

Análisis automático de textos: Reconocimiento de incisos

Automatic Text Analysis: Recognition of Inserts

Walter Adrián Koza

Grupo INFOSUR- UNR- Becario de CONICET

Rosario, Argentina

walter_koza@yahoo.com.ar

Abstract

This paper is part of a Ph.D. Thesis in progress within the Infosur Project in Computational Linguistics. Its goal is to work on the analysis of inserts. The insert is a construction inserted in a clause, which breaks the construction of the phrase. This category includes two-member appositions, vocatives, secondary information and explanations of a heterogeneous nature. The first part of the paper proposes a classification of this type of construction that distinguishes those inserts requiring an antecedent from those which do not, while considering, at the same time, their likely occurrences in the clause. Next we present the development of the computational treatment for the recognition of two-member appositions. With that purpose we conducted a linguistic modeling that allowed the automatic analysis and resorted to Smorph and Mps (Post Smorph Module). Smorph is an analyzer and a generator that, in a single stage, conducts the delimitation and morphological analysis of the target segments. The MPS software (Post-Smorph Module) allows the recognition of appositions in the texts, through the application of specific rules.

Key words: Automatic analysis, Comma, Punctuation, Insert, Apposition.

Resumen

El presente trabajo es parte de la tesis doctoral que en estos momentos me encuentro realizando y se enmarca dentro del proyecto Infosur de Lingüística computacional. Se pretende abordar el análisis de los incisos. El inciso es una construcción inserta dentro de una cláusula y es un elemento que rompe la construcción de la frase. Se incluyen en esta categoría las aposiciones bimembres, los vocativos, las informaciones secundarias y explicaciones de diversa índole. En la primera parte del trabajo, se propone una clasificación de este tipo de construcciones en la que se diferencia a aquellos incisos que requieren un antecedente de los que no, a la vez que se consideran sus posibilidades de ubicación en la cláusula. Posteriormente, se presenta el desarrollo del tratamiento informático para el reconocimiento de las aposiciones bimembres. A tales efectos, se realizó una modelización lingüística, que permitió el análisis automático y se recurrió a los softwares Smorph y Mps (Módulo Post Smorph). Smorph es un analizador y generador textual que, en una única etapa, realiza la delimitación previa de los segmentos a considerar y el análisis morfológico. El software MPS (Módulo Post-Smorph) permite reconocer, mediante reglas específicas, la presencia de aposiciones en los textos.

Palabras claves: Análisis automático, Coma, Puntuación, Inciso, Aposición.

1. INTRODUCCION

Se presenta en este artículo, una aproximación al reconocimiento de incisos en el ámbito de la lingüística informática. Este trabajo forma parte de la tesis doctoral que en estos momentos estoy realizando sobre el análisis de los signos de puntuación, con la dirección de la doctora Zulema Solana y la investigación está subvencionada por una beca otorgada por el CONICET.

Los antecedentes que se han tenido en cuenta remiten, por un lado, a los estudios realizados sobre la puntuación y, por otro, a la utilización de formalismos y softwares declarativos. El reconocimiento de los incisos forma parte del análisis de las funciones de la coma en la medida en que una de las tareas de esta marca de puntuación es, precisamente, delimitar tales construcciones.

Las comas, al igual que los paréntesis y las rayas, permiten la ubicación de incisos y funcionan como delimitadores de éstos. Si bien no se ha encontrado una definición unívoca para este término, llamado también frase extracolocada por algunos autores como Boula de Mareüil y Maillebau [1], en rasgos generales, puede señalarse que se trata de una construcción *inserta* dentro de una cláusula y es un elemento que rompe la construcción de la frase. El contorno melódico aísla los incisos dentro del enunciado global y en la lectura en voz alta debe indicarse con un tono más grave que el del resto del enunciado. El inciso remite a indicaciones adjuntas como ser el nombre de la persona que se está expresando (discurso directo) o comentarios. Se incluyen en esta categoría las aposiciones explicativas, los vocativos, las informaciones secundarias y explicaciones de diversa índole.

