

## **Análisis automático de ambigüedades en español: ‘preposición’**

**Automatic analysis of ambiguities in Spanish: ‘preposition’**

**Prof. Stella Maris Moro**

Universidad Nacional de Rosario

Rosario, Argentina

smmoro@yahoo.com.ar

### **Abstract**

Automatic labeling tools for texts in natural language consider ambiguities as an obstacle difficult to treat. In previous works we have proposed a strategy of phased disambiguation and considered its effectiveness in assigning 'Noun', 'Adjective', 'Verb' labels to strings of characters corresponding to one or the other, according to their position in the sentence. This time we propose the possibility of extending this strategy to ambiguous chains involving the category 'Preposition': “ante”, “entre”, “para”, “sobre”, “fuera (de)”, “aparte (de)”. We declare the rules that work on a new stage of labeling, using the results of previous stages as input. Finally we analyze the effectiveness of this proposal.

**Keywords:** ambiguity, automatic analysis, disambiguation, interlanguage, spanish second language.

### **Resumen**

Las herramientas de etiquetado automático de textos escritos en lenguaje natural encuentran las ambigüedades como un obstáculo de difícil tratamiento. En trabajos anteriores hemos propuesto una estrategia de desambiguación por etapas y consideramos su efectividad al asignar etiquetas ‘Nombre’, ‘Adjetivo’, ‘Verbo’ a cadenas de caracteres a las que correspondía una u otra de acuerdo a su posición en la oración. En esta ocasión consideramos la posibilidad de ampliar esta estrategia a cadenas ambiguas que involucran la categoría ‘Preposición’: “ante”, “entre”, “para”, “sobre”, “fuera (de)”, “aparte (de)”. Declaramos reglas que funcionan en una nueva etapa de etiquetado, utilizando como input los resultados de las etapas anteriores. Finalmente analizamos la efectividad de esta propuesta.

**Palabras claves:** ambigüedad, análisis automático, desambiguación, preposición.

## 1. INTRODUCCION

Las herramientas de etiquetado automático de textos escritos en lenguaje natural encuentran las ambigüedades como un obstáculo de difícil tratamiento.

En trabajos anteriores [1][2][3] hemos propuesto una estrategia de desambiguación por etapas y consideramos su efectividad al asignar etiquetas ‘Nombre’, ‘Adjetivo’, ‘Verbo’ a cadenas de caracteres<sup>1</sup> a las que correspondía una u otra de acuerdo a su posición en la oración.

En esta ocasión consideramos la posibilidad de ampliar esta estrategia a cadenas ambiguas que involucran la categoría ‘Preposición’<sup>2</sup>: “ante”, “entre”, “para”, “sobre”, “fuera (de)”, “aparte (de)”.

La tabla 1 presenta las diferentes categorías que pueden asignarse a estas cadenas<sup>3</sup>:

Tabla 1. Ambigüedades que involucran la categoría ‘Preposición’

<i>Cadena de Caracteres</i>	<i>Categorías asignables</i>
‘ante’	‘Nombre’, ‘Preposición’
‘entre’ ‘para’	‘Verbo’, ‘Preposición’
‘fuera’ ‘aparte’	‘Verbo’, ‘Adverbio’, ‘Preposición’
‘sobre’	‘Nombre’, ‘Verbo’, ‘Preposición’

En este trabajo, declaramos reglas que operan desambiguando automáticamente estas cadenas.

Por último, analizamos la efectividad de esta propuesta, considerando las salidas obtenidas por el etiquetador, a partir de las reglas declaradas.

## 2. HERRAMIENTAS Y METODOLOGÍA DE ETIQUETADO

Al igual que en los trabajos previos mencionados, utilizamos dos herramientas computacionales:

- a.- *SMORPH* (desarrollado por Aït-Mokhtar [4]): autómata de estados finitos que permite lematizar<sup>4</sup>[5] y analizar morfológicamente las cadenas de caracteres, dando como salida la asignación categorial y morfológica correspondiente a cada ocurrencia, de acuerdo con los rasgos que se declaran.

<sup>1</sup> Denominamos “cadena” a la sucesión de caracteres que corresponden a lexemas en la lengua natural, y “frase” a la sucesión de cadenas que constituyen partes de oración y/o sintagmas nominales, verbales, adjetivales, adverbiales y preposicionales.

