



Revista Signos

ISSN: 0035-0451

revista.signos@ucv.cl

Pontificia Universidad Católica de Valparaíso
Chile

Koza, Walter

Funciones gramaticales de la coma. Clasificación e implantación computacional

Revista Signos, vol. 47, núm. 86, diciembre-, 2014, pp. 412-434

Pontificia Universidad Católica de Valparaíso

Valparaíso, Chile

Disponible en: <http://www.redalyc.org/articulo.oa?id=157032730004>

- Cómo citar el artículo
- Número completo
- Más información del artículo
- Página de la revista en redalyc.org

redalyc.org

Sistema de Información Científica

Red de Revistas Científicas de América Latina, el Caribe, España y Portugal

Proyecto académico sin fines de lucro, desarrollado bajo la iniciativa de acceso abierto



Funciones gramaticales de la coma. Clasificación e implantación computacional

*Grammatical functions of the comma.
Classification and computational implementation*

Walter Koza

PONTIFICIA UNIVERSIDAD CATÓLICA DE VALPARAÍSO
CHILE
walter.koza@ucv.cl

Recibido: 20-III-2013 / **Aceptado:** 18-IV-2014

Resumen

El objetivo del presente artículo es clasificar los usos de la coma enfocada en el aspecto gramatical y desde la perspectiva de la lingüística computacional. A partir de este objetivo, se exponen algunas consideraciones teóricas sobre dicho signo de puntuación y se establecen las siguientes funciones: (i) función indicadora: señala enumeraciones y elipsis; (ii) función delimitadora: enmarca todo tipo de inciso (aposiciones, vocativos, etcétera), y (iii) función desambiguadora: evita confusiones en expresiones que pueden llegar a tener más de una interpretación. Posteriormente, se realiza una modelización y una implantación computacional de dicha clasificación, a fin de detectar automáticamente las funciones establecidas, mediante los programas *Smorph* y Módulo *Post Smorph* (MPS). *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. Por su parte, MPS toma como *input* el *output* de *Smorph* y analiza la cadena de lemas resultante del análisis morfológico. Se trabajó con un corpus textual compuesto por artículos periodísticos que sumaban un total de 10.000 palabras. Los resultados arrojaron medidas de precisión, cobertura y medida-f lo suficientemente adecuados para confirmar la validez de la clasificación propuesta.

Palabras Clave: Puntuación, coma, funciones gramaticales, análisis automático, lingüística computacional.

Abstract

Objective of present paper is to classify the comma uses focused in grammatical aspects and from a computational linguistic perspective is proposed. From this objective, some theoretical aspects based on grammatical criteria are showed, and the following classification of the comma functions are established: (i) indicator comma: it points enumerations and ellipsis, (ii) bounding comma: it delimits incidental clauses (appositions, vocatives, etcetera), and (iii) comma for disambiguation: it avoids confusion in expressions that could present more than a interpretation. Afterwards, a formalization and a computational implementation are made with the objective of getting a method of automatic detection for comma functions. In relation to the computational work, the software Smorph and Post Smorph Module (MPS) were used. Smorph analyzes the characters chain morphologically, giving an output with the morphological and categorical assignation for each occurrence according to the features declared. The Smorph output is the MPS input, and MPS analyzes the lemmas chain given by the morphological analysis, with composition, decomposition, and correspondence rules declared by the user. It was worked with a corpora composed of journalist articles that had 10,000 words. Results showed measures of precision, recall and f-measure that are adequate enough to confirm the validity of the proposed classification.

Key Words: Punctuation, comma, grammatical functions, automatic analysis, computational linguistic.

INTRODUCCIÓN

La problemática acerca del sistema de los signos de puntuación ha recibido un tardío y, en cierta medida, escaso interés en los estudios lingüísticos, ya sea desde el punto de vista del análisis del discurso, de la relación con la enseñanza-aprendizaje o bien desde el alcance semántico y pragmático de su empleo. Al respecto, Figueras (1997) menciona que una posible causa del exiguo interés que ha despertado entre lingüistas y psicólogos del lenguaje se deba a que tiene la condición de ser un recurso propio de la escritura. Dicha concepción implicaría que, por el solo hecho de constituirse como un aspecto de lo escrito, las condiciones del empleo de los signos de puntuación estarían reguladas por la normativa y, a tales efectos, dicho uso estaría dado por un conjunto de reglas unívocas y uniformes que no presentarían problemas de aplicación. No obstante, es posible apreciar que la utilización de estas marcas no siempre responde a aspectos normativos, sino que, en su uso, subyacen decisiones estilísticas y de significado. La puntuación, señala Figueras (2001), es un sistema de marcas vinculado a la construcción del texto, y eso dificulta la posibilidad de establecer normas.

Al respecto, si bien la RAE (2010) dedica un apartado a la puntuación en la ‘Ortografía de la lengua española’, señala que su uso va más allá de los límites de la ortografía, en la medida en que la mayoría de los signos que componen el sistema

poseen usos no lingüísticos. Allí se define a la puntuación como el conjunto de signos ortográficos que organiza el discurso para facilitar la comprensión de este, evitando ambigüedades y señalando el carácter especial de ciertos fragmentos (RAE, 2010). La puntuación señala segmentaciones gramaticales tales como oraciones, cláusulas, frases y sintagmas, junto con marcas de significado, como, por ejemplo, exclamaciones, que sustentan, clarifican y amplían los mensajes escritos favoreciendo la actividad del lector.

