayor interés de los padres en respetar al docente y su profesionalidad.

Creo que es el arma fundamental para sacar de raíz los grandes males en la sociedad como la delincuencia y la droga adicción. Gente educada y con trabajo tiene mayores posibilidades y menos necesidad de caer en estos mundos

Es un proceso fundamental para el desarrollo de todos los miembros de la sociedad pero hoy en día se encuentran varias fallas sistemáticas, ya sean institucionales como profesionales (educador). Creo que debería haber más exigencia hacia los educando y, también al docente. Las nuevas tecnologías se deben usar para poder llegar, de mejor forma, a los alumnos del siglo XXI.

Creo que la educación es lo primero que se necesita para llevar adelante a un país. Con educación y trabajo digno todo es posible. Soy psicopedagoga, recién recibida y lo que veo es que la educación en la Argentina está muy pobre, y con calidad bajísima, tanto en la escuela pública como privada. La educación permite que uno piense por cuenta propia, que razone, la educación abre las puertas al mundo.

Personalmente, considero que la Educación es la base de toda sociedad, ya que a través de ella se pueden aprender e incorporar conocimientos, valores y aptitudes para ser cada día mejores ciudadanos. Hoy en día en nuestro país, la Educación no está pasando por su mejor momento y creo que tanto los gobernantes, los maestros, los padres y cada uno de nosotros debemos colaborar para mejorar cada día la Educación de las futuras generaciones

Pienso que no está enfocada a lo que realmente necesitamos los ciudadanos para la vida y que la escuela pública está con un nivel muy bajo debido a los docentes sin vocación y al gran ausentismo docente

Que usa un método anticuado. La educación de hoy en día debería preocuparse por desarrollar individualmente las actividades en las que cada niño se destaca. El hecho de evaluar a todos por igual, como es hoy en día y como fue siempre, es un método que está quedando obsoleto

Creo que en la educación hace falta incluir el uso de nuevas tecnologías, sobretodo en los niveles iniciales, dado que los niños del siglo xxi son nativos digitales, aprenden de otra forma, distinta al modelo vigente.

«Learning with NooJ»

From Computational Linguistics to Language Teaching

Andrea Rodrigo and Rodolfo Bonino

Author Information

Andrea Rodrigo: PhD in Humanities and Arts with orientation in Linguistics, Facultad de Humanidades y Artes, Universidad Nacional de Rosario. Member of the Academic Committee of Specialization for Teaching Spanish as a Second Language, Facultad de Humanidades y Artes. Universidad Nacional de Rosario. Teacher at the Universidad Nacional de Rosario and the Instituto Superior N° 28 «Olga Cossettini».

Rodolfo Bonino: PhD in Humanities and Arts with orientation in Linguistics, Facultad de Humanidades y Artes. Universidad Nacional de Rosario. Teacher at the Instituto Superior N° 28 «Olga Cossettini».

Collaborators

Research Team led by Andrea Rodrigo (2015-2017)

Silvia Reyes

Luciana Andrín

Alejandra Rodríguez

María Andrea Fernández Gallino

Paula Alonso

Soledad Harriet

Participation of the students:

Malen D'Urso (Instituto Superior del Profesorado N° 28, «Olga Cossettini»)

Camila Ferramondo (Facultad de Humanidades y Artes. Universidad Nacional de Rosario).

Scientific Advisory Committee

Dr. Max Silberztein, Université de Franche-Comté, France

Dr. Xavier Blanco, Universidad Autónoma de Barcelona, Spain

Dr. Gerardo Kahan, Universidad Nacional de Rosario, Argentina

TABLE OF CONTENTS

Learning with NooJ (English version)

Preface by Xavier Blanco Escoda 7

FIRST PART

Chapter 1

Introduction	79
1. Who we are	79
2. Book Structure	80
3. What is Computational Linguistics?	80
4. The NooJ Platform	81
5. What does NooJ do?	81
6. NooJ Dictionaries	84
7. NooJ Grammars	84

Chapter 2

The Corpora

1. Journalistic Texts	86
2. Analysis of the Corpus of Journalistic Texts	86
3. Contrast with the Texts Written by Young People: Strengths and Weaknesses. Data from <i>Aprendo con NooJ</i>	90
4. Analysis of the Texts Written by Young People	
5. Comparison Between Corpora: Journalistic Texts in Contrast to Young People Texts	92
6. Students' Opinions	92
• Personal Experience Accounts of Attendants to the « <i>Applied linguistics Workshop: Using NooJ as a Tool for the Study of Spanish Morphology</i> » (2017)	92
• Personal Experience Accounts of Attendants to the Linguistics-and-Languages-oriented Seminar "Contributions of Computational Linguistics to the Teaching of Spanish as L1 and L2", given by Silvia Reyes and Andrea Rodrigo (2017)	93

Conclusions 95

PREFACE

The book you are now holding in your hands, dear reader, written by Andrea Rodrigo y Rodolfo Bonino, is the result of many years of fruitful work. It has been nourished with the contributions of the research team led by Dr. Andrea Rodrigo, which address the linguistic engineering platform NooJ created by Professor Max Silberztein (Université de Franche-Comté).

This volume is a sort of condensation of various works. On the one hand, it is an excellent manual for the NooJ user, in Spanish and English, since the book is to a large extent bilingual. On the other hand, it is also a work of computational analysis of the Spanish language, containing numerous commented examples and the computational description of different orthographic, morphological and syntactic features of Spanish, in its Rioplatense variety, which has so much contributed to the common heritage of our language, and continues to do so. It is moreover a manual of language teaching, since it presents important theoretical as well as practical elements of the pedagogical application of NooJ to the teaching/learning of mother and foreign tongues, which even include the personal experience accounts of students after attending applied linguistics workshops based on this computer environment. It likewise is a manual of discourse analysis that also features and analyses real texts that are representative of a 'youth grammar', collected for an original and specific research. Finally, but not least important, *Aprendo con NooJ* is also a book of exercises that, based on explicitly stated goals and on a meticulously planned pedagogical progression, will allow the reader to go into the universe of computational linguistics and its applications with sure step.

This book is a live work as well, for it feeds on the permanent teaching profession of the large team of the professors and researchers who sign it, and also on the experiences that teachers and students shared on the Facebook page *Aprendo con NooJ*, on the series of videos available at YouTube, which compose the tutorial for the use of NooJ in Spanish, and on the NooJ Spanish Module Argentina available at the web page of the above mentioned computer programme.

I can therefore only celebrate and applaud the appearance of this work that will surely constitute a milestone in the spheres of computational linguistics and of language teaching in Argentina.

Xavier Blanco

Universitat Autònoma de Barcelona

FIRST PART

Chapter 1

Introduction

1- Who we are

We are a team of research professors from the *Universidad Nacional de Rosario* and the *Instituto de Enseñanza Superior* Nº 28, «Olga Cossettini». In November 2014, we presented a research project named *Software tools for language teaching: an application from a didactic approach in teacher training*, which became the chief cornerstone for the foundations of the works in this book. In 2015, this Project was accredited by the Secretary of State for Science, Technology and Innovation of the Province of Santa Fe [1].