<sup>2</sup> La cadena ‘bajo’ puede recibir 5 etiquetas: ‘Nombre’, ‘Adjetivo’, ‘Verbo’, ‘Adverbio’, ‘Preposición’. Los límites de este artículo no nos permiten desarrollar su tratamiento, razón por la cual será abordada junto con la categoría ‘Adverbio’ en futuros trabajos.

<sup>3</sup> Tomamos en cuenta las categorías léxicas atribuidas a cada palabra en el *Diccionario de la Lengua Española*, Real Academia Española, 22ª edic., en <http://www.rae.es/rae.html>.

<sup>4</sup> Lema: cadena que designa por convención toda la serie de variantes morfológicas de un lexema. Por ej.: ‘libro’ es el lema que designa el paradigma nominal: ‘libro / libros’, mientras que ‘librar’ es la denominación del paradigma verbal ‘libro / libras / librás / libra /...’

b.- Módulo Post-Smorph (*MPS*), implantado en máquina por Faiza Abbaci [5]: tiene como input la salida de Smorph y, por medio de reglas de recomposición, descomposición y correspondencia que declara el operador, analiza la secuencia de lemas y rasgos gramaticales que se obtiene como output de Smorph.

En el procesamiento de información, ambas herramientas utilizan fuentes declarativas en archivos que son cargados y pueden ser modificados continuamente por el operador.

Para la aplicación de SMORPH utilizamos los archivos de fuentes cargadas por el equipo Infosur, sobre los que se operaron modificaciones en la declaración de los rasgos necesarios para el tratamiento automático de las ambigüedades.

En cuanto a MPS opera con tres tipos de reglas que especifican secuencias posibles de lemas: reglas de **reagrupamiento**, **descomposición** y **correspondencia** [1].

Tal como ya presentamos en [1], la metodología de análisis en la que estamos trabajando realiza un análisis por etapas: cada una de ellas articula la información necesaria para las etapas posteriores. Para lograrlo, nuestra propuesta reagrupa los tres tipos de reglas en cuatro conjuntos diferentes:

- **reglas de ejecución preliminar:** incluye reglas de reagrupamiento y de descomposición que facilitan la aplicación de las reglas siguientes. (Ejec0)
- **reglas de ejecución primaria:** reglas que analizan las cadenas y secuencias no ambiguas. Incluye reglas de recomposición que operan sobre la formación de sintagmas que incluyen cadenas no ambiguas, y también desambigua cadenas resolubles en primera instancia, tales como ‘el libro’. (Ejec1)
- **reglas de postergación:** opera sobre cadenas ambiguas, no resolubles en primera instancia (tales como ‘la amenaza’). Se trata de reglas de correspondencia que asigna nuevas etiquetas que estarán disponibles para las etapas de análisis posteriores. (Ejec1)
- **reglas de ejecución secundaria:** operan sobre cadenas y secuencias que han sido postergadas en ejecuciones anteriores, y aprovecha para esto el etiquetamiento de otras secuencias analizadas en etapas anteriores. (Ejec2 y Ejec3)

### 3. ANÁLISIS DE DATOS LINGÜÍSTICOS Y REGLAS PROPUESTAS

#### 3.1. Ambigüedad Nombre / Preposición: ‘ANTE’

El analizador deberá poder aplicar a esta cadena de caracteres la etiqueta ‘Nombre’ o ‘Preposición’ según corresponda:

- (1) Elegantes zapatos con piel combinada en **ante** y tafilete.
- (2) Deberá comparecer **ante** la ley.

Se aplicó en el análisis de estas oraciones las reglas que habían sido declaradas en MPS en etapas anteriores de trabajo, y el resultado obtenido fue el siguiente:

Tabla 2. Output inicial del analizador<sup>5</sup> para la cadena 'ante'

'Elegantes zapatos'.
[ 'elegante zapato', 'EMS', 'SN', 'EMS', 'A+°N' ].
'con piel'.
[ 'con piel', 'EMS', 'SP', 'EMS', 'P+SN' ].
'combinada'.
[ 'combinar', 'EMS', 'Ppio' ].
<b>'en ante'.</b>
[ <b>'en ante', 'EMS', 'SP', 'EMS', 'P+SN' ]</b> .
'y'.
[ 'y', 'EMS', 'cop' ].
'tafilete'.
[ 'tafilete', 'EMS', 'SN', 'EMS', '°N' ].
'.
'Deberá comparecer'.
[ 'deber comparecer', 'EMS', 'SV', 'EMS', 'V+Vfin' ].
<b>'ante'.</b>
[ <b>'ante', 'EMS', 'SN', 'EMS', '°N' ]</b> .
'la ley'.
[ 'el ley', 'EMS', 'SN', 'EMS', 'D+°N' ].
'.