Asimismo, y a pesar del mencionado desinterés, es posible encontrar algunos estudios relevantes en el área, sobre todo en la década del noventa, a partir del trabajo de Numberg (1990), quien desarrolla una lingüística de la puntuación focalizado en las unidades textuales que los signos delimitan. Este trabajo posibilitó la aparición de nuevos estudios sobre el tema, ya sea en el ámbito de la lingüística tradicional (no computacional) como así también en el procesamiento del lenguaje natural (PLN). Dentro de los primeros, se destacan los trabajos de Ehrlich (1992), Hall y Robinson (1996) y, dentro del español, Figueras (2001). Para el caso de los trabajos en el área de la lingüística computacional, pueden mencionarse los llevados a cabo por Briscoe y Carroll (1995), White (1995) y Jones (1996). Sin embargo, se puede notar que, en los últimos tiempos, los estudios sobre puntuación han decrecido notoriamente y, en la actualidad, han aparecido solo unos pocos trabajos que la tratan en relación con la enseñanza y el aprendizaje (Roselló, 2010; Wallace, 2012), o bien desde la perspectiva del análisis del discurso (Rojas, 2011). En lo concerniente al ámbito del PLN, actualmente los trabajos llevados a cabo están direccionados hacia el área de la asignación automática de la puntuación en textos orales transcritos (Chian, Wan & Chen, 2012).

Ahora bien, dentro del sistema de la puntuación, la coma es el signo más dinámico e indica una diversidad de significaciones. Su residencia intersticial establece conexiones de sentido para organizar la comunicación contenida en las palabras escritas (Cipriano, 1976). Desde ya que esta función no es privativa de la coma, puesto que también la cumplen, en mayor o menor grado, los otros signos de puntuación, sin embargo, su asiduidad le otorga un lugar preponderante.

El objetivo del presente trabajo es clasificar los usos de la coma enfocada en el aspecto gramatical y desde la perspectiva de la lingüística computacional. A partir de este objetivo, se exponen algunas consideraciones teóricas sobre este signo de puntuación y se establecen las siguientes funciones: (i) función indicadora: señala enumeraciones y elipsis; (ii) función delimitadora: enmarca construcciones incidentales (alteraciones en el orden regular, aposiciones, subordinadas relativas explicativas, etcétera), y (iii) función desambiguadora: evita confusiones en construcciones que pueden llegar a presentar más de un significado, como por ejemplo:

(1) No debes hacerlo.

(2) No, debes hacerlo.

En el primer caso, se da a entender que alguien no debe hacer algo, mientras que en (2), con las mismas palabras, se indica lo contrario. Posteriormente, para validar dicha clasificación, se realiza una modelización y una implantación computacional a fin de detectar automáticamente las funciones establecidas. Para ello, se recurrió a los *softwares Smorph* (Äit-Mokthar, 1998) y Módulo *Post Smorph* (MPS) (Abacci, 1999). Los resultados obtenidos fueron evaluados mediante las medidas de precisión, cobertura y medida F.

1. Estudios sobre el sistema de la puntuación

Los estudios sobre la puntuación se enmarcan, fundamentalmente, en dos perspectivas. La primera de ellas se focaliza en el aspecto prosódico, es decir, a partir de la relación de los signos con las pausas de la oralidad y la entonación. Dentro de este ámbito, se pueden apreciar los trabajos de Luque (2006) y Zamudio (2004). En dichos estudios se describe la manera en que surge el sistema de los signos de puntuación en la escritura, como representación de la prosodia de las frases, y en relación con factores tonales, duracionales, de intensidad, rítmicos, etcétera, en unidades mayores a la palabra (Luque, 2006).

No obstante, más allá de que la puntuación haya surgido como una herramienta para reproducir los elementos prosódicos de la lengua oral (Figueras, 2013), es posible observar un desplazamiento de esta función a la de organizar el discurso a fin de facilitar su comprensión. De este modo, la segunda perspectiva de análisis se relaciona con el significado textual, en la medida que cumple un papel determinante en la distribución de los elementos verbales del texto a fin de garantizar una correcta interpretación. Así, la puntuación se fundamenta en criterios gramaticales y pragmáticos en la medida en que brinda información respecto de la identificación y la jerarquización de las unidades lingüísticas, la modalidad de los enunciados y la omisión de algunos elementos. Si bien estas cuestiones se manifiestan también en la oralidad, hay que observar que la organización de la cadena oral responde, además, a condicionamientos propios que no comparten con la cadena escrita y viceversa (RAE, 2010).

En relación con este segundo punto de vista, hay que mencionar que la principal motivación por la que algunos investigadores dirigieron sus estudios al terreno de la puntuación fue la aparición, en 1990, de la monografía de Nunberg (1990), *The linguistics of punctuation*. En ella, el autor ataca la opinión generalizada de que la puntuación es prescriptiva y que únicamente sirve para reflejar la entonación en la escritura. Por el contrario, señala que, después de la divergencia entre el lenguaje escrito y la oralidad, la puntuación ha devenido en un sistema lingüístico con sus propias características. Los planteos de Nunberg (1990) han sido el punto de partida para numerosas investigaciones. Dentro de ellas, y para el caso del español, se puede mencionar el trabajo de Figueras (2001), quien, enmarcada en la teoría de la relevancia, propone un

enfoque pragmático para desarrollar una gramática de los signos de puntuación. Para Figueras (2001), la puntuación es considerada un mecanismo cuya función principal es controlar eficientemente la interpretación del lector.

Con respecto a los estudios realizados en el marco de la lingüística computacional, la caracterización de Nunberg (1990) dio inicio a diversos trabajos. Entre ellos, señala White (1995), se pueden apreciar dos ramas: una perteneciente al terreno de la síntesis, cuyo interés está enfocado en la manera de formalizar las funciones discursivas de la puntuación, y otra que se centra en el análisis en donde el objetivo es producir mejoras en el análisis robusto.

1.1. Primeras investigaciones. La lingüística de la puntuación de Nunberg (1990)

Nunberg (1990) asocia la puntuación a otros recursos gráficos y de diseño que se utilizan para dar cuenta del sistema de categorías textuales. No obstante, argumenta además que se trata de un subsistema lingüístico y, por lo tanto, debe considerársela como parte del vasto sistema del lenguaje escrito. Las reglas de puntuación son verdaderamente complejas y sistemáticas, no accesibles a la introspección casual, y se organizan a lo largo de líneas que sugieren la influencia de principios de organización que subrayan la estructura de otros sistemas de lenguaje natural.