With computational linguistics, we seek to frame our line of work within a different approach. The underlying assumption is that through metalinguistic reflection and with the aid of software, it is possible to generate an explicit knowledge of language to prove different linguistic hypotheses. From our perspective, we think that computational linguistics is not only an efficient resource to teach the grammar of a language but also an efficient tool to gather significant information about how students use written language. These last two points will enable us to elaborate teaching materials aimed at strengthening students' understanding as well as encouraging their incorporation of new knowledge.

We believe that the use of software tools in teaching a language is a truly innovative practice, since, as so far as known, there are no formal records in the Province of Santa Fe. With this work, we seek to contribute to the training of language teachers by presenting some of the multiple uses that can be made with these resources.

Our first aim is to give a general overview of computational linguistics but without so many technicalities in order to ease the way for higher education teachers and students who are new to this subject. Once they become familiar with the basics, our following aim is to show how NooJ can be used as a platform to compare texts written by students with journalistic texts in the Rioplatense variety. This gives us better diagnostic elements to elaborate didactic strategies for our assumption that there is a «grammar used by groups of young people», i.e. a 'youth grammar'. Finally, we lay out a series of simple exercises based on automatic text analysis, building dictionaries and electronic grammars for Spanish and in a smaller extent, for English. Such exercises are focused on two possibilities: that students use the available resources in NooJ's website (nooj-association.org) for the Spanish Module Argentina, or that students create their own grammars and dictionaries. For English, the exercise only uses the resources available NooJ English module.

Our central attempt is to provide an analysis tool which allows us to start from a state of situation when faced with linguistic questions such as «*what grammar do young students produce?, what grammatical knowledge have they been able to build?*». We believe that only from a diagnostic like this it is possible to project teaching materials for Spanish. A sample exercise to teach English as a foreign language is also included, but for illustrative purposes only. Our endeavour is to enrich grammatical knowledge by starting from present lexical categories or syntactic structures and help

1. Code 2060-005-14, Resolution 117/2015, March 18th, 2016 to September 18th, 2017.http://www.santafe.gov.ar/index.php/web/content/download/219348/1141506/file/Resultados*%20Convocatoria%202014%20%20Programas%20de%20Promoci%C3%B3n%20de%20las%20Actividades%20Cient%C3%ADfico-Tecnol%C3%B3gicas%20y%20de%20Innovaci%C3%B3n.pdf

to develop absent or underdeveloped ones. For instance, if nominal or verb categories or structures are more developed, the key is how to turn other less developed categories or structures into more developed ones.

Some considerations about the French nucleus verb phrase made by Gabriel Bès (1999) have served us as a starting point for the theoretical framework of this book. By considering the morphosyntactic categories that are part of the nucleus phrases, it is possible to determine their beginning, end and nucleus with very precise order constraints, since they are considered to be inseparable blocks. Thus, the definition of nucleus phrases is as follows:

- * *la casa* (nucleus noun phrase)
- * *muy hermosa* (nucleus adjective phrase)
- * *tan valientemente* (nucleus adverbial phrase)
- * *de su interés* (nucleus prepositional phrase)
- * *escuchaba* (nucleus verb phrase)

These first units are part of major units which expand up to the sentence level.

- * *la casa muy hermosa* (Noun phrase)
- * *escuchaba tan valientemente* (Verb phrase)
- * *eran de su interés* (Verb phrase)

Within natural language processing (NLP), the analysis of nucleus phrases allows us to reduce the ambiguity of the morphosyntactic categorization from the moment it is possible to concatenate internal expressions. The linguistic hypotheses about the expressions in Spanish are validated when «entering them into the computer», that is, when comparing them with the dictionaries and grammars designed in NooJ with real texts. Finally, it is important to note that in order to determine the categories noun, adjective and verb, it is necessary to register the different combinations in which such categories appear, since our conception focuses on a context-sensitive grammar.

2- Book Structure

This book aims to discuss some general issues about the use of NooJ from a Computational Linguistics approach, as well as other works about Natural Language Processing, and then to immerse ourselves in a series of exercises with different levels of complexity.

3- What is Computational Linguistics?

The relationships between Linguistics and Computer Science are varied. Theorists have suggested many different names for the interdisciplines involving these two fields of knowledge (computational linguistics, linguistic informatics, etc.). The different denominations attempt to account for the different objects of study that can be addressed from interdisciplinary perspectives and the goals pursued. The field of study is broad, versatile and dynamic, with many approaches from different –and even contradictory- viewpoints.

Our work intends to present a clear perspective of the goals to be developed together with the names that suit them better. We consider that Computational Linguistics as a discipline that involves many other fields of study whose object is to present the best formalization of the human language. These formalizations can be processed with the use of software without the implication that the automatic processes have to be analogous to the mental processes of the speakers of a language.

Although our formalizations are developed according to the resources provided by a particular computer programme, the previous existence of software is not a previous requisite for research. In fact, programmers do their work based on specific hypotheses about the formal nature of language. We are inclined to believe that Linguistics and Informatics do not have a one-way relationship. Instead, we dare say that their relationship is dynamic: linguistic hypotheses are used to develop software, software functions restrict the type of grammar that is created and, at the same time, these grammars enrich software as they place new demands for a better formalization.

4- The NooJ Platform

Although in the original project we thought of the use of computer tools «specifically, XFST and NooJ», we soon decided to put aside XFST because we only had a trial version offering us a little scope for entering texts, and focused on NooJ, which had become a platform. The following list presents the advantages that seduced us:

- ♦ NooJ is readily available online and free for academic purposes,
- ♦ NooJ is not a black box. Instead, NooJ is an empty and totally interactive box in which it is possible to create texts, dictionaries and grammars, or use the resources available in its modules — in this case the Spanish Module Argentina,
- ♦ NooJ is a platform that has been designed to be used with different languages, allowing multiple common work starting from a common base,
- ♦ it is possible to consult NooJ's creator, Dr. Max Silberstein, who is leading the NooJ project —and our project is, then, intertwined with his proposal,
- ♦ as we have access to the Spanish Module Barcelona (Spain-UAB) by kind permission of Dr. Xavier Blanco, it is possible to complete and improve the dictionaries and grammars of the Spanish Module Argentina.

NooJ enables the formalization of different linguistic levels: orthography, morphology and syntax. Using NooJ in teacher training spheres also helps students to be mindful of the need of linguistic formalization and the advantages it brings with it. In our opinion, NooJ allows us to be in the vanguard of research in Computational Linguistics.

5- What does NooJ do?