Como puede observarse, para (1) el output resultó correcto, porque la existencia de la regla:

Preposición + Nombre → SP

declarada previamente en el archivo de Reglas permitió desambiguar adecuadamente la cadena en la frase encabezada por preposición. En cambio, el ejemplo (2) fue analizado de forma errónea.

De acuerdo con la estrategia de etiquetado que venimos proponiendo hasta aquí, se requiere una regla de postergación, que permita postergar la atribución de etiqueta a la cadena 'ante' hasta disponer del análisis del resto de la secuencia de palabras.

Para esto, en el archivo de entradas se asigna el rasgo ambNP a la cadena 'ante', y se propone la siguiente regla de postergación:

R1.ambNP → AmbNP

Se trata de una regla de correspondencia que opera en Ejec1 de análisis sólo para postergar el etiquetado hasta que en Ejec2 dispongamos de la etiqueta SN para 'la ley' obtenida en el output de Ejec1.

Agregamos la regla

R2.AmbNP + SN → SP

para que se ejecute en Ejec2. El resultado obtenido es el siguiente:

<sup>5</sup> Etiquetas:

'EMS': Etiqueta Morfo-Sintáctica (a la que se asocian rasgos de Clase de Palabra o de Tipo de Sintagma)

SN=Sintagma Nominal. SV=Sintagma Verbal. SP=Sintagma Preposicional. A=Adjetivo. N=Nombre. Ppio=Participio.

D=Determinante. V=Verbo. Cop= conjunción copulativa.

El símbolo ° señala lo que se considera el núcleo del SN.

Tabla 3. Output obtenido con las Reglas de Desambiguación para la cadena ‘ante’

'Deberá comparecer'. [ 'deber comparecer', 'EMS', 'SV', 'EMS', 'V+Vinf' ]. 'ante la ley'. [ 'ante el ley', 'EMS', 'SP', 'EMS', 'P+N' ]. '.
--

### 3.2. Ambigüedad Verbo / Preposición: ‘ENTRE’, ‘PARA’

Veamos los diferentes contextos oracionales en que puede aparecer cada una de estas cadenas:

- (3) Le pidió que **entre**. [‘Verbo’]
- (4) **Entre** todos pudieron lograrlo. [‘Preposición’]
- (5) Está **entre** la espada y la pared. [‘Preposición’]

Para el caso de ‘entre’, hay que considerar que en la secuencia de (4) la preposición ‘entre’ selecciona un complemento nominal (en este caso un pronombre) necesariamente plural, pero que en (5), exige la aparición de dos términos coordinados, que pueden ser singular o plural. En (3) es verbo.

De acuerdo a las reglas que teníamos declaradas, el output obtenido era el siguiente:

Tabla 4. Output inicial del analizador para la cadena ‘entre’

'Le dijo'. [ 'lo decir', 'EMS', 'SVN', 'EMS', 'Cl+°V' ]. 'que'. [ 'que', 'EMS', 'rel' ]. [ 'que', 'EMS', 'sub' ]. ' <b>entre</b> '. [ ' <b>entrar</b> ', 'EMS', 'SV', 'EMS', 'V' ]. '. ' <b>Entre</b> '. [ ' <b>entrar</b> ', 'EMS', 'SV', 'EMS', 'V' ]. 'todos'. [ 'todo', 'EMS', 'SN', 'EMS', '°N' ]. 'pudieron lograr'. [ 'poder lograr', 'EMS', 'SV', 'EMS', 'V+Vinf' ]. 'lo'. [ 'lo', 'EMS', 'encl' ]. '. 'Está'. [ 'estar', 'EMS', 'SV', 'EMS', 'V' ]. ' <b>entre</b> '. [ ' <b>entrar</b> ', 'EMS', 'SV', 'EMS', 'V' ]. 'la espada'. [ 'el espada', 'EMS', 'SN', 'EMS', 'D+°N' ]. 'y'. [ 'y', 'EMS', 'cop' ]. 'la pared'. [ 'el pared', 'EMS', 'SN', 'EMS', 'D+°N' ]. '.
--

Como vemos, si bien en el primer caso el etiquetado resulta adecuado, no lo es en los restantes.