“The term ‘punctuation’ is generally used to refer to a category defined in partially graphic terms: a set of non-alphanumeric characters that are used to provide information about structural relations among elements of a text, including commas, semicolons, colons, periods, parentheses, quotations marks and so forth. From the point of view of function, however, punctuation must be considered together with a variety of other graphical features of the text, including font-and-face-alternations, capitalization, indentation and spacing, all of which can be used to the same sorts of purposes” (Nunberg, 1990: 17).

Debe notarse, sin embargo, que los indicadores de categoría no pueden ser clasificados solo sobre la base de sus propiedades gráficas, puesto que el mismo recurso gráfico podría tener múltiples funciones. Por ejemplo, el punto funciona tanto como delimitador de oración, como marcador de una abreviación lexical y una mayúscula inicial sirve tanto para delimitar oraciones como para marcar nombres propio. Asimismo, la clasificación de un recurso gráfico particular puede depender del género particular de escritura en el que aparece. Así, una nueva línea sirve como una indicadora de categoría en listas (donde separa elementos de la categoría ‘ítem’) y en un poema (donde separa elementos de la categoría ‘línea’). Pero las cadenas de caracteres separadas por saltos de línea en la prosa corrida no son elementos de ninguna categoría textual.

Nunberg (1990) plantea que el sistema de indicadores de categorías textuales es una ‘aplicación’ del lenguaje natural, lo que significa que surge a través de un

tipo de aplicación de los principios de diseño de dicho lenguaje, como medio de imponer una cierta organización en el contenido léxico del texto escrito. Visto en profundidad, el sistema no tendría análogo en el lenguaje oral, puesto que es producto de la estandarización e intelectualización de la escritura, como una respuesta a los requerimientos comunicativos particulares de los textos de lenguaje escrito y como un aprovechamiento de los recursos expresivos propios que la presentación gráfica hace posible.

Asimismo, este autor distingue dos clases de gramática diferentes, “o, si se prefiere, dos niveles de descripción gramatical distintos” (Nunberg, 1990: 19), relevantes para determinar la distribución de elementos explícitos en la escritura. A la primera de ellas la denomina gramática lexical y a la segunda, gramática textual. La gramática lexical de la escritura es la responsable de describir las dependencias existentes entre los ítems lexicales del texto. Se trata de la misma clase de sistema que los lingüistas llaman generalmente gramática, en relación con el lenguaje hablado, pero, según el autor, habría ciertas diferencias a tener en cuenta. En primer lugar, la gramática lexical de un lenguaje escrito particular podría contener palabras específicas y construcciones que no se dan en la variedad oral y viceversa. Más importante, la gramática lexical de la escritura podría contener categorías morfológicas, semánticas o sintácticas que deben ser explícitamente marcadas en la escritura, pero que no están tratadas distintivamente en la oralidad. Por ejemplo, la categoría expresión propia (nombres propios), expresadas con mayúsculas, no tiene equivalente, claro está, en la oralidad. La gramática textual, por su parte, remite a las reglas que describen la distribución de ciertas categorías explícitamente marcadas (como, por ejemplo, el párrafo o la cláusula textual) que clasifican el papel del contenido de los componentes léxicos en relación con una determinada estructura de argumento y contexto de interpretación. Las condiciones de buena formación impuestas por la gramática textual pueden ser establecidas en modos que son independientes del contexto de las categorías lexicales, aunque la relación entre las dos gramáticas es compleja. Con respecto a ello, Nunberg (1990) señala que ciertas clases de categorías del lenguaje escrito, tales como las expresiones parentéticas, deben considerarse categorías tanto de la gramática textual como de la lexical, y más aún, que ellas están sujetas a diferentes restricciones e interpretaciones indistintamente del nivel en el que se las introduce.

Aquí, el término gramática es entendido como un conjunto de reglas que determinan relaciones sintácticas entre elementos explícitos formales (como opuestos para describir esencialmente relaciones semánticas o pragmáticas de coherencia y similares). Por supuesto, las categorías de la gramática textual como la cláusula textual y el párrafo se asocian con reglas de interpretación semántica, pero la gramática textual, tal como aquí es utilizado el término, es la responsable de asignar tales interpretaciones solo en la medida en que ellas se correspondan con la presencia de categorías explícitamente formales.

1.2. Perspectiva pragmática. Los aportes de Figueras (2001)

Enmarcada en la teoría relevantista, Figueras (1997, 2001) presenta un enfoque pragmático para desarrollar una gramática del sistema de los signos de puntuación. La puntuación, en este caso, se considera como un mecanismo cuya función primordial es controlar eficientemente la interpretación del lector. A tales efectos, la autora especifica de qué modo cada signo dirige el proceso de comprensión textual.

“Los signos de puntuación codifican información procedimental que dirige el proceso de recuperación del contenido explícitamente transmitido por el texto (las explicaturas en el modelo relevantista). Los signos de puntuación contribuyen a fijar la forma proposicional de cada uno de los enunciados del texto y minimizar el esfuerzo de procesamiento del lector, optimizando, de este modo, la relevancia global del texto” (Figueras, 1997: en línea).

De esta forma, Figueras (2001) propone una gramática del sistema de la puntuación, especificando el modo en que cada uno de los signos dirige el proceso de comprensión textual. Vale aclarar que dichos signos delimitan las distintas unidades lingüísticas que conforman el discurso escrito. Por ende, la puntuación debe concebirse como un mecanismo más de organización del texto, según la autora, estaría en el mismo nivel que el de los marcadores discursivos, puesto que permiten delimitar distintas unidades textuales y señala de qué modo las debe interpretar el lector.

A tales efectos, clasifica a los signos de puntuación en dos grupos, el primer régimen de signos lo constituyen el punto final, el punto y aparte, el punto y seguido, el punto y coma, los dos puntos y la coma; son los signos mediante los cuales, el autor organiza la información del texto. El segundo régimen lo conforman las rayas, los paréntesis y las comillas, por un lado, y los signos de exclamación, interrogación y puntos suspensivos, por otro. Con los primeros, es posible insertar un segundo discurso en el propio, además, las rayas y los paréntesis delimitan incisos secundarios, que pueden suprimirse sin graves repercusiones sintácticas, pero que el autor juzga necesario proporcionar. Los segundos no delimitan un segundo discurso, sino que son indicadores de modalidad.