En lo que atañe al trabajo informático, se va a ilustrar el reconocimiento automático de incisos, presentando un caso de balizamiento de aposiciones explicativas. Para ello, se recurrió a los programas Smorph y el Módulo Post Smorph (MPS). Smorph es un analizador y generador textual que en una única etapa realiza la delimitación previa de los segmentos textuales a considerar (*tokenización*) y el análisis morfológico (*lematización*) dando como resultado las formas correspondientes a un lema con los valores correspondientes. Este programa es una herramienta declarativa, y la información utilizada está separada de la maquinaria algorítmica. Esto hace que se la pueda adaptar al uso que quiera darse, ya que con el mismo software se puede tratar cualquier lengua si se le cambia la información lingüística.¹

Con respecto a MPS (Módulo Post-Smorph), este software ha sido especificado en el GRILL por Caroline Hagège, José Rodrigo, Gabriel Bès y Faizza Abacci, e implantado en C++ en un contexto de Windows por Faiza Abacci. Posteriormente, fue extendido en Pasmó (cf. Caroline Hagège, Nuno Namede & Joana Paulo, Pasmó-pós análise morfológica, Technical report, INESC-ID, Lisboa), en donde se le adicionaron otras funcionalidades.

MPS realiza tratamientos previos a los de la sintaxis general de la oración, con el objetivo de normalizar la entrada de la sintaxis estándar, como ser fechas, cantidades, cuestiones relativas a la sufijación y prefijación, el tratamiento de los clíticos y de las contracciones.

Este programa, al igual que SMORPH, también es una herramienta declarativa, con la que, mediante ciertas reglas, se pueden expresar los valores de entradas (sobre dos o más estructuras de datos de la salida de Smorph) y los valores de salida sobre la estructura reagrupada.

En las indicaciones sobre el trabajo de MPS, Gabriel Bès y Zulema Solana explican:

¹ Cf. Aït Mokthar [2].

“MPS compara una *oa* [ocurrencia analizada] en la entrada a analizar con el primer elemento de *c/una* de sus reglas. Si encuentra una regla cuyo primer elemento subsume el *oa*, verifica el *oa* siguiente con el elemento siguiente en la entrada de la regla, y así siguiendo. MPS aplica la primer regla en cuya entrada todos los elementos subsumen un *oa* en la entrada a analizar.” [3]

En primer lugar, se presentarán algunos lineamientos teóricos generales respecto del inciso y, posteriormente, se hará una demostración del reconocimiento de aposiciones bimembres en textos a partir de la utilización del software MPS.

2. ALGUNAS CONSIDERACIONES TEÓRICAS ACERCA DEL INCISO

2.1. Estado de la cuestión

El inciso ha sido estudiado desde diversos puntos de vista. De acuerdo con lo que plantean Asuaje, Blondet, Mora y Rojas [4], pueden encontrarse análisis sobre él, no sólo en la gramática tradicional, sino también en la sintaxis de la lengua oral, el análisis del discurso, la informática y las tecnologías de síntesis y reconocimiento del habla. Los autores que se han consultado para la elaboración del presente trabajo entablan una estrecha relación entre el inciso y la pausa fónica o el cambio de entonación.

Sobre esta cuestión, Marta Payá Canals [5], en su estudio sobre las cláusulas incidentales en el catalán hablado, señala que los incisos se producen a menudo entre silencios y en un registro más bajo que la elocución principal en la que se insertan. Se pronuncian con menor intensidad y generalmente se expresan con un tono de voz descendente final, pero las construcciones que poseen un valor de modalización, tienden a finalizar en un tono alto.

En lo que atañe al español, Asuaje, Blondet, Mora y Rojas observan que el hablante interrumpe su línea discursiva de base para introducir información incidental, considerada usualmente como una digresión:

“(...) esta interrupción es señalada prosódicamente mediante cambios en la velocidad de la elocución y la entonación. Consideramos que una mayor velocidad de elocución y un descenso del promedio de frecuencia fundamental de la voz del hablante, en relación con su línea discursiva de base, constituyen dos indicadores prosódicos de la codificación de información incidental en el habla espontánea.” [4]

El inciso se encuentra codificado de manera prosódica a partir de un aumento de la velocidad de elocución y un descenso de la entonación con respecto a la línea discursiva de base.