Para estos casos, requeriremos para SMORPH agregar a la secuencia 'entre' el rasgo de ambigüedad preposición / verbo 'ambPV' en entradas, y declarar para MPS las siguientes reglas:

- R3. ambPV → AmbPV (regla de postergación ejecutable en Ejec1)
- R4. AmbPV + SN → AmbSPSV (regla de postergación en Ejec2)
- R5. AmbPV + SN + cop + SN → SP (id.)
- R6. Subordinante + AmbPV → Sub + V (regla que, por posición, va a operar en Ejec2 siempre y cuando no se den las anteriores)
- R7. AmbSPSV + SV → SP+SV (regla ejecutable en Ejec3)
- R8. SV + SN + AmbSPSV → SV + SN + SP (id.)
- R9. SV + SA + AmbSPSV → SV + SA + SP (id.)
- R10. SV + AmbSPSV → SV + SP (id.)

Veamos los resultados de declarar estas reglas en el analizador:

Tabla 5. Output obtenido con las Reglas de Desambiguación para la cadena 'entre'

<p>'Le dijo'.          [ 'lo decir', 'EMS', 'SVN', 'EMS', 'Cl+°V' ].  <b>'que entre'</b>.          [ 'que entre', 'EMS', 'OSub', 'EMS', 'Sub+V' ].          ':  <b>'Entre todos pudieron lograr'</b>.          [ 'entre todo poder lograr', 'EMS', 'SP+SV' ].          'lo'.          [ 'lo', 'EMS', 'encl' ].          ':          'Está'.          [ 'estar', 'EMS', 'SV', 'EMS', 'V' ].  <b>'entre la espada y la pared'</b>.          [ 'entre el espada y el pared', 'EMS', 'SP', 'EMS', 'P+SN+Cj+SN' ].          ':</p>
---

La regla R4, evita que se asigne de manera apresurada la etiqueta SP o SV hasta tanto no se tenga el contexto, y opera como Regla de Postergación en Ejec2. En la siguiente ejecución (Ejec3) se determina si la frase ambigua está precedida o seguida por un SV, en cuyo caso se desambigua como SP por las reglas R7a R10.

R8 y R9 permiten tratar cadenas más complejas, del tipo:

- (6) Está la joven entre la espada y la pared.
- (7) Está atrapada entre la espada y la pared.

Esta metodología de etiquetado asegura que oraciones elípticas tales como:

- (8) Entre los automóviles.

permanezcan ambiguas, tal como resulta en el lenguaje natural, pues sólo son factibles de desambiguación en el contexto.

Tabla 6. Output con ambigüedad en lenguaje natural, cadena ‘entre’

'Entre los automóviles'.  
 [ 'entre el automóvil', 'EMS', 'AmbSPSV' ].  
 '.

Con respecto a la secuencia ‘para’, tenemos los siguientes contextos oracionales posibles:

- (9) El perro no **para** de ladrar. [‘Verbo’]
- (10) **Para** Juan resulta posible. [‘Preposición’]
- (11) Lo hizo **para** ganar. [‘Preposición’]
- (12) Lo hizo **para** que no lo descubrieran. [‘Preposición’]

En el primer caso (9), la secuencia corresponde a un verbo, pero en las tres restantes se trata de diferentes usos de la preposición: en (10) selecciona un SN (en este caso un Nombre Propio), mientras que (11) y (12) son construcciones de finalidad, la primera con infinitivo y la segunda con una Subordinada.

Veamos el output sin las reglas de desambiguación:

Tabla 7. Output inicial del analizador para la cadena ‘para’

'El perro'.  
 [ 'el perro', 'EMS', 'SN', 'EMS', 'D+°N' ].  
 'no'.  
 [ 'no', 'EMS', 'Adv' ].  
**'para'**.  
 [ 'parar', 'EMS', 'SV', 'EMS', 'V' ].  
 'de ladrar'.  
 [ 'de ladrar', 'EMS', 'SP', 'EMS', 'P+Vinf' ].  
 '.