NooJ basic function is to recognize Unicode characters and identify them as letters, numbers, punctuation marks and blank spaces. This makes it the best eligible option for processing many natural languages. Any string of characters corresponding to letters that may occur among other characters with other values such as blank spaces, punctuation marks or numbers is recognized as a word form regardless of what these characters are. This function allows to provide statistical information even without having any specific linguistic information:

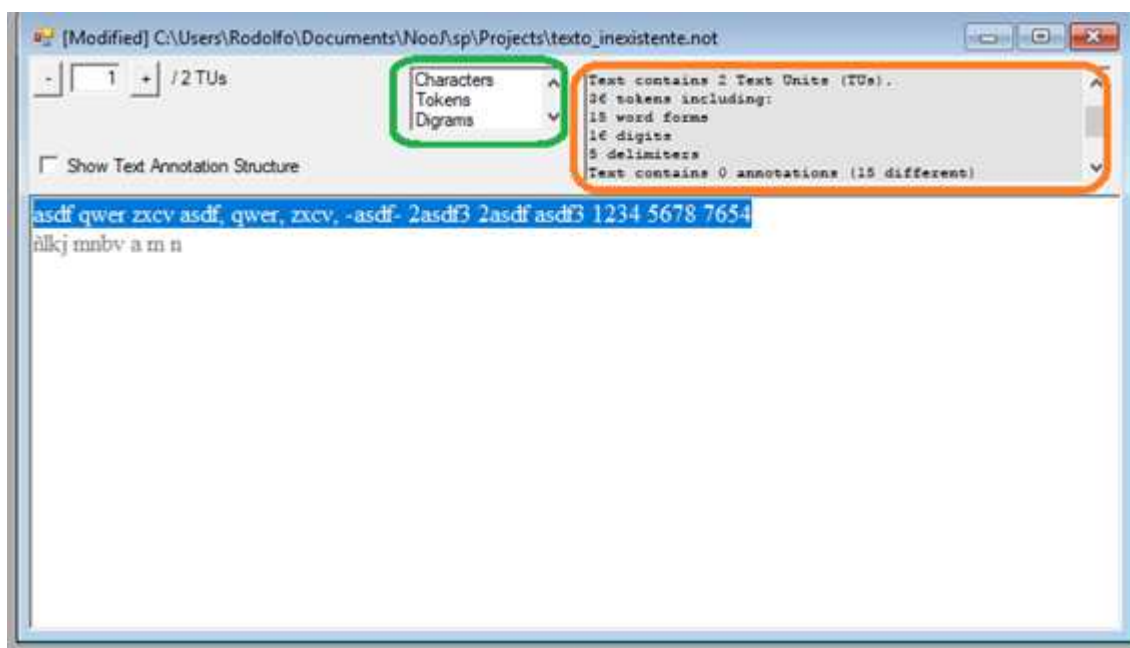


Figure 1 Analysis of non-existent sequences.

Figure 1 shows the analysis of a series of character strings which do not match any word of any language. In the upper right orange box, we can observe that two text units (*Text Units*), 36 linguistic units (*tokens*) which include 15 word forms (*word forms*), 16 figures (*digits*) and 5 punctuation marks (*delimiters*). Since we have a brief sample text, it is possible to reproduce the automatic analysis manually. By doing so, we can see that:

- line changes (*return or enter*) separate text units
- if they occur between non-alphabetical characters, letters are considered as *tokens*; an alphabetical *token* can be made up by one letter or more than one
- figures and punctuation marks are counted as *tokens* individually whether they are separated or not by other characters
- blank spaces are not counted as *tokens*

In the upper green box in the center of the Figure 2, we see there is a chart presenting other statistical data from the text:

Text has 31 characters.			
Freq	Char	Type	Unicode
16		Blk	0020
7	a	Let	0061
6	s	Let	0073
6	d	Let	0064
6	f	Let	0066
3	,	Del	002C
3	v	Let	0076
3	3	Dig	0033
3	2	Dig	0032
2	7	Dig	0037
2	m	Let	006D
2	e	Let	0065
2	q	Let	0071
2	x	Let	0078
2	n	Let	006E
2	4	Dig	0034
2	z	Let	007A
2	-	Del	002D
2	5	Dig	0035
2	w	Let	0077
2	r	Let	0072
2	6	Dig	0036
2	c	Let	0063
1	j	Let	006A
1	1	Dig	0031
1	l	Let	006C
1	8	Dig	0038
1	b	Let	0062

Figure 2 Text's characters include frequency, type and code of each character.

8 different tokens	
Freq	Tokens
6	asdf
2	qwer
2	zxcv
1	ñlkj
1	n
1	m
1	mnbv
1	a

Figure 3 Tokens present unit frequency

As it has been said before, only single letters or sequences of letters are considered to be *tokens*.

Freq	Digrams
2	asdf qwer
2	qwer zxcv
2	zxcv asdf
3	asdf asdf

Figure 4 Digrams recognize sequences of two tokens that appear repeatedly in the text

Additionally, NooJ have other functions that allow the user to add dictionaries and grammars of different natural languages. NooJ neural network includes sets of universal properties which are common to many or to large groups of languages such as the following:

- despite there are still many difficulties to define languages at a theoretical level, it is a common truth that every language has words,

- in numerous languages, the graphic representation of words is made by letters (and hence this is why one of NooJ basic functions is to recognize letters or sequences of letters as words),
- words can be grouped according to their lexical, morphological and syntactic properties,
- in many languages, words have inflected forms,
- many languages use derivative affixes to form new words from already existing ones,
- morphological variations sometimes imply root variations (irregular verbs, suppletive forms, graphic variations and stress changes, etc.),
- all natural languages are recursive, that is, units integrate in upper levels (morphemes create words, words form phrases and phrases integrate into clauses or sentences).

Stated briefly, NooJ does not «know» any language; instead NooJ «knows» many general properties of natural languages which allow users to «teach NooJ» different languages. In this sense, it is possible to speak about two different programming levels: general programming, which are the available functions to develop dictionaries and grammars, and programming per se, which is done by NooJ users with the dictionaries and grammars of specific languages.

6- NooJ Dictionaries

NooJ identifies any letter or sequences of letters delimited by non-alphabetical characters as a word. If we want NooJ to recognize the real words of a specific language, we must create dictionaries. These dictionaries have to include at least category information. Each user may define in a file – without any restrictions- which categories suit it best for their needs: it is possible to use standard labels (noun, adjective, verb, etc.) with any subclassification considered to be appropriate) as well as other labels created for specific purposes (category1, category2, category3, etc.).

A dictionary including only category information will be able to recognize the entries declared. For example, if the (Spanish) dictionary contains only the noun *libro* (singular), *libros* (plural) will not be recognized. In order to do so, it will be necessary to include all the morphological variants of the word –an uneconomical option, by the way– or rather create grammars with inflectional paradigms. For an option like this, the dictionary should include the inflectional paradigm for each entry with the category label. As for entries –or lemmas–, the user can also choose to put entries in the conventional manner by entering the infinitive form of verbs, and the masculine and the singular form of nouns and adjectives, or any other morphological variant.