Ya en el ámbito de la lingüística computacional, Boula de Mareüil y Maillebauu [1] proponen un método de captura automática de los incisos en francés, a la vez que presentan un modelo prosódico específico para la síntesis de la palabra a partir del texto.

A partir de la definición de inciso que propone Grevisse, los autores mencionados distinguen los ‘incidentes’ de los ‘incisos’. Incidente es la subfrase insertada en el interior de otra frase superior pero que no cumple el rol de sujeto o complemento. Por su parte, el inciso es un tipo especial de incidente que indica que se traen las palabras o pensamientos de alguien; se colocan dentro de la cita o después de ésta. Cf. Boula de Mareüil y Maillebauu [1].

Los incisos son capturados automáticamente de acuerdo con criterios léxicos, sintácticos y *puntuacionales*. Para ello, elaboraron un método de expresión regular y heurística que se apoya en elementos tales como comillas, comas, verbos discursivos, etcétera.

Por otro lado, también presentan un modelo prosódico en el que se reconocen los incisos en la lectura en voz alta de un texto a partir de las variaciones en la entonación y la modulación.

2.2. Clasificación de los incisos

A diferencia de los estudios mencionados, en este trabajo no se tuvieron en cuenta las características de la oralidad, sino que por el contrario, el estudio se basa, además de criterios sintácticos, en elementos de puntuación. Un elemento a considerar en el reconocimiento automático es la posibilidad de que la cláusula incidental se halle delimitada por comas o alguna otra marca, como ser los paréntesis o las rayas.

Con el propósito de presentar una descripción exhaustiva de los incisos, se distinguió a aquellos que requieren de un antecedente, como ser las aposiciones y demás construcciones predicativas, de los que no lo requieren, tal el caso de los marcadores discursivos, las interjecciones, etcétera.

Por otro lado, se tuvieron en cuenta las distintas posiciones que pueden ocupar en la cláusula en la que se hayan insertos. De este modo, se observan incisos que sólo pueden aparecer en determinado lugar; como se el caso de la aposición que se ubica inmediata al sintagma nominal que modifica y otros que poseen una mayor movilidad, como ser el caso de las interjecciones.

De este modo, las construcciones incidentales se clasifican en dos grupos; el grupo I, que incluye a los incisos con antecedentes, y el grupo II, que reúne a las construcciones incidentales sin antecedente.

2.2.1. Grupo I: Incisos con antecedentes

En este grupo se encuentran las aposiciones bimembres, las construcciones nominales no apositivas, las construcciones adjetivales, preposicionales y adverbiales.

2.2.1.1. Construcciones nominales apositivas

Con respecto a la aposición –del latín *apositio*, derivado de *apponere*, equivalente a “colocar junto a”–, ésta remite a las construcciones en las que un sustantivo explica o especifica el significado de otro sustantivo, sin que medie entre ellos una preposición.

Las aposiciones introducen una predicación secundaria de carácter nominal usada para especificar o explicar otra palabra de la misma especie y pueden ser *especificativas* (o *restrictivas*) o *explicativas* (o *no restrictivas*). Asimismo, y de acuerdo con la existencia o no de pausas recibirán el nombre de *unimembres* cuando conformen un solo grupo fónico y *bimembres*, cuando las aposiciones aparezcan segmentada (generalmente por comas) en dos².

Las aposiciones bimembres pueden dividirse, a su vez, en dos subgrupos:

Grupo A

² Véase Suñer Gratacós [6].

Nombre propio + Nombre común

(1) [Juan Manuel de Rosas, el Restaurador de las leyes, fue vencido en la batalla de Caseros.]

Nombre común + Nombre propio

(2) [La presidenta, Cristina Fernández de Kirchner, se reunió con su par brasileño.]

Nombre común + Nombre común

(3) [El león, el rey de la selva, se aparee diecisiete veces al día.]

Estas aposiciones se caracterizan en que, generalmente, sus miembros son sintácticamente intercambiables, el núcleo puede ser aposición y viceversa. No obstante, el término apositivo debe aparecer inmediatamente al sintagma que modifica.