**'Para'**.  
 [ 'parar', 'EMS', 'SV', 'EMS', 'V' ].  
 'Juan'.  
 [ 'Juan', 'EMS', 'SN' ].  
 'resulta'.  
 [ 'resultar', 'EMS', 'v', 'MODOV', 'ind', 'PERS', '3a', 'NUM', 'sg', 'TPO', 'pres' ].  
 [ 'resultar', 'EMS', 'v', 'MODOV', 'imper', 'PERS', '2a', 'NUM', 'sg', 'TPO', 'pres', ].  
 'posible'.  
 [ 'posible', 'EMS', 'adj', 'GEN', ' ', 'NUM', 'sg' ].  
 '.

'Lo hizo'.  
 [ 'lo hacer', 'EMS', 'SV', 'EMS', 'CI+°V' ].  
**'para ganar'**.  
 [ 'para ganar', 'EMS', 'SP', 'EMS', 'P+Vinf' ].  
 '.

'Lo hizo'.  
 [ 'lo hacer', 'EMS', 'SV', 'EMS', 'CI+°V' ].  
**'para'**.  
 [ 'parar', 'EMS', 'SV', 'EMS', 'V' ].  
 'que'.  
 [ 'que', 'EMS', 'rel' ].  
 [ 'que', 'EMS', 'sub' ].  
 'no'.  
 [ 'no', 'EMS', 'Adv' ].  
 'lo descubrieran'.  
 [ 'lo descubrir', 'EMS', 'SV', 'EMS', 'CI+°V' ].  
 '.

En la primera oración, el etiquetado es correcto. El tercero también, pues opera una regla declarada previamente de prep + inf, que en etapas anteriores de la investigación se utilizó para desambiguar infinitivos tales como 'parecer', 'ser', etc. [1]

Para etiquetar adecuadamente los dos restantes, en primer término se declara un rasgo ambPV en entradas para la cadena 'para', tal como se hizo con 'entre'. Las reglas R4 y R7 a R10 aplican también en este caso. Sólo resta agregar:

R11.AmbPV + Subordinante → Prep + Cláus. Finalidad

Para el caso de 'parar de + inf', se declara el rasgo 'mod2' para el verbo 'parar' y en Ejec0 se declara la regla:

R12.VMod2+ 'de' + Infinitivo → SV

El resultado obtenido fue el siguiente:

Tabla 8. Output obtenido con las Reglas de Desambiguación para la cadena 'para'

<p>'El perro'.          [ 'el perro', 'EMS', 'SN', 'EMS', 'D+°N' ].          'no'.          [ 'no', 'EMS', 'Adv' ].  <b>'para de ladrar'.</b>          [ 'parar de ladrar', 'EMS', 'SV', 'EMS', 'V+P+Vinf' ].          '.  <b>'Para Juan resulta'.</b>          [ 'para Juan resultar', 'EMS', 'SP+SV' ].          'posible'.          [ 'posible', 'EMS', 'adj', 'GEN', '_, 'NUM', 'sg' ].          '.          'Lo hizo'.          [ 'lo hacer', 'EMS', 'SV', 'EMS', 'CI+°V' ].  <b>'para que'.</b>          [ 'que para', 'EMS', 'CFin', 'EMS', 'P+CIFin' ].          'no'.          [ 'no', 'EMS', 'Adv' ].          'lo descubrieran'.          [ 'lo descubrir', 'EMS', 'SV', 'EMS', 'CI+°V' ].          '.</p>
---

Al igual que con la cadena 'entre', para una oración como la siguiente es necesario mantener la ambigüedad, para lo cual siguen operando las reglas ya declaradas:

(13) Para los automóviles.

Tabla 9. Output con ambigüedad en lenguaje natural, cadena 'para'

<p><b>'Para los automóviles'.</b>          [ 'para el automóvil', 'EMS', 'AmbSPSV' ].          '.</p>
---

### 3.3. Ambigüedad Nombre / Verbo / Preposición: 'SOBRE'

Aquí los contextos posibles de aparición serían:



- (14) Trajo un **sobre** cerrado. [‘Nombre’]
- (15) El mensaje está **sobre** la mesa. [‘Preposición’]
- (16) No quiero que **sobre**. [Verbo]

Con las reglas declaradas en la etapa previa de nuestra investigación, el resultado obtenido era el siguiente:

Tabla 10. Output inicial del analizador para la cadena ‘sobre’