In addition to inflectional and category information, the user can add any other type of information that may be relevant to calculate the grammatical properties of an entry, i.e. whether the entity named by the noun is animate or inanimate, whether the word belongs to a specific linguistic variety, etc.

7- NooJ Grammars

Today's NooJ version allows us to create different types of grammars:

1. morphological grammars
2. productive morphological grammars
3. syntactic grammars

Besides describing the inflectional paradigms of a word, morphological grammars are used to formalize inflectional morphology. In both cases, the user has to specify all the transformations which a word may undergo. For instance, let's take the word *tragar* (swallow): to change the infinitive suffix *-ar* for the third person suffix for the present tense of the subjunctive mood, it is necessary to delete two characters: the thematic vowel *a* and the consonant *r*, and then add the letters *u* and *e*, which do not have morphological charge but have orthographical charge. The resulting word is *trague*. Another instance to illustrate this is the noun *chico* (child): to turn it into the diminutive, it is necessary to replace the consonant *c* by *qu* and add the morpheme *-ito*. Examples like these let us infer that we are in front of a precise grammar which demands a thorough work. The use of grammars in manually created dictionaries allows NooJ to automatically generate dictionaries including all the inflectional and derivative variations intended. These dictionaries are to be used for the analysis of the texts.

Productive morphological grammars are used to identify and categorize sequences of letters. For example, the suffix *-ito* adds the diminutive to a noun, while the suffix *-aba* in a verb forms the first or third person singular of the Imperfect tense for the Indicative Mood of the first conjugation. These grammars are applied directly in text analysis and complement the analysis produced by previous grammars. However, productive morphological grammars can be very imprecise as they are likely to identify nouns such as *apetito* (appetite) and *ámbito* (field) as diminutives, and *sílaba* (syllable) as a verb. But NooJ allows us to set an order of priorities. If they are assigned a low level of priorities, these grammars can be applied only to the words which have not been analyzed by the dictionary. In practice, if the word *apetito*, *ámbito* and *sílaba* are listed in the dictionary, the grammar will skip their analysis.

Syntactic grammars are useful to formalize the relationships among the words which constitute a phrase. Besides, syntactic grammars are recursive. This means that formalized phrases can be part of other structures which may include them. Phrases can be formalized as words or as labels assigned by the dictionary. If we choose the first option, we will get a grammar with a poor performance, which will only recognize the sequence of words specified by the user. If we choose the second option, we will get grammars of complex performance, which will be capable of recognizing category sequences.

Chapter 2

The Corpora

1- Journalistic Texts

During our preparatory work, we collected a series of journalistic texts to build the corpus which will serve us as a source. We believe that in these texts there is a cautious use of language and that they do reflect a «real and fresh» Spanish which allow us to avoid falling into the language of literature or the academic Spanish of grammars. Texts have been selected from many Argentine newspapers in order to work on a specific language variety, Rioplatense Spanish. We understand that these texts may have some orthographic deviations, but their occurrence is considered as valid. Full texts are included in the Annex at the end of this book. The following translation of the original text is only an excerpt.

Dorrego was friend of «those» with deep roots and good relationship with the provinces. But this fact spelt his death as so did the letters written by Juan Cruz Varela and Salvador María Del Carril. Why did these letters become famous and why did they transcend when so many facts were usually concealed? Not long ago, in this newspaper, Priest Domingo Bresci wrote about the files held by the Church during the years of the dictatorship. He requested that these files should be disclosed and brought into light. It was a hard task to perpetrate. The files were under lock and key. The Church had participation in the political power and was an accomplice in the affairs of the dictatorship. Why that power –which describes history at their will, that of the victors– did not save Del Carril and Varela, who were figures of the «civilization»? To save the brave and glorious General Lavalle, who was the «other» San Martín, that of the inner front in the battle, that of the civil wars, the soldier of the illustrated city of Buenos Aires. The first who draw his sword to spill the blood of his brothers. First, the blood of Dorrego, as it was requested by his friends, the Unitarians. Well acquainted with the brutal deed he requested to Lavalle, Juan Cruz Varela, wrote with his own hands: «Letters like this, they have to be destroyed». But they were not destroyed. The shooting of Dorrego was an unpardonable crime; he was the «someone» who had to pay the price to save the evident executor, Lavalle. And the decision was made to tell the history in this way, by publishing the letters of a missing poet and a rough and hard Unitarian.

The analysis of texts will allow us to set a parameter on how to build a well-formed text in the Rioplatense Spanish variety.

2- Analysis of the Corpus of Journalistic Texts

NooJ allows us to formalize the description of natural languages to analyze aspects of Spanish inflectional and derivational morphology, terminology, syntax and semantics in linguistic corpora. In this point, our attempt is to use the **Linguistic Analysis** function to analyze a corpus of 3,883 words from journalistic texts. The texts were selected from different newspapers from Argentina. The aim is to observe which lexical categories occur with more frequency. With the aid of this function, we can gather information about the amount of characters, tokens, known and unknown words among other data, as we can see in the picture below:



Figure 7 Linguistic Analysis of a text.

Then, we use TEXT > Export annotated text as an XML document.