Cada signo de primer régimen demarca una unidad de distinto rango dentro de la estructura jerárquica que vertebra el texto. Así, el punto final señala la unidad mayor, el texto, y la coma, la unidad menor, el sintagma. Los demás signos de primer orden (punto y seguido, punto y coma y dos puntos) demarcan unidades intermedias y cada nueva unidad queda integrada en un rango superior. Tomando como base el planteo de Nunberg (1990), Figueras (2001) determina las siguientes unidades textuales:

- **El texto:** delimitado por la sangría de la primera línea y la mayúscula inicial, que lo abre, y el punto final, que lo cierra.
- **El párrafo:** delimitado por la sangría y la mayúscula al principio, y el punto y aparte al final.

- **El enunciado textual:** definido mediante el punto y seguido.
- **La cláusula textual:** unidades menores al enunciado textual definidas por punto y coma.
- **El enunciado oracional:** unidad interior a la cláusula textual, anunciada por los dos puntos.
- **El sintagma:** unidad intraoracional definida por la coma.

Con respecto a las instrucciones que transmite la coma, Figueras (2001) señala que es difícil marcar una única instrucción, porque son muchas las funciones que desempeña y porque, además, es un signo directamente relacionado con la estructura sintáctica del período. Generalizando, podría decirse que la coma sirve básicamente para anunciar, delimitar y definir sintagmas que se pueden calificar de periféricos o circunstanciales. Es decir, la coma separa unidades sintagmáticas que no pertenecen al núcleo oracional básico formado por el sujeto y el predicado (este último constituido por el verbo más sus complementos fuertes). La coma, desde esta perspectiva, contribuye decisivamente a guiar la interpretación de la estructura y significado del enunciado: advierte al lector que el sintagma que delimita tiene un nivel de complementación más externo, más periférico, más ‘circunstancial’ que el de los complementos ‘fuertes’ (Figueras, 2001).

A partir de la propuesta de Figueras (2001), se han elaborado diversos trabajos que tratan a la puntuación en relación con su adquisición durante la escolaridad primaria (Dávalos & Alvarado, 2009), su uso en el nivel medio y superior (Rojas, 2011) y su enseñanza en el nivel universitario (Sánchez, 2004). Asimismo, también se pueden encontrar estudios relacionados con el uso de la puntuación en la prensa digital (García, 2013).

En el presente trabajo, se toma en consideración dos de las funciones principales que señala Figueras (2001): (i) la de delimitar incisos; las comas que cumplen esta función, se denominarán Comas Delimitadoras, y (ii) la de anunciar, aunque más precisamente indicar, para los casos de enumeraciones y elipsis, a estas se las denominarán Comas Indicadoras.

1.3. Análisis de la puntuación en el marco de la lingüística computacional

1.3.1. Estudios basados en la propuesta de Nunberg (1990)

Entre los trabajos en el ámbito de la lingüística informática realizados a partir de los aportes de Nunberg (1990), se destacan los llevados a cabo por Briscoe y Carroll (1995), quienes presentan una aproximación al análisis del dominio robusto sintáctico de ocurrencias irrestrictas naturales del inglés. La técnica empleada involucra el análisis de secuencias de partes del discurso y niveles de puntuación usando una unificación de base gramatical acoplada con un etiquetador probabilístico. Estos

autores construyen una gramática del texto, como la que intenta Nunberg (1990), a partir de la tokenización de los signos de puntuación, separadamente de las palabras, y usan una gramática de unificación en conjunción con un parser para ciertos usos lexicográficos. Aquí, la puntuación es vista como un uso no solo para segmentar el texto en unidades adecuadas para el análisis, sino también para resolver estructuras ambiguas.

Otro de los trabajos destacados es el de White (1995), que examina a la puntuación desde un punto de vista de la Generación de Lenguaje Natural (GLN). El autor investiga cómo las aproximaciones de Nunberg (1990) estarían reflejadas si se incorporan en un sistema de GLN. Asimismo, extiende y critica los análisis de la puntuación de Nunberg (1990) y propone una arquitectura abocada para la implementación con tres componentes: sintáctico, morfológico y gráfico. Estos se ocupan de las reglas de la presentación de la puntuación para la jerarquía, la adyacencia, y la forma gráfica, respectivamente. De este modo, apunta a poner reglas en curso de generación de la puntuación en acción, superando así algunos de los defectos del marco de Nunberg (1990).

Un tercer trabajo fundamental es el de Jones (1996), quien describe una tarea computacional de análisis relacional también basado en el marco teórico de Nunberg (1990). No obstante, el autor se abstiene de usar una gramática de dos niveles, considerando que las interacciones entre ambos complejizan innecesariamente el sistema. Por el contrario, establece la necesidad de una nueva teoría de la puntuación, que sea adecuada para la implementación computacional y examina las funciones sintácticas de los signos en el texto. Su gramática revisada produce similares (o en todo caso, mejores) resultados que los obtenidos por Briscoe y Carroll (1995). Jones (1996) también da una teoría esquemática de la puntuación en la que él clasifica sintáctica, semántica y pragmáticamente los usos de los signos.

1.3.2. Trabajos computacionales enfocados en la coma

Bayraktar, Say y Akman (1998), retomando la propuesta de Jones (1996), presentan uno de los primeros análisis enfocados exclusivamente en la coma. En él, se crea una detallada lista de patrones sintácticos que se corresponden a los diferentes roles que la coma puede tener. En la medida en que la función principal de la coma es fijar límites sintácticos, la información contenida en un análisis arbóreo sería suficiente para hacer la clasificación de sus usos. A tales efectos, los autores, en primer lugar, construyen una base de datos de todos los patrones sintácticos que contienen una o más comas y, en segundo lugar, realizan una clasificación en la que asignan los patrones sintácticos establecidos a las clases apropiadas.

Con respecto a trabajos para el español, se puede mencionar la propuesta de Garat (2006), quien construye, tomando como base los planteos de Bayraktar et al. (1998), un analizador sintáctico de superficie que combina métodos estadísticos y simbólicos.