<p>'Trajo'.                  [ 'atraer', 'EMS', 'SV', 'EMS', 'V' ].                  '<b>un sobre</b>'.                  [ '<b>un sobre</b>', 'EMS', 'SN', 'EMS', 'D+°N' ].                  'cerrado'.                  [ 'cerrar', 'EMS', 'Ppio' ].                  '.'                  'El mensaje'.                  [ 'el mensaje', 'EMS', 'SN', 'EMS', 'D+°N' ].                  'está'.                  [ 'estar', 'EMS', 'SV', 'EMS', 'V' ].                  '<b>sobre</b>'.                  [ '<b>sobre</b>', 'EMS', 'ambPVN' ].                  'la mesa'.                  [ 'lo mesa', 'EMS', 'AmbSNSV' ].                  '.'                  'No'.                  [ 'no', 'EMS', 'Adv' ].                  'quiero'.                  [ 'querer', 'EMS', 'SV', 'EMS', 'V' ].                  'que'.                  [ 'que', 'EMS', 'rel' ].                  [ 'que', 'EMS', 'sub' ].                  '<b>sobre</b>'.                  [ '<b>sobre</b>', 'EMS', 'ambPVN' ].                  '.'</p>
---

La oración (14) es resuelta sin dificultades por una regla de buena formación de SN ya declarada:

$$\text{Det} + \text{N} \rightarrow \text{SN}$$

Ahora bien, aún cuando se había declarado un rasgo de ambigüedad Preposición / Verbo / Nombre ‘ambPVN’, la ambigüedad no había sido resuelta.

Como en los casos anteriores, corresponde declarar una regla de postergación, que permita proceder a un etiquetamiento secundario:

$$\text{R13.ambPVN} \rightarrow \text{AmbPVN}$$

y las reglas de desambiguación correspondientes para la Ejec2:

$$\text{R14. AmbPVN} + \text{SN} \rightarrow \text{AmbSPSV}$$

$$\text{R15. Sub} + \text{AmbPVN} \rightarrow \text{OSub}$$

$$\text{R16. AmbPVN} \rightarrow \text{SV}$$

La regla R14 funcionará paralela a R4, y activará en la ejecución posterior las R7 a R10. La R15 corresponde al caso de (16) en que el verbo aparece sin complementos dentro de la subordinada. Finalmente la R16 declara que no puede haber otro caso que pueda interpretarse como preposición sin término.

El caso de (15) incluye una dificultad extra, pues la cadena ambigua 'sobre' está seguida por una frase también ambigua: 'la mesa', lo que nos obliga a declarar:

R17.AmbPVN + AmbSNSV → AmbSPSV

que implica no resolver la ambigüedad de toda la frase hasta considerar la presencia de otro SV en la oración, para lo cual se activan nuevamente las R7 a R10.

Los resultados de aplicar estas reglas al archivo de prueba son los siguientes:

Tabla 11. Output obtenido con las Reglas de Desambiguación para la cadena 'sobre'

'El mensaje'.
[ 'el mensaje', 'EMS', 'SN', 'EMS', 'D+°N' ].
'está sobre la mesa'.
[ 'sobre lo mesa estar', 'EMS', 'SV+SP' ].
'.
'No'.
[ 'no', 'EMS', 'Adv' ].
'quiero'.
[ 'querer', 'EMS', 'SV', 'EMS', 'V' ].
'que sobre'.
[ 'que sobre', 'EMS', 'OSub', 'EMS', 'Sub+V' ].
'.

### 3.4. Ambigüedad Verbo / Adverbio / Preposición: 'FUERA', 'APARTE'

Los contextos posibles de aparición para estas cadenas serían<sup>6</sup>:

- (17) Ojalá **fuera** invierno. ['Verbo']
- (18) Lo dejó **fuera**. ['Adverbio']
- (19) Lo echó **fuera** de la casa. ['Preposición']
- (20) No lo **aparte** por favor. ['Verbo']
- (21) Lo buscó **aparte**. ['Adverbio']
- (22) No hay nada **aparte** de eso. ['Preposición']
- (23) Si **fuera** de mi país, lo entendería. ['Verbo']
- (24) Pedían que se **fuera** de Argentina. ['Verbo']

<sup>6</sup> No vamos a considerar aquí la diferencia entre 'fuera' como subjuntivo de 'ir' y como subjuntivo de 'ser'. En esta etapa nos interesa sólo el reconocimiento de la posición como verbal.