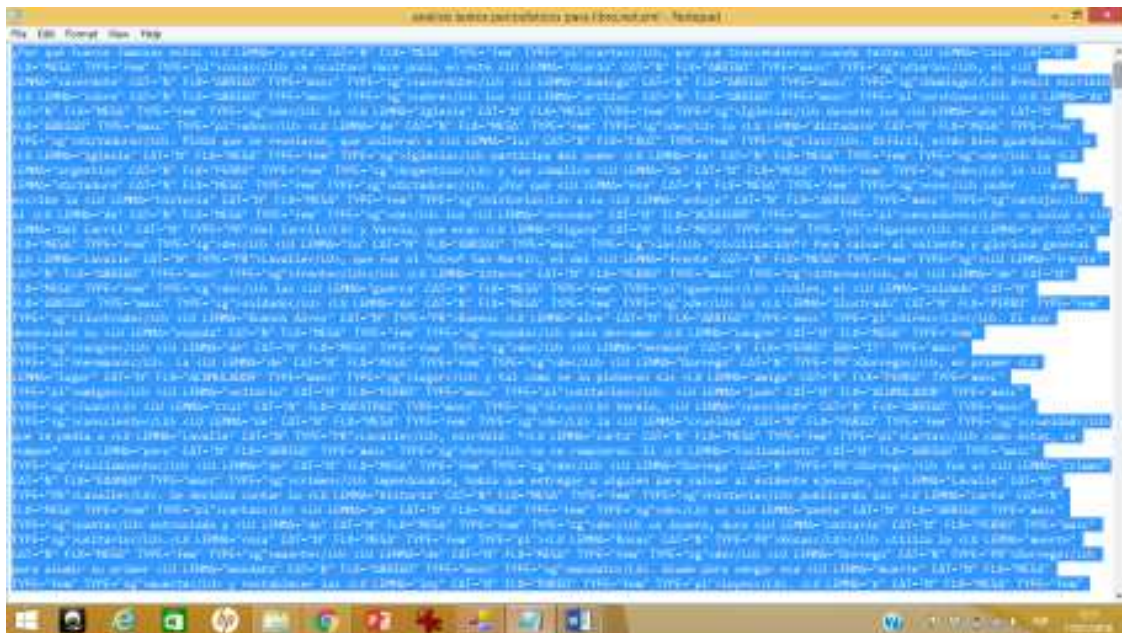


Figure 8 XLM formatted text

Once there, we click on the second option² and we type <N>, <V>, <ADJ>, <ADV>

The resulting XML text will serve us as a new source for a new analysis. We copy and paste it in a new file. Again, we perform **Linguistic Analysis** and then **Locate** to identify all the occurrences of the lexical categories we are interested in. For this sample, we type N, V, ADJ, ADV without angle brackets. In this way and by using **Statistical Analysis**, we obtain the following result:

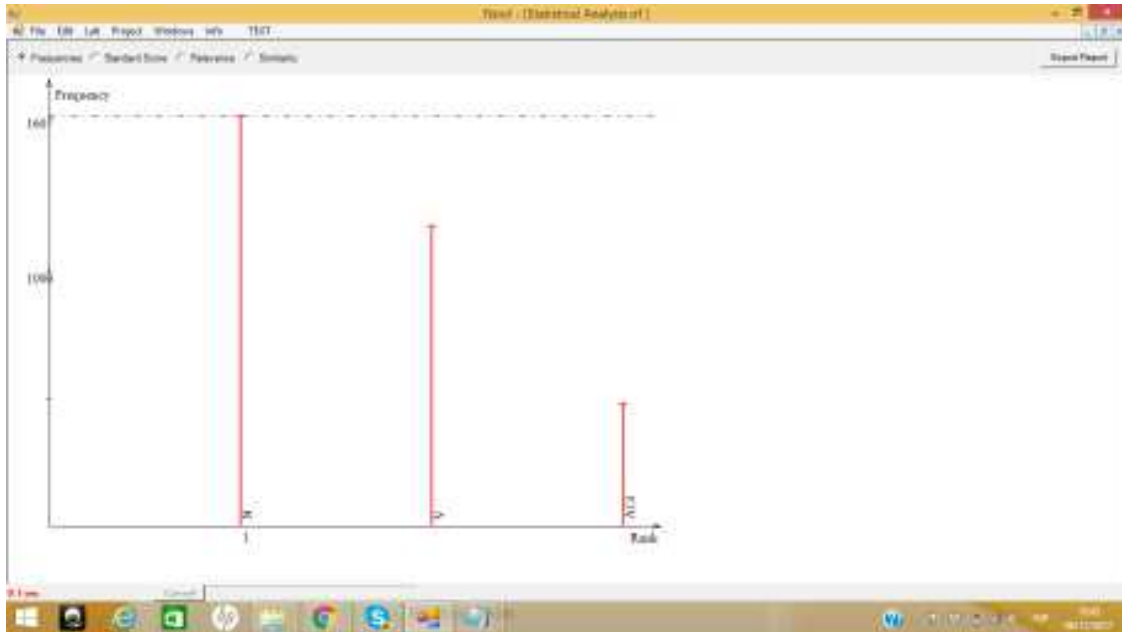


Figure 9 Frequency of Nouns, Verbs and Adjectives.

As an early conclusion, we observe that on a total of 3,883 words, there are 1,518 nouns, 878 verbs, 467 adjectives and only 150 adverbs. This gives us an idea of where to aim teaching strategies and shows that it is necessary to reinforce the adverb and adjective categories due to the poor amount of the input.

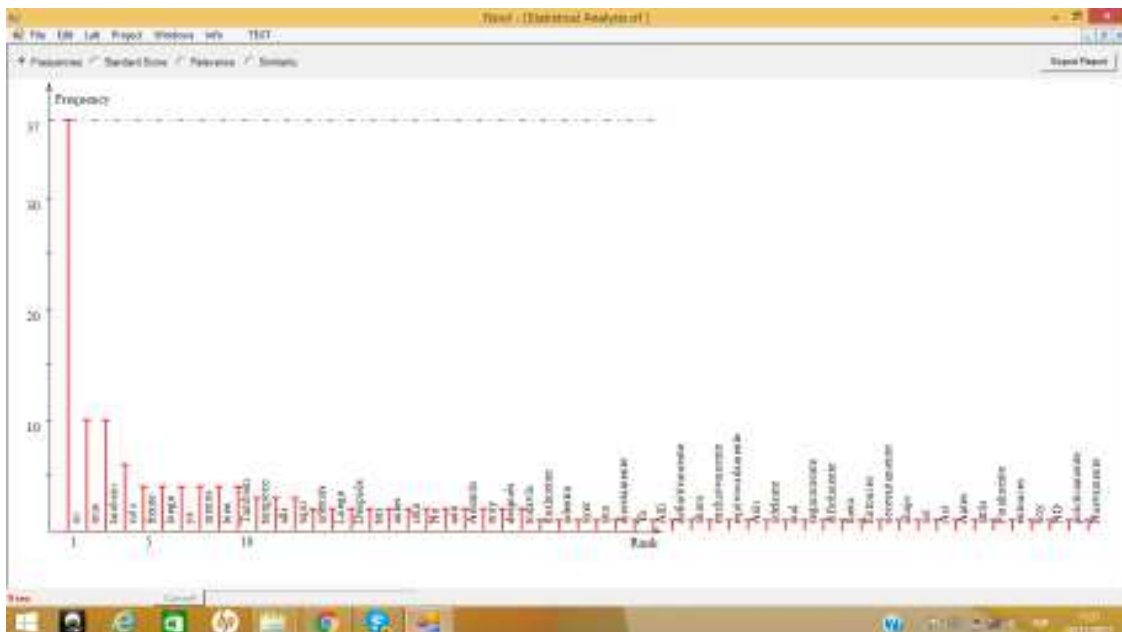


Figure 10 Frequency of adverbs in journalistic texts.