De este modo, obtiene, por un lado, un evaluador de las funciones de la coma mediante técnicas de aprendizaje automático y, por otro, escribe reglas de análisis que aprovechan la clasificación realizada por el evaluador, utilizando formalismos de reglas de reescritura.

Los trabajos actuales sobre este signo de puntuación se focalizan, como se mencionó en la introducción, en la asignación automática de comas en textos orales transcritos, aquellos elaborados a partir de información lingüística se basan en el etiquetado de las palabras; su longitud y la longitud de las cláusulas, y la posición y distribución de los sintagmas en el texto (Chian et al., 2012). Asimismo, también es posible encontrar propuestas basadas en elementos prosódicos y de entonación (Shriberg, Stolcke, Hakkani-Tür & Tür, 2000), mixtos (Favre, Grishman, Hillard, Ji, Hakkani-Tür & Ostendorf, 2008) y en modelos predictivos (Lu & Tu, 2010; Chian et al., 2012).

2. Propuesta de clasificación de las funciones de la coma

Para el tratamiento informático de la coma, fue necesario, desde un primer momento, determinar fehacientemente cuáles eran sus funciones. A tales efectos, se partió de los análisis sobre este signo en las gramáticas tradicionales y se trabajó a partir de las clasificaciones propuestas por Alcoba (2000), Simone (1991) y Figueras (2001). Asimismo, también se consideró las de Cipriano (1976), del diccionario Panhispánico de Dudas (RAE, 2005) y de la Ortografía de la lengua española (RAE, 2010).

A partir de dichos estudios, se propuso una clasificación propia, en la que estuvieran contemplados los usos gramaticales de este signo, pero que, a la vez, fuera pertinente con el trabajo de implantación computacional. Se necesitó establecer una clasificación distinta a la de los demás autores, en primer lugar, porque aquellas eran demasiado complejas para el trabajo computacional y, en segundo lugar, porque se observaron algunas cuestiones problemáticas en fenómenos que no estaban del todo claros como, por ejemplo, los incisos. Al respecto, los autores consultados (Alcoba, 2000; Figueras, 2001; RAE, 2005, entre otros) coinciden, en mayor o menor medida, en que el inciso es una expresión que interrumpe la cláusula para explicar algo relacionado con ella y que posee independencia sintáctica. Para ofrecer un mejor análisis de este planteo, es pertinente desmembrar dicha definición en tres partes fundamentales. Entonces, una construcción incidental: (i) interrumpe la cláusula; (ii) tiene independencia sintáctica, y (iii) explica algo sobre la cláusula que la contiene. A continuación se analizará cada ítem, con el objetivo de demostrar la incompatibilidad entre ellos.

Sobre el primer aspecto, autores como Alcoba (2000: 155) relacionan al inciso con “cualquier expresión que interrumpe la oración”. Al respecto, Desinano (2004) señala que es posible que sea muy difícil explicar con cierto rigor a qué se refiere este autor con esta idea de interrupción. Tanto en los ejemplos del propio Alcoba (2000), como así también en los de Desinano (2004), los incisos constituyen una parte del *continuum*

oracional, en donde, más allá de la segmentación, siguen constituyendo una unidad. Ejemplos serían los casos de aposiciones explicativas (3) o subordinadas relativas explicativas (4).

(3) [Juan, el marido de María, solicitó el divorcio.]

(4) [Juan, que ya sabía todo, pidió el divorcio.]

En relación con el segundo ítem, otra de las características atribuidas al inciso es que estos parecerían no tener relación sintáctica de ningún tipo con la cláusula en la que están incluidos. Como quedó demostrado en el párrafo anterior, esto no siempre se aplica, pues incisos como la aposición y la subordinada relativa explicativa, claro está, dependen del sintagma al que modifican. En el ejemplo siguiente, tomado de Alcoba (2000), se pueden apreciar otros tipos de dependencia sintáctica.

(5) [Los caballeros, vestidos con sus armas, con el paso firme, se fueron al combate.]

Esta cláusula contiene dos incisos. Por un lado, ‘vestidos con sus armas’, que modifica al sujeto y es una construcción participial, cuyo comportamiento es similar al de una cláusula subordinada relativa explicativa. Por otro, ‘con el paso firme’, que se trata de un complemento circunstancial del verbo desplazado hacia la izquierda. Como puede observarse, ambos incisos poseen una función sintáctica que depende de la cláusula que los contiene.

Además, el caso de ‘con el paso firme’ remite a la cuestión de la alteración del orden regular o canónico. Se trata de un recurso que a menudo se utiliza con el propósito de destacar un sintagma dentro de la cláusula y está estrechamente relacionado con el fenómeno de la topicalización (Morales, 2011). Aquí es preciso señalar que, en la bibliografía consultada, autores como Cipriano (1976) o el Panhispánico de Dudas (RAE, 2005), entre otros, al momento de proponer una clasificación de las funciones de la coma, presentan a la de delimitar incisos y a la de indicar alteraciones en el orden canónico, como dos fenómenos claramente reconocibles y diferenciables. Tanto Alcoba (2000) como el diccionario Panhispánico de Dudas (RAE, 2005) mencionan que la alteración del orden canónico o regular se da cuando un sintagma no aparece en su orden habitual, entendiendo por esto, al orden dado por el esquema sujeto + verbo + objeto directo + complementos verbales. He aquí otro ejemplo tomado de Alcoba (2000):

(6) [Bombardearon, los aviones de la OTAN, una aldea de montañas.]

Alcoba (2000) observa en (6) que el elemento dislocado está delimitado por comas y estas deben ser ‘interpretadas con la melodía típica de los incisos’ (Alcoba, 2000). Sin embargo, De la Fuente (2005) no se refiere a ‘alteración del orden’, sino que considera construcciones como las ejemplificadas dentro del fenómeno de los incisos en general. De acuerdo con los planteos de este autor, se estaría ante ‘incisos reubicables’ (De la

Fuente, 2005), puesto que pueden ser reordenados a fin de mantener el orden regular y, en algunas ocasiones, se podría prescindir de la puntuación. En el caso de ‘con el paso firme’, su reubicación daría la siguiente cláusula:

(7) [Los caballeros se fueron al combate con el paso firme.]