La Tabla 12 muestra la salida obtenida previa la declaración de reglas específicas de desambiguación:

Tabla 12. Output inicial del analizador para ‘fuera’ y ‘aparte’

<p>'Ojalá'.          [ 'ojalá', 'EMS', 'ij' ].          'fuera'.          [ 'ser', 'EMS', 'SV', 'EMS', 'V' ].          'invierno'.          [ 'invierno', 'EMS', 'SN', 'EMS', '°N' ].          '.          'Lo dejó'.          [ 'lo dejar', 'EMS', 'SV', 'EMS', 'CI+°V' ].  <b>'fuera'</b>.          [ 'ser', 'EMS', 'SV', 'EMS', 'V' ].          '.          'Lo echó'.          [ 'lo echar', 'EMS', 'SV', 'EMS', 'CI+°V' ].  <b>'fuera'</b>.          [ 'ser', 'EMS', 'SV', 'EMS', 'V' ].          'de la casa'.          [ 'de lo casa', 'EMS', 'SP', 'EMS', 'P+D+N' ].          '.          'No'.          [ 'no', 'EMS', 'Adv' ].  <b>'lo aparte'</b>.          [ 'lo apartar', 'EMS', 'SV', 'EMS', 'CI+°V' ].          'por favor'.          [ 'por favor', 'EMS', 'SP', 'EMS', 'P+SN' ].          '.          'Lo buscó'.          [ 'lo buscar', 'EMS', 'SV', 'EMS', 'CI+°V' ].  <b>'aparte'</b>.          [ 'apartar', 'EMS', 'SV', 'EMS', 'V' ].          '.          'No'.          [ 'no', 'EMS', 'Adv' ].          'hay nada'.          [ 'haber nada', 'EMS', 'SV+SN', 'EMS', '°V+°N' ].  <b>'aparte'</b>.          [ 'apartar', 'EMS', 'SV', 'EMS', 'V' ].          'de'.          [ 'de', 'EMS', 'SP' ].          'de eso'.          '.</p>
---

Como se observa, el analizador etiquetó todas las ocurrencias como ‘Verbo’.

Existe además una variante de uso que corresponde a oraciones que indican procedencia u origen:

Al igual que en los casos previos, agregamos en entradas un rasgo de ambigüedad Verbo / Preposición / Adverbio ‘ambVPA’ para las cadenas ‘aparte’ y ‘fuera’. Declaramos las reglas siguientes para MPS:

- R18.ambVPA → AmbVPA [Regla de postergación en Ejec1]
- R19.SV + AmbVPA + SP → SV+SP
- R20.SV + AmbVPA → SV + Adv

R21.SV+SN + AmbVPA + SP → SV+SN+ SP  
 R22.SV+SN + AmbVPA → SV+ Adv  
 R23.AmbVPA → V

Al igual que la R4, la R18 posterga la desambiguación hasta tener información del contexto oracional. Las reglas siguientes actúan cuando este dato ya está disponible en el output, o sea en Ejec2 o Ejec3. R19 y R20 actúa en casos como (18), (19) y (21). R21 opera en casos en los que entre el verbo y el adverbio 'aparte' o 'fuera' se ubica algún tipo de SN, como en (22). R22 se aplica en frases como (25)

(25) No hay nada aparte.

Por último, R23 operará en el resto de los contextos, como en: (17), (20), (23) y (24). El resultado obtenido se detalla en la Tabla 13:

Tabla 13. Output obtenido con las Reglas de Desambiguación para las cadenas 'fuera' y 'aparte'

<p>'Lo dejó fuera'.          [ 'lo dejar fuera', 'EMS', 'SV+Adv' ].          '.          'Lo echó fuera de la casa'.          [ 'lo echar fuera de lo casa', 'EMS', 'SV+SP' ].          '.          'Lo buscó aparte'.          [ 'lo buscar aparte', 'EMS', 'SV+Adv' ].          '.          'No'.          [ 'no', 'EMS', 'Adv' ].          'hay nada aparte de eso'.          [ 'haber nada aparte de eso', 'EMS', 'SV+SN+SP' ].          '.</p>
---

## 6. EVALUACIÓN DE RESULTADOS

Por último, se procedió a realizar una primera evaluación de cómo operan las reglas declaradas en un sub-corpus del Corpus Infosur [6]. Sobre un total de 137 cadenas ambiguas, al momento los resultados obtenidos son

Precisión: 94%

Cobertura: 57.66%

El descenso de la cobertura se debe a que en la salida de la última ejecución de MPS se detectaron casos que no se han considerado aquí: la presencia de incisos y marcadores discursivos, los contextos oracionales de subordinación, y signos de puntuación tales como comillas y comas, interrumpen algunas frases y dificultan su tratamiento, como así también contextos de negación. En todos estos casos, el analizador no les asigna etiquetas erróneas, sino que aún las etiqueta como ambiguas.