Grupo B

Pronombre + Nombre

(4) [Tú, razón de mis desvelos, no me castigues con la indiferencia.]

Infinitivo + Nombre

(5) [Dormir, el placer más sublime, me ha sido vedado.]

Nombre + Subordinada completiva

(6) [Pretendo una sola cosa, que no me molesten.]

En este caso, las posibilidades de intercambio son más reducidas. Al igual que las del grupo A, la aposición debe aparecer inmediata al antecedente.

2.2.1.2. *Construcciones nominales no apositivas*

Al igual que las aposiciones bimembres, también se subdividen en dos grupos.

Grupo A: Cuando el inciso es un SN con núcleo adjetival.

(7) [Juan, el pobre, se quedó sin el pan y sin la torta.]

Estas construcciones poseen más posibilidades de ubicación ya que pueden aparecer inmediatos al sintagma nominal que modifican o al final de la cláusula:

(8) [Juan se quedó sin el pan y sin la torta, el pobre.]

Grupo B: Cuando el inciso es un SN sin determinante.

(9) [Alumno ejemplar, Domingo Faustino Sarmiento nunca faltó a clases.]

En este caso, la cláusula puede ir antepuesto al sintagma que modifica, pospuesto o en construcciones tácitas:

(10) [Domingo Faustino Sarmiento, alumno ejemplar, nunca faltó a clases.]

(11) [Alumno ejemplar, nunca faltaba a clases.]

2.2.1.3. Construcciones adjetivales

Pueden ser de dos tipos:

Cláusulas relativas explicativas.

(12) [Juan, que era tonto, se quedó sin el pan y sin la torta.]

Si bien no necesariamente deben aparecer inmediatas al antecedente, no pueden encontrarse muy alejadas de éste.

(13) [Juan, el marido de Luisa, que era tonto, se quedó sin el pan y sin la torta.]

Cláusulas participiales:

(14) [Vestido para la ocasión, el novio llegó puntual.]³

Tienen gran movilidad dentro de la cláusula; pueden aparecer antes o después del sintagma que modifican y al final de la cláusula:

(15) [El novio, vestido para la ocasión, llegó puntual.]

(16) [El novio llegó puntual, vestido para la ocasión.]

2.2.1.4. Construcciones preposicionales y adverbiales

Tanto en las construcciones preposicionales como en las adverbiales, las cláusulas deben aparecer inmediatas al antecedente.

(17) [Se casó con quien menos imaginábamos, con su prima.]⁴

(18) [Las cosas son así, como yo digo.]

2.2.2. Grupo II: Incisos sin antecedentes

Está conformado por incisos interruptores, marcadores discursivos, los verbas dicendi y las interjecciones y locuciones interjectivas.

³ Ejemplo tomado de Suñer Gratacós.

⁴ Ídem anterior.

2.2.2.1. Incisos interruptores

Son construcciones que interrumpen la oración insertándose en la cláusula. Pueden ser de dos tipos:

Incisos Interruptores del orden canónico:

(19) [Pensé, como siempre, que iba a perder.]

Incisos interruptores de un sintagma núcleo

Los sintagmas núcleos son sintagmas conformados por categorías morfosintácticas, las que posibilitan determinar dónde comienzan, dónde terminan, cómo están compuestos y cuál es su núcleo. A partir de determinadas propiedades de linealidad, se restringen las posibilidades combinatorias de sus elementos. La cualidad principal de estos sintagmas es que finalizan en su núcleo. Un sintagma núcleo es un bloque casi inseparable, y esto hace que su extracción no presente demasiados problemas. Zulema Solana y Andrea Rodrigo [7] proporcionan los siguientes ejemplos de los distintos sintagmas núcleos que pueden encontrarse en el español:

- *no los ha visto* (sintagma verbal núcleo –svn–);
- *las lejanas playas* (sintagma nominal núcleo –snn–);
- *muy hermoso* (sintagma adjetival núcleo –sadjn–).