The figure below shows the general result in a graph:

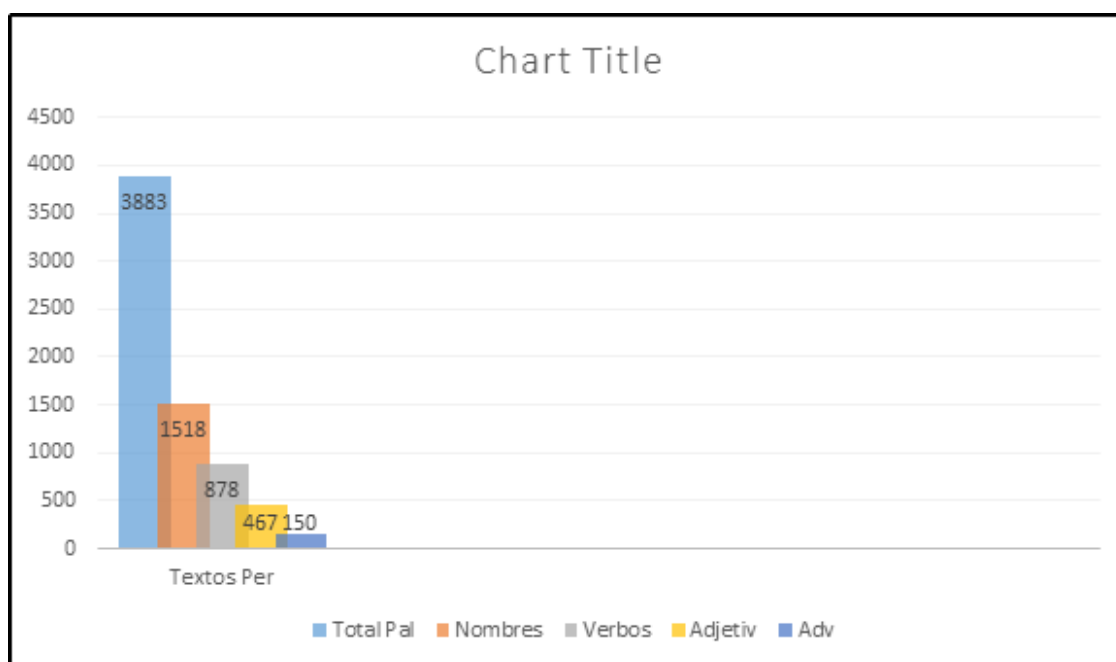


Figure 11 Frequency graph of lexical categories in journalistic texts

3- Contrast with the Texts Written by Young People: Strengths and Weaknesses

Our goal is to work on a corpus of digital texts. With the aid of colleagues and students, we created a Facebook page, *Aprendo con NooJ*² (Learning with NooJ), in order to give visibility to our project among other colleagues, students and members of the community. Then, a Google Docs with a question open to the public: *What do you think about education?*, was uploaded to our Facebook page. Based on the answers and according to the following variables:

Age: 18 to 25 years old

Language: Spanish

we got the following text:

la educacion es fundamental para el desarrollo de la persona

(education is essential to the development of the person)

In this corpus, we observe that there are some written words which do not follow orthographic rules. This may be due to typing or electronic spellchecker errors. Next, we present a fragment of the text (and its translation), but the whole corpus is in the Annex.

Para mí la educación es un derecho individual y un deber del estado hacerla llegar a toda su población. La educación en Argentina la considero muy buena, ya que le da la posibilidad a personas con bajos recursos interiorizarse, tener una opinión, una postura y una visión sobre el entorno que nos rodea. No debemos desperdiciar la oportunidad que nos da nuestro país de poder estudiar gratuitamente. Me parece repugnante que vengan extranjeros a usurpar cupos en universidades Argentinas, si su país no les brinda una educación pública gratuita por qué nosotros si?

2. Link to Aprendo con NooJ: https://docs.google.com/forms/d/1maxe0DMzNcEPzPYL7ARw_sLes25xN6V9Z1gYH6EyXM/edit

(For me, education is an individual right and an obligation of the State to extend it to the whole population. I consider that the education in Argentina is of very good quality, as it gives people with low incomes and resources the possibility to have access to it, to form them and the chance to hold an opinion, a position and a vision about the world around us. We should not waste the opportunity which gives our country to receive free education. I find utterly unpleasant that students from other countries come and take seats in Argentine universities, which keeps me wondering that if their countries of origin do not provide them public and free education, why do we have to do so?)

Data from *Aprendo con NooJ*

The following graph shows the age of the young people who left their answers in the Google document:

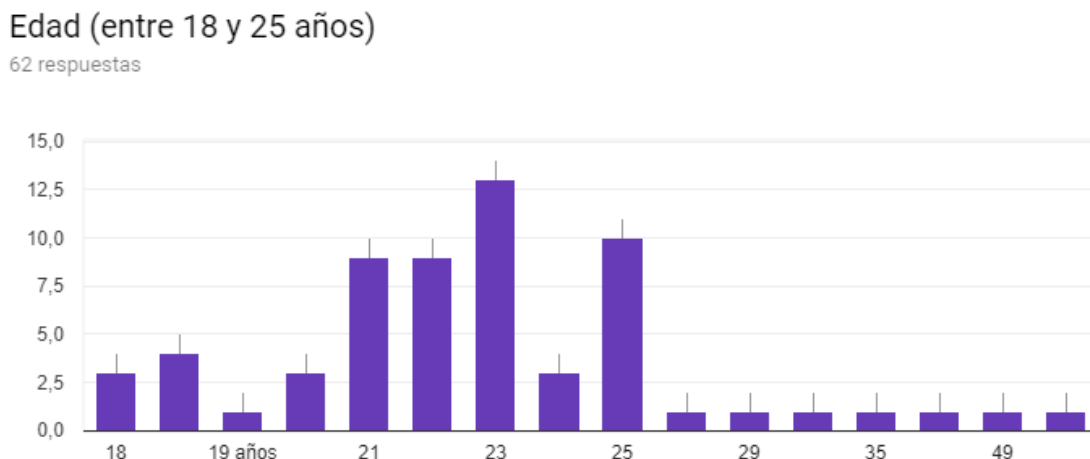


Figure 12 Age bar graph

We observe that the higher percentages are for young people between 21 and 25 years old.

4-Analysis of the Texts Written by Young People

According to the word frequency, we observe the following:

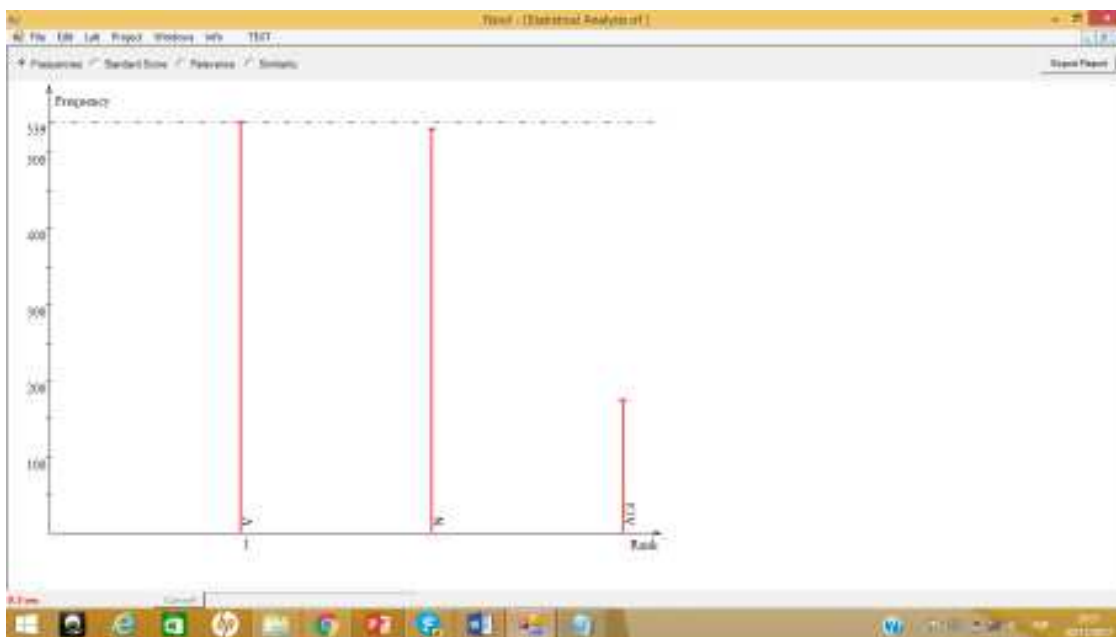


Figure 13 Word frequency in texts written by young people.

This analysis is the result of using the same dictionary for the analysis of journalistic texts as well as of the analysis of the texts written by young people. However, our intention here is not to consider those words whose orthography does not match «the rules» as being wrong. Therefore, it is necessary to enter new words, which will be shown with the tag [+jov]. The words below may help us to illustrate this point:

educacion/N+jov
interiorizarce/V+jov
trasmite/V+jov
autocriticas/ADJ+jov
concientes/ADJ+jov
esta/V+jov

Concordance mistakes that are more related to syntactic issues are commonly registered, such as the following example:

Creo que debería haber más exigencia hacia los educando y, también al docente.

A lack of knowledge of the form that an indirect interrogative clause takes is also observed:

Gracias a la educación podemos decidir que hacer con nuestras vidas y expresarnos de distintas maneras.

5- Comparison Between Corpora: Journalistic Texts in Contrast to Young People Texts

In both cases, these are opinion texts, that is to say, the basis of comparison rests on the type of text. Similarities and differences can be observed:

- word frequency is similar
- there are major orthographic deviations in the text produced by young students
- syntactic concordance is an aspect to be reinforced

6- Students' Opinions

Next, we present the personal experience accounts of some students who have taken part in a workshop and a seminar given by teachers of our research team. We believe it would be useful to share them here.

♦ Personal Experience Accounts of Attendants to the «Applied Linguistics Workshop: «Using NooJ as a Tool for the Study of Spanish Morphology» (2017)

«My approach to this workshop and NooJ was motivated by curiosity. I usually visit the «Olga Cossettini» Institute's web page looking for new training or proficiency courses and in one of these visits, I found an invitation which seemed to me very innovative. Never before had I heard about NooJ. Even so, the proposal sounded promising, especially that part which offered its users the possibility of creating their own dictionaries. Frankly speaking, I decided to take part without knowing what I was going to find. As a novice translator, my greater expectation was that of thinking NooJ as a CAT tool. I should say that at the very beginning, this did not happen but I found myself pleasantly surprised when I began to use it. Fearing that I was the only one unacquainted with this tool, I visited NooJ's web page to clarify my ideas. My first impression was vague. I found NooJ's web page somewhat chaotic, with computing terminology which was difficult for me to understand. I questioned myself if it was a good choice to take part in this course. The following day, after a little panic, I decided to give another opportunity to the tool. I took my time to surf through the web page's tabs and I found some tutorials and explanations. I watched some tutorials in Spanish. Soon I discovered that the woman hosting the videos was one of the workshop's teachers, Andrea Rodrigo, and a sense of comfort returned to my body. My choice wasn't wrong. It came the day to attend our first lesson of the «Applied Linguistics Workshop: Using NooJ in the study of Spanish morphology». The atmosphere was warmer than the fridity of the course's name. I saw few familiar faces, many smiling. We began with a round of mates while downloading the tool. In that moment, I discovered that many, if not all, of my peers had the same inexperience and curiosity as me. This was a relief; I felt more confident

and comfortable with the group, and a less fearful approach to the tool. I must emphasize the importance of using NooJ in group activities and not in a solitary learning environment. Professor Rodolfo Bonino is a person with a great deal of knowledge in Informatics and Linguistics. But this does not mean that he acted as a figure of authority nor was he distant to us. On the contrary, he, and Paula Alonso and Andrea Rodrigo explained to us the concepts all the times that were necessary for us to understand and accompanied us with the processes of discovering the tool».

«With regard to the theory given in the course, the general notions about the computer processing of natural languages were both clarifying and stimulating. Besides, the teachers made a brief but concise historical sketch of this field. This helped us to get a closer approach to this new phenomenon. On the same page, the teachers gave us basic explanations about applied linguistics and the different schools of thought about computing tools. With regards to NooJ itself, I came to find that I had conflicting emotions. I must admit that, initially, I felt somewhat disappointed when I realized that this software wouldn't be as helpful as a tool to speed up my work. In fact, on my way back home after the first lesson, I kept thinking and thinking how this tool could contribute to my work. As a novice translator, I have some basic knowledge about computer tools and none of them serve as a resource to learn grammar. At the beginning of the course, and because our occasional lack of knowledge, we found the practice exercises were quite simple and repetitive for the sole purpose of learning how the tool worked. Despite its simplicity, these activities offered a challenge. There were moments where difficulties and little frustration came to our minds, but putting them into words or questions to our teachers was of great use. When someone had an issue, the issue was shared with the group and we helped each other to improve and get the best of it. Our teachers were always there to lead us in the right way. In some cases, we took advantage of other classmates' mistakes to avoid making the same ones, and we celebrated others' hits and discoveries to improve our performance. It took me two lessons to find what NooJ had to offer. Unlike other tools, NooJ enables the user to control intermediate processes besides the input and output of information. This allows placing particular segments and formalizations of language. In other words, users can freely create their own grammars, dictionaries and categories to organize the lexical units.»

«One interesting point I discovered with NooJ is the notion of recursivity. While I was a student, I studied Noam Chomsky's generative grammar. One of the aspects that called my attention from the very beginning was recursivity. With it, a grammar can generate an infinite amount of syntactic constructions from a finite number of rules and abstract units. Initially, I found it difficult to observe it in specific examples and even harder to build examples of my own. When we began loading the data in NooJ, I came to understand better the nature of this concept. For example, in our grammars we added general explanations under the names GÉNERO 1' or 'NÚMERO 3' and then we labeled the words starting from these categories. Consequently, I realized there is a limited number of patterns to create the gender or number, and that all the words can be encapsulated in these groups. Evidently, from a set of finite elements it is possible to generate and interpret an infinite number of syntactic constructions.»