## 7. CONCLUSIONES Y PERSPECTIVAS DE DESARROLLO

A medida que avanza nuestra investigación, la tarea de desambiguación va limitando cada vez más el número de ambigüedades.

La declaración de 23 reglas permitió desambiguar cierto tipo de cadenas que se caracterizan por tener más de un uso en el lenguaje natural, resultando efectiva en casos en que otras herramientas, sobre todo las que utilizan métodos estadísticos, otorgan salidas de análisis inadecuadas.

Sin duda, la mayor complejidad oracional requerirá reglas que abarquen nuevos contextos. Los coordinantes y subordinantes, en este caso, resultarán un elemento crucial para el etiquetado. Los incisos [7] y los marcadores discursivos [8] constituyen también un elemento a considerar en la desambiguación, y se abordarán en etapas futuras de la investigación, como así también las enumeraciones.

Entre los casos no abordados, aparecen secuencias como:

(26) ... para las que...

(27) ... entre los que...

El abordaje de estos ejemplos requiere una de dos alternativas: considerar el inicio de una subordinada o plantear una regla de desambiguación tal como:

R24.AmbPV + Relacionante → Prep+OSub

Incisos tales como “..., entre ellos...”, “..., para mí, ...” etc... puede declararse una regla tal como:

R25.ambVP + PronPersonal → SP

Una opción es declarar reglas que incluyan los signos de puntuación que delimiten incisos y marcadores. Estos contextos se abordarán en próximas etapas de la investigación.

De la misma manera, la negación requiere la declaración de reglas que contemplen la aparición de la cadena ‘no’ en los contextos posibles (esto implicará simplemente duplicar una parte de las reglas declaradas hasta aquí).

Por último, se requerirá una evaluación global de las reglas elaboradas hasta aquí en un corpus mayor.

## Referencias

- [1] Moro, S.M. (2008) “Análisis automático de ambigüedades en español: las categorías ‘nombre’ y ‘verbo’”, en *Infosur* 2:15-26, <http://www.infosurrevista.com.ar>
- [2] Moro, S.M. (2008b) “Análisis automático y ambigüedades en la asignación de la categoría ‘adverbio’ en español”, ponencia al *XV Congreso Internacional de la Asociación de Lingüística y Filología de América Latina (ALFAL)*, Montevideo.
- [3] Moro, S.M. (2008c) “Análisis automático y ambigüedades en la asignación de la categoría ‘verbo’ en español”, ponencia al *XI Congreso de la Sociedad Argentina de Lingüística (SAL)*, Coloquio “J.A.L.I.M.I.”, U. N. del Litoral, Santa Fe.

- [4] Aït-Mokthar S. (1998) *L'analyse présyntaxique en une seule étape*. Tesis doctoral dirigida por Gabriel G. Bès en el GRIL, Université Blaise-Pascal, Francia, 1998.
- [5] Abbaci F. *Développement du Module Post-Smorph*. Memoria del DEA de Linguistique et Informatique. Universidad Blaise-Pascal/GRIL, Clermont-Fd, 1999.
- [6] Este corpus está compuesto por textos periodísticos de diarios publicados en Argentina en los últimos años. Para la primera etapa de evaluación del etiquetado se consideró un subconjunto del corpus que asciende a un total de 10.190 palabras.
- [7] Koza, W.A. (2008) “Análisis automático de textos: Reconocimiento de incisos”, en *Infosur* 2:73-84
- [8] Koza, W.A. () “Marcadores discursivos del español. Descripción y propuesta de detección automática”, en *Revista de Epistemología y Ciencias Humanas*, Nº 3, en <http://www.revistaepistemologi.com.ar/biblioteca/11.KOZA.pdf>