Abney [8] justifica la noción de sintagma núcleo por razones prosódicas y psicolingüísticas por un lado y a la vez porque permiten un análisis automático del texto con menores dificultades. A estas construcciones, las denomina *chunks* y a partir de Aït Mokhtar y Chanod [9] se han hecho investigaciones de análisis sintáctico automático utilizando la plataforma XIP que recurre al análisis en *chunks* o sintagmas núcleos

Se considera que el análisis en este tipo de sintagmas permite reducir significativamente la ambigüedad de la categorización morfosintáctica al concatenar las expresiones internas. Por ejemplo, puede señalarse en el caso de *no los ha visto* que *los* no puede ser artículo.

A tales efectos, un inciso interruptor de un sintagma núcleo se ubica entre el comienzo y el núcleo del sintagma:

(20) [Pedro ha, probablemente, engañado a su esposa.]

2.2.2.2. Marcadores discursivos

Son construcciones invariables que no ejercen una función sintáctica en el marco de la predicación oracional –se trata de elementos marginales– y la función es guiar, de acuerdo con sus distintas propiedades morfosintácticas, semánticas y pragmáticas, las inferencias que se realizan en la comunicación. (Zorraquino y Portolés, 1999)

Ejemplos de marcadores discursivos:

(21) [Se casó con quien menos imaginábamos, es decir, con su prima.]

(22) [Juan descubrió que su esposa le era infiel, por ende, le pidió el divorcio.]

Por lo general, tienen gran movilidad dentro de la cláusula. No obstante, pueden darse casos en los que las posiciones estén determinadas. Por ejemplo, una construcción como “en primer lugar”, cuando cumple la función de marcador discursivo no puede aparecer al final de la cláusula.⁵

2.2.2.3. *Verbas dicendi*

Remite a las construcciones de discurso directo. En este caso, el *verba dicendi*, cumple la función de verbo principal de la oración, como puede verse en el ejemplo.

(23) [“¿Sabés una cosa”, dice ella, “sos un imbécil: si te digo que vivo ahí es porque vivo ahí”.]
(Leo Maslíah, “La cita”)

La cláusula principal está constituida por lo que dice el personaje y se indica a través del uso de comillas. El inciso ‘dice ella’, contiene al sujeto y al verbo del predicado.

Otros verbos que presentan el mismo fenómeno: ‘imaginar’, ‘pensar’, ‘asegurar’, etcétera.

(24) [Ramona está contenta, me imagino, por el regreso de su novio.]

(25) [“Cuando menos lo esperen”, pensaba Carolina, “junto mis cosas y me mando a mudar.]

(26) [Para el viernes voy a tener todo el trabajo terminado, te lo aseguro.]

2.2.2.4. *Interjecciones y locuciones interjectivas*

Este tipo de construcciones puede ocupar diversas posiciones en la cláusula.

(27) [No sé, ¡ay de mí!, cuánto tiempo más voy a poder soportarlo.]

En el párrafo siguiente, se presentará un acercamiento a la implantación en máquina y el reconocimiento de las aposiciones bimembres a través del software Smorph y Mps.

3. IMPLANTACIÓN EN MÁQUINA

Se trabajó con un texto de diez mil palabras de extensión. Para el análisis de las aposiciones explicativas, se procedió, en primer lugar, a realizar el análisis morfológico de las palabras y el reconocimiento de los signos de puntuación. En segundo lugar, se elaboraron las reglas de Mps para conformar los sintagmas nominales (de nombre comunes y de nombres propios) de las construcciones apositivas. Posteriormente se determinaron las reglas para su reconocimiento.

3.1. Análisis morfológico y reconocimiento de signos de puntuación

En la primera etapa del trabajo, se procedió a determinar los elementos textuales a considerar. Con el programa Smorph se obtuvo el análisis morfológico de cada término y el reconocimiento de los signos de puntuación.

⁵ Para el tratamiento informático de los marcadores discursivos, Cf. Prada [10].

A modo de ejemplo, dado el siguiente fragmento:

En un clima tenso, Carlos Juárez se negó a declarar.

Smorph dio como resultado:

'En'.

['en', 'EMS','prep'].

'un'.

['un', 'EMS','det', 'TINDF1','indf1a'].

'clima'.