«There is a point which I wish to remark with the use of NooJ and it is concerned with systematic irregularities. As it is necessary to use a series of commands to tell the system how to create new words from a base –or root- I felt compelled to re-explain to myself how Spanish morphology works. Perhaps it is a strange thing to say but as a native speaker I did not use to question the different irregularities that the Spanish language may have. But with the different works in the workshop I could see with other eyes the many and different types of verb irregularities, graphical changes, diphthongs, combinations of irregularities and many others aspects with a clearer differentiation. For instance, we used the especial command («<U>») to make diphthongs for words such as «acertar '!acierto». I find this option extremely useful for those people who teach foreign languages or Spanish as a second language. As language is explored in so much detail like, it is easier to understand common –and not so common- mistakes of language learners.»

Personal Experience Accounts of Attendants to the Linguistics-and-Languages-oriented Seminar «Contributions of Computational Linguistics to the Teaching of Spanish as L1 and L2», given by Silvia Reyes and Andrea Rodrigo (2017)

«We started reading the Tutorial for Spanish. This work was designed by the teachers Rodolfo Bonino and Andrea Rodrigo, and it offers a series of indications and exercises to initiate the language learner in NooJ. The Tutorial is an essential tool for inexperienced users in NLP (Natural Language Processing) programs. NooJ is freely available on the Internet and, for Spanish, users can download a written version of the tutorial. NooJ's webpage includes the tab Tutorial. In it, Andrea Rodrigo will guide the users in a series of videos to learn the trade from the ground up towards more challenging activities.»

*«Of singular interest was a proposal made in lesson 2 around building an inflectional or derivational grammar and the commands which can be used in that process. In the Tutorial, one of the examples shows how to use these commands –and specially how to combine them. One example to illustrate this is the Spanish verb **lucir**: the combination of the commands LUCIR=<B2><L>z<R>o/pte+1era+sg yields the verb **luzco** as a conjugated verb with its corresponding inflectional features. A close observation of this sequence led us to consider other possibilities to represent the conjugations of the Spanish verb paradigm in a much more economical manner. In other words, one of the goals is to seek much more compact formulas with fewer commands to express the different labels for the Spanish verb inflection.»*

«NooJ is a remarkable tool, open access and free! It has multiple activities which are user-friendly to any language learner, and it is supported by many operating systems common languages and various keyboards. I consider it as an innovative approach to study natural languages. NooJ does not work with memory databases of grammar rules. In contrast, NooJ operates with grammatical trials and tasks in a ludic oriented way that allows students to learn language while manipulating, creating and analyzing texts. I feel that NooJ is a positive and refreshing resource for learning Spanish as a first or second language.»

«My first approach to NooJ was in an attempt to focus on the verb category from another standpoint to determine how much regular or irregular a verb lexical item can be, the aspects and knowledge which can be derived from them using this tool and how these can be used to keep on working and expanding the verb inflectional model. In my experience, I should pinpoint the function «Lab Morphology». This function is easy to understand and use, and can be operated with many language forms. In my case, it was possible to apply it to the whole Spanish verb paradigm, which for our educational purposes it was of great value. We could detect the formants of a lexical item, their variations and their regularities and/or irregularities. In other words, I dare say that «Lab Morphology» allowed us to «play» with the lexical item while learning.»

CONCLUSIONS

Throughout this book, we have accepted a fundamental supposition, understanding that it is possible to generate an explicit knowledge of language by using a computer tool. From this assumption and depending on the use of the NooJ platform we believe that it is possible to make hypotheses about language, in this case about Spanish, since knowledge is produced by corroborating different hypotheses on the computer.

We decided to carry out observations on young people's written texts that would allow us to consider whether a youth grammar exists or not. For that reason, we made use of the Facebook page *Aprendo con NooJ*³. It became an effective resource because a Google document was inserted therein so that young people between 18 and 25 years old would answer a question. And their responses to that question enabled us to elaborate a diagnosis about their texts and to establish our first working conclusions.

In the second section, we present specific exercises for the use of NooJ in Spanish. Each class simultaneously provides two kinds of learning situations: how to work with the computer tool and, at the same time, how to deal with linguistic data in Spanish. In fact, we return to what is already known, for example, gender, verbal paradigms, syntactic constructions, concordance relationships, and so on.

This book in no way pretends to be a fully comprehensive study of the grammar young people use, and simply constitutes a first approach to the subject. It supports the hypothesis that when starting from a youth grammar, it must be elucidated during the language class. An 'error', in this sense, becomes a place for analysis, about which it is necessary to think together. We are undoubtedly in a process of permanent change, and it would even be necessary to register how iconicity is gradually finding its way among young people, in terms of chats, of emoticons, of the famous 'skere' as an expression coming from rap. The acceptance of the other is part of accepting how young people talk and write.

In this context, the neutrality of a computer tool devoid of contents offers an ideal land to load data about Spanish, as well as the chance of checking what happens in other languages by the use of a common format, and consequently of envisioning the acquisition of a «multilinguistic» proficiency. This delimits a new path that leads us, as teachers and researchers, to confront issues in their increasing complexity, and encourages a permanent search for stressing the proactive nature of teaching, which cannot be achieved without the real and effective participation of students.

3 Available at the following link: <https://www.facebook.com/indagacioneslinguisticas>

Esta primera edición del libro
Aprendo con NooJ
de *Andrea Rodrigo* y *Rodolfo Bonino*
se terminó de imprimir el 15 de agosto de 2019
en los talleres gráficos de Editorial Ciudad Gótica



...“Se trata, por una parte, de un excelente manual del usuario de NooJ, en español y en inglés, puesto que el libro es, en gran medida, bilingüe. Por otra parte, es también una obra de análisis computacional de la lengua española, que contiene numerosos ejemplos comentados de descripción informática de distintos aspectos ortográficos, morfológicos y sintácticos del español, en su variedad rioplatense, que tanto ha aportado y sigue aportando al patrimonio común de nuestro idioma. Constituye, asimismo, un manual de didáctica de las lenguas, ya que presenta importantes elementos, tanto teóricos como prácticos, de aplicación didáctica de NooJ a la enseñanza/aprendizaje de lenguas maternas y extranjeras, que incluyen incluso relatos de experiencias de estudiantes tras los talleres de lingüística aplicada basados en este entorno informático.”...

... “It is an excellent manual for the NooJ user, in Spanish and English, since the book is to a large extent bilingual. On the other hand, it is also a work of computational analysis of the Spanish language, containing numerous commented examples and the computational description of different orthographic, morphological and syntactic features of Spanish, in its Rioplatense variety, which has so much contributed to the common heritage of our language, and continues to do so. It is moreover a manual of language teaching, since it presents important theoretical as well as practical elements of the pedagogical application of NooJ to the teaching/learning of mother and foreign tongues, which even include the personal experience accounts of students after attending applied linguistics workshops based on this computer environment”...

Xavier Blanco

Universitat Autònoma de Barcelona

ISBN 978-987-597-398-5



CG
Editorial