Lo que se puede observar es que se está ante la confluencia de dos cuestiones dada por un elemento que altera el orden canónico y se ubica en una posición incidental. El hecho de considerar a ‘con el paso firme’ únicamente como una alteración o como un inciso implica caer en un análisis parcial, en donde se dejaría de percibir uno u otro fenómeno. Este hecho se puede comparar con la metáfora del significado y el significante, que Saussure (2007) los presenta como dos caras de una misma hoja de papel, que no se pueden separar y que, a la vez, una vale por la existencia de la otra. Pues, en este caso, se puede ver que hubo una alteración, porque hay un inciso, y, a la inversa, hay un inciso porque se ha dado un caso de alteración.

A partir de los ejemplos presentados, y en relación con la definición clásica de inciso expuesta más arriba, por lo menos para los que se hayan delimitados por comas, se puede apreciar que muchas construcciones que pueden considerarse incidentales poseen función sintáctica, no interrumpen la oración y, en algunos casos, como ser la aposición, ni siquiera alteran el orden regular.

Además de lo que se ha planteado en esta sección, también, por el contrario, es posible encontrar incisos que no cumplen función sintáctica y no aportan información a la cláusula que los contiene, como ser el vocativo (8) y la construcción interjectiva (9).

(8) [Hoy, Juan, tengo ganas de ir al cine.]

(9) [La próxima vez, ¡quiera dios!, me llame María.]

Con todo lo dicho, se desprende que las tres características primordiales que se enumeraron más arriba sobre el inciso son incompatibles cuando este se haya delimitado por comas. Pues, no siempre interrumpen la cláusula; cuando aportan información a esta, no pueden considerarse como independientes, sintácticamente hablando, y en los casos en los que hay independencia sintáctica, se tratan de elementos marginales que no aportan información (Koza, 2013).

Tanto por esta problemática referente al inciso en particular, como así también inconvenientes de otras índoles, como, por ejemplo, la enumeración, sumado a la complejidad que presentaban las clasificaciones analizadas, se optó por establecer una clasificación propia, compatible con la descripción gramatical y el trabajo computacional.

El primer paso en la elaboración de la nueva clasificación fue englobar aquellas

funciones con características similares en una sola, por ejemplo, la seriación de cláusulas se consideró dentro de la enumeración. Luego, se establecieron tres funciones principales: función indicadora, función delimitadora y función desambiguadora. Estas concentran las múltiples funciones descritas por los autores consultados. Así, la coma indicadora señala enumeraciones (10) y elipsis (11); la coma delimitadora enmarca incisos (12), y la coma desambiguadora evita posibles confusiones en construcción que pueden presentar más de una interpretación, como, por ejemplo, el caso de los marcadores discursivos (13) y (14) (Prada, 2001).

(10) [Escribió una obra de teatro, dos libros de poesía y una novela romántica.]

(11) [El niño pidió una hamburguesa y una Coca-Cola; el hombre, solo un café.]

(12) [El trayecto era tranquilo, hasta que, ¡oh, Jesús!, aparecieron las víboras.]

(13) [Sin embargo, yo creo que es un buen tipo.]

(14) [La operación se realizó sin embargo de bienes.]

En segundo lugar, se procedió a analizar las distintas construcciones que están afectadas por la coma en las diferentes funciones gramaticales que este signo posee. A tales efectos, hubo que realizar, además, clasificaciones para los tipos de enumeraciones e incisos, a fin de realizar, luego, una adecuada modelización para la implantación computacional. De este modo, las enumeraciones fueron clasificadas de acuerdo con la naturaleza y función sintáctica de los elementos enumerados, como así también, según la naturaleza la condición de abierta o cerrada de la enumeración. A continuación, se presenta dicha clasificación.

De acuerdo con la naturaleza de los componentes

Sintagmáticas: Nominales, verbales, adjetivales, preposicionales y adverbiales.

(15) [María es alta, rubia y simpática.]

Clausales: subordinadas y no subordinadas.

(16) [Quiero que me quieran, que no me juzguen y que me esperen.]

(17) [Juan canta, María baila y Pedro zapatea.]

Mixtas:

(18) [Quiero una mujer linda, de cabello lacio y que sepa cocinar.]

De acuerdo con el número de componentes

Cerradas:

(19) [Compró pan, verduras y carne.]

Abiertas:

(20) [Platero es pequeño, peludo, suave, etcétera.]

(21) [Platero es pequeño, peludo, suave...]

Los incisos, por su parte, se clasificaron de acuerdo con si tenían o no función sintáctica y, para el caso de que los tuvieran, si requerían o no de un antecedente.

Con función sintáctica:

Con antecedente: aposiciones, subordinadas relativas explicativas, etcétera

(22) [Juan, el marido de María, salió a pasear.]

(23) [Juan, que ya sabía todo, pidió el divorcio.]

Sin antecedente: Alteraciones en el orden canónico y construcciones dicendi

Los primeros remiten a elementos de la oración, generalmente complementos verbales, desplazados de su lugar habitual. Las construcciones dicendi son aquellas utilizadas por el autor para insertar en el texto palabras provenientes de un discurso ajeno (Koza, 2009).

(24) [María, mientras Juan trabajaba, tenía una aventura.]

(25) [“Juan”, dijo María, “vayamos al cine”.]

Sin función sintáctica

Construcciones interjectivas

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

Vocativos

(27) [Una vez que se descarta lo imposible, querido Watson, lo que queda es la verdad.]

Los casos de elipsis serán tratados en trabajos futuros y, para los marcadores discursivos, se tomó la clasificación que presentan Zorraquino y Portolés (1999).

Posteriormente, a partir de las clasificaciones de las funciones de la coma y de las expresiones afectadas por ella, se procedió a la modelización que permitió la implantación computacional.