['clima', 'EMS','nom', 'GEN','masc', 'NUM','sg'].

'tenso'.

['tenso', 'EMS','adj', 'GEN','masc', 'NUM','sg'].

['tensar', 'EMS','v', 'MODOV','ind', 'PERS','1a', 'NUM','sg', 'TPO','pres', 'TR','r', 'TC','c1', 'TDIAL','estrpi'].

','.

['cc', 'EMS','coma'].

'Carlos'.

['Carlos', 'EMS','npr'].

'Juárez'.

['Juárez', 'EMS','npr'].

'se'.

['lo', 'EMS','cl', 'TPCRF','rflse'].

'negó'.

['negar', 'EMS','v', 'MODOV','ind', 'PERS','3a', 'NUM','sg', 'TPO','prets', 'TR','r', 'TC','c1', 'TDIAL','estrpi'].

'a'.

['a', 'EMS','prep'].

'declarar'.

['declarar', 'EMS','v', 'MODOV','infin', 'TR','r', 'TC','c1'].

'

'.

['linsig', 'EMS','pun'].

3.2. Reconocimiento de sintagmas nominales

Una vez obtenidos los resultados de Smorph, se constituyeron las reglas de reconocimiento de sintagmas nominales. El texto analizado contenía veintinueve aposiciones bimembres, todas ellas

conformadas por un nombre propio más un sintagma nominal común, o un sintagma nominal común más un nombre propio. A tales efectos, se distinguieron dos tipos de sintagmas nominales:

- Sintagma Nominal (SN);
- Sintagma Nominal de Nombre Propio (snnpr).

3.2.1. Sintagmas nominales comunes

En este caso, puede tratarse de un sintagma nominal núcleo (snn) o no, cuando se extiende más allá de su núcleo. Es decir, se le pueden adicionar al núcleo diversos modificadores como ser un sintagma adjetivo núcleo (sadjn⁶), un sintagma preposicional (sp⁷) o un sintagma adjetivo núcleo más un sintagma preposicional. Gráficamente, el SN ofrece las siguientes variables:

- snn (*el libro*)
- snn + sadjn (*el libro nuevo*)
- snn + sp (*el libro de cocina*)
- snn + sadjn + sp (*el libro nuevo de cocina*)

3.2.2. Sintagmas nominales con nombres propios

Los sintagmas nominales con nombres propios estaban conformados de la siguiente manera:

- Un nombre propio (o varios) más un apellido: *José Manuel de la Sota*;
- Un SN más un nombre (o varios) más un apellido: *El abogado Juan Manuel López*;
- Un artículo más un nombre propio: *La corte suprema*.

3.3. Reconocimiento de aposiciones

Finalmente y una vez definidos estos dos tipos de sintagmas nominales (los SN y los snnpr) se realizaron las reglas de reconocimiento de aposiciones. Las aposiciones explicativas estaban conformadas de la siguiente manera:

- SN + coma + snnpr + coma: *El ex fiscal, Carlos Lanuse,*
- SN + coma + snnpr + punto y coma: *El ex fiscal, Carlos Lanuse;*
- SN + coma + snnpr + punto: *El ex fiscal, Carlos Lanuse.*
- SN + coma + snnpr + dos puntos: *El ex fiscal, Carlos Lanuse:*
- snnpr + coma + SN + coma: *Carlos Lanuse, el ex fiscal,*
- snnpr + coma + SN + punto y coma: *Carlos Lanuse, el ex fiscal;*
- snnpr + coma + SN + punto: *Carlos Lanuse, el ex fiscal.*

⁶ Se reconocen a partir de reglas similares a las del snn.

⁷ Formados por una preposición más un snn, un snn más un sadjn o más un nombre propio.

- snnpr + coma + SN + dos puntos: *Carlos Lanuse, el ex fiscal:*

A manera de ejemplo, pueden observarse las siguientes aposiciones reconocidas por Mps:

'el ex gobernador santiagueño , Carlos Juárez ,'

['el ex gobernador santiagueño cc Carlos Juárez cc', 'EMS', 'aposición'].

(...)

'China , el país de moda ,'

['China cc el país de moda cc', 'EMS', 'aposición'].

(...)

'el gobernador de Santa Cruz , Sergio Acevedo .'

['el gobernador de Santa_Cruz cc Sergio Acevedo pfp', 'EMS', 'aposición'].

Los resultados obtenidos arrojaron un 90% de precisión y un 93,10% de cobertura.

4. CONSIDERACIONES FINALES

En la primera parte del artículo se trajeron a colación algunas consideraciones teóricas sobre el inciso para luego proponer una clasificación de éstos en la que se diferenciaba a aquellos incisos que requerían antecedentes de los que no, considerando al mismo tiempo sus posibilidades de ubicación en la cláusula.

Posteriormente, se presentó el desarrollo del tratamiento informático a partir del análisis morfológico de Smorph y el reconocimiento sintagmático con Mps. A partir de allí, se elaboraron las reglas de reconocimiento de aposiciones.

Sobre un total de veintinueve aposiciones que tiene el texto analizado, se reconocieron veintisiete y se señalaron de manera errónea tres. Esto implica un 90% de precisión y 93,10% de cobertura.

Actualmente, la investigación está focalizada en continuar con el tratamiento informático a partir de las demás funciones y usos de la coma y completar el tratamiento informático a partir de los softwares descriptos y con el programa Xfst (Xerox Finite State Tools).

Referencias

- [1] Boula de Mareüil, P. y Maillebau, E. ; 2002, "Traitement des incisives en français : capture automatique et modèle prosodique", en *XXIVèmes Journées d'Étude sur la Parole, Nancy, 24-27 juin*.
- [2] Aït Mokthar, S.; 1995, *SMORPH: Guide d'utilisation. Rapport technique*, Clermont-Fd. : Universidad Blaise Pascal/GRILL.
- [3] Bès, G. y Solana, Z.; 2004, "Análisis morfológico y gramáticas locales: introducción y una aplicación concreta", En I Jornadas Argentinas de Lingüística Informática: Modelización e Ingeniería, Rosario: Facultad de Humanidades y Artes, Universidad Nacional de Rosario.

- [4] Asuaje, R., Blondet, M., Mora, E. *et al*; 2005, “Codificación Prosódica de la Información Incidental en el Discurso Espontáneo: Un estudio de caso”, en *Rev. Vzlan. de Soc. y Ant.* [online]. dic. 2005, vol.15, no.44, p.449-460. Disponible en la Web: http://www2.scielo.org.ve/scielo.php?script=sci_arttext&pid=S0798-30692005000300010&lng=es&nrm=iso.
- [5] Payá, M.; 2002, “Incidental Clauses in Spoken Catalan: Prosodic Characteristics and Pragmatic Function”, en B. Bel i I. Marlien (eds.), *Proceedings of Speech Prosody 2002 Conference*, Aix-en-Provence, Laboratoire Parole et Langage, pp. 559-562.
- [6] Suñer Gratacós, A.; 1999, “La aposición y otras relaciones de la predicación en el sintagma nominal”, en Bosque, I. y Demonte V. (Dirs.), *Gramática descriptiva de la lengua española, Tomo I*, Madrid: Espasa Calpe.
- [7] Solana Z. y Rodrigo A.; 2005, “El sintagma nominal núcleo”, en Castel Víctor (comp.) *Desarrollo, Implementación y utilización de modelos para el procesamiento automático de textos; Trabajos de la II Jornadas Argentinas de Lingüística Informática: Modelización e Ingeniería*, Mendoza: Facultad de Filosofía y Letras, Universidad Nacional de Cuyo.
- [8] Abney, S.; 1991, *Parsing by Chunks*, en Berwick et al., *Principle-Based Parsing*, Dordrecht: Kluwer Academic Publishers.
- [9] Aït Mokthar, S. y Chanod J.; 1997, “Incremental Finite-State Parsing”, en *Proceedings of the 8th Conference on Applied Natural Language Processing, ANLP-97*, Washington, 72-79.
- [10] Prada, J.; 2001, *Marcadores del discurso en español –análisis y representación*. Tesis de Maestría, Universidad de la República, Montevideo, Uruguay.