3. Metodología

Se desarrolló un método computacional para el reconocimiento automático de las funciones gramaticales de la coma, de modo de comprobar la clasificación propuesta. A tales efectos, fue necesaria la creación de reglas de análisis que permitieran señalar la manera en que se relacionaban ciertas unidades (palabras, sintagmas, cláusulas, otros signos de puntuación) con la coma dentro de un texto. En esta ocasión, se tomó como base la noción del sistema de los signos de puntuación propuesta por Nunberg (1990) en cuanto a que los signos delimitan unidades textuales y, a tales efectos y tratándose de la coma, se focalizó en el nivel del sintagma (Figueras, 2001).

Para el trabajo informático, se recurrió a los programas *Smorph* (Aït-Mokthar, 1998) y Módulo *Post Smorph* (MPS) (Abacci, 1999). *Smorph* permite analizar morfológicamente la cadena de caracteres, dando como salida la asignación categorial y morfológica correspondiente a cada ocurrencia de acuerdo con los rasgos declarados. MPS tiene como *input* el *output* de *Smorph* y, a partir de reglas de recomposición, descomposición y correspondencia declaradas por el usuario, analiza la cadena de lemas resultante del análisis morfológico. Se trabajó con un corpus compuesto por textos periodísticos que sumaban un total de 10.000 palabras.

La modelización y la posterior implantación computacional se focalizaron en: (i) enumeraciones sintagmáticas nominales, verbales, adjetivales y preposicionales completas con conjunción y abiertas; (ii) incisos: interjecciones, aposiciones, subordinadas relativas explicativas, construcciones dicendi y otros incisos ubicados al inicio de la cláusula o incrustados en ella (aquí se incluyen incisos que modifican a toda la cláusula, alteraciones en el orden regular, etcétera), y (iii) marcadores discursivos.

El trabajo informático se llevó a cabo de la siguiente manera:

0. Etapa previa: Análisis morfológico y reconocimiento de los signos de puntuación con *Smorph*, detección de sintagmas nominales y adjetivales, y sintagmas verbales núcleos, que empiezan en el primer elemento del sintagma (por ejemplo, un pronombre) y finalizan en el núcleo, es decir, el verbo conjugado (Bès & Solana, 2004) a partir de reglas de reagrupamiento con MPS;

1. Detección de Enumeraciones;
2. Detección de interjecciones, aposiciones, subordinadas relativas explicativas y construcciones dicendi;
3. Detección de sintagmas preposicionales y adverbiales, posterior detección de expresiones delimitadas por comas no detectadas en 2;
4. Desambiguación y reconocimiento de marcadores discursivos.

A continuación, se describe el trabajo realizado.

3.1. Análisis morfológico y reconocimiento de los signos de puntuación

Con el programa *Smorph*, se obtuvo el análisis morfológico de cada término y el reconocimiento de los signos de puntuación. Aquí, la coma debe declararse en el archivo *entradas.txt* que es el diccionario fuente que utiliza el programa para proceder al análisis morfológico. Por ejemplo, dado el siguiente fragmento:

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

Smorph presenta el siguiente resultado:

‘En’.

[‘en’, ‘EMS’,’prep’].

‘un’.

[‘un’, ‘EMS’,’det’].

‘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’,’cl’].

‘,’.

[‘cc’, ‘EMS’,’coma’].

‘Carlos’.

[‘Carlos’, ‘EMS’,’npr’].

‘Juárez’.

[‘Juárez’, ‘EMS’,’npr’].

‘se’.

[‘lo’, ‘EMS’,’cl’].

‘negó’.

[‘negar’, ‘EMS’,’v’, ‘MODOV’,’ind’, ‘PERS’,’3a’, ‘NUM’,’sg’, ‘TPO’,’prets’, ‘TR’,’ir’, ‘TC’,’cl’].

‘a’.

[‘a’, ‘EMS’, ‘prep’].

‘declarar’.

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

‘

‘.

[‘linsig’, ‘EMS’, ‘pun’].

Una vez realizado el análisis de *Smorph*, se procedió a la elaboración de reglas para el reconocimiento de sintagmas y, posteriormente, de las funciones gramaticales de la coma.

3.2. Modelización y creación de reglas para MPS

El *output* generado por *Smorph* es el *input* con el que trabaja MPS. Una vez obtenidos los resultados del primero, se constituyeron las reglas para formar los sintagmas. A partir de ellos, fue posible establecer, luego, las reglas para el reconocimiento de las funciones de la coma, dado que el signo afecta, precisamente, a estas unidades.

Se ejemplificará con la enumeración, puesto que, en los textos, es la función más clara de observar y, dentro del análisis computacional, la más sencilla de modelizar y reconocer automáticamente. Si se toma el ejemplo (10)

(10) [Escribió una obra de teatro, dos libros de poesía y una novela romántica.],

una modelización pertinente para la enumeración sería:

SN + coma + SN + cop. + SN = Enumeración Nominal Completa

Esto quiere decir que, si en el texto se encuentra un SN seguido de coma, más otro SN seguido de una conjunción copulativa (cop), más otro SN; entonces hay una enumeración nominal completa.

De esta manera, se desarrolló un conjunto de reglas que modelan los casos posibles de enumeraciones sintagmáticas, las cuales eran las siguientes:

$(SX + coma) \geq 1 + SN + conjunción + SN = ESX$

$(SX + coma) \geq 2 + punto\ suspensivos = ESX$

$(SX + coma) \geq 2 + 'etcétera' = ESX$

$(SX + coma) \geq 2 + 'entre\ otras\ cosas' = ESX$

$(SX + coma) \geq 3 = ESX$ (para los casos de asíndeton)

He aquí algunas de las enumeraciones detectadas en el corpus:

(...) El menú permitido incluye nada más que **‘carnes desgrasadas , verduras hervidas y gelatinas’**. [**‘carne desgrasado cc verdura hervido y gelatina’**, **‘EMS’**, **‘Enum-Nom-Comp’**]

(...) útiles para las dirigencias que aspiran **‘a dominar la calle , a organizar pobladas o a presionar a las autoridades’**. [**‘a dominar el calle cc a organizar poblada o a presionar a el autoridad’**, **‘EMS’**, **‘Enum-Prep-Comp’**]. mediante la protesta violenta o la alteración del orden público. (...)

(...) porque ha conseguido **‘crecer , expandir su comercio exterior y lograr un inédito superávit externo’**.

[**‘crecer cc expandir su comercio exterior y lograr un inédito superávit externo’**, **‘EMS’**, **‘Enum-Verb-Comp’**]. sin ceder a las presiones. (...)

De manera similar, se trabajó con las demás funciones, es decir, el reconocimiento de incisos y las alteraciones en el orden regular, para la función delimitadora y el de marcadores discursivos, para la desambiguadora. Los resultados obtenidos se evaluaron mediante las medidas de precisión, cobertura y medida F. En el parágrafo siguiente, se detalla dicha evaluación.

4. Resultados obtenidos

Se evaluaron los resultados de la implantación computacional realizada a partir de las medidas de precisión, cobertura y medida F. A continuación se presenta una tabla (Tabla 1) con los porcentajes obtenidos.

Tabla 1. Resultados obtenidos.

Elementos a reconocer	Total de expresiones	Elementos detectados	No detectados	Detecciones erróneas	Precisión	Cobertura	Medida F
Enumeración	33	29	4	4	87%	93%	89,90%
Aposición	29	27	2	3	90%	93,1%	91,52%
Sub. Rel. Exp.	32	31	1	6	81,25%	97%	88,43%
Const. Dic.	9	8	1	0	100%	89%	94,17%
Del. al inicio	72	62	10	2	97,2%	86%	91,25%
Del. incrust.	50	46	4		96%	92%	93,96%
Marc. Disc.	34	34	0	0	100%	100%	100%

Como se puede apreciar, las mejores precisiones fueron para el reconocimiento de las construcciones dicendi y los marcadores discursivos, 100% en ambos casos, y la mejor cobertura, para los marcadores discursivos, 100%, seguida del reconocimiento de las subordinadas relativas explicativas, 97%. Por otro lado, la precisión más baja fue para las subordinadas relativas explicativas, 81,25% y la cobertura más baja, para delimitaciones ubicadas al inicio, 86%. En el caso de la medida F, la mejor remitió a los marcadores discursivos, obteniéndose en ellos 100%, seguida de las construcciones dicendi con 94,17%. Asimismo, la más baja fue la correspondiente a las subordinadas relativas explicativas, 84,43%. A tales efectos, los resultados obtenidos permiten argumentar que la implementación automatizada es válida y, con ello, se justifica la clasificación propuesta.

Ahora bien, también debe señalarse que se vislumbró un porcentaje bajo de errores y omisiones que se debieron a tres factores:

- Palabras desconocidas, no incluidas en el diccionario fuente: la presencia de neologismos, términos específicos y extranjerismos no pudieron ser analizados por *Smorph*, lo que afectó a la reconocimiento de algunas funciones de la coma.
- Estructuras complejas que no pudieron ser contempladas al momento de elaborar las reglas: algunos casos presentaban estructuras complejas, no tenidas en cuenta al momento de la elaboración de las reglas.
- Fallos en la puntuación por parte del autor del texto: Los errores en la puntuación ocasionaron etiquetados erróneos en las funciones de la coma.

CONCLUSIONES

Se presentó el análisis del uso de la coma en relación con la sintaxis y, posteriormente, se estableció una clasificación de las funciones gramaticales de este signo de puntuación basada en la relación que establece la coma con los otros elementos del texto en el nivel sintáctico. A partir de esta clasificación, se presentó un método de detección automática de estas construcciones, a partir de la formalización y creación de reglas. Los porcentajes de precisión, cobertura y medida F obtenidos en la evaluación son lo suficientemente adecuados como para validar la efectividad de las reglas elaboradas. Esto último, se puede apreciar especialmente en la medida F más baja, que superó el 88%.

Durante el análisis, se pudo apreciar que, si bien es posible establecer un uso normativo de la puntuación, en los textos reales se observan desviaciones a dicha normativa sin que eso implique una escritura errónea. Así por ejemplo, es posible observar en el corpus fragmentos de este estilo:

(...) Desde hace un mes el presidente Néstor Kirchner vive de roce en roce con la dirigencia de su Partido Justicialista (...)

Aquí hay una alteración del orden regular que debiera estar delimitada por coma. No obstante, es necesario advertir, en estos casos, se puede prescindir de la puntuación, si la expresión es breve (RAE, 2005). Quedaría, entonces, al criterio del autor, cuándo una expresión es breve o extensa. Algo similar ocurriría con las rematizaciones, en donde el uso o no de coma depende, en gran medida, del autor (Figueras, 2001).

(28) Una pavada dijo Juan la semana pasada.

(29) A María la vi la semana pasada.

No obstante, y más allá de estas particularidades, el presente trabajo constituiría un aporte tanto a los estudios gramaticales, con respecto al rol de la coma en la sintaxis, como así también a la lingüística computacional. Se pretende que la presente investigación sea el inicio para investigaciones futuras en el área y, para ello, se cree que, con la clasificación propuesta, se contarán con nuevos elementos de análisis. Por otro lado, siguiendo la perspectiva de la lingüística computacional, se logró establecer un método que permitiera detectar las funciones gramaticales de la coma en textos de lenguaje natural, pero que, además, con ello validara la descripción y el análisis previo, con lo cual, puede señalarse que el presente trabajo constituye una contribución de la formalización de estructuras sintácticas.

En relación con trabajos futuros, se plantea analizar textos no periodísticos, lo que podría suscitar modificaciones a las reglas y a las categorías aquí propuestas. Por otro lado, se ha podido apreciar que la puntuación puede ser útil para tareas de extracción automática de información en textos de lenguaje natural, por lo que se intentará probar el método de detección propuesto en dicho ámbito.

REFERENCIAS BIBLIOGRÁFICAS

- Abacci, F. (1999). *Développement du Module Post Smorph*. Clermont-Fd.: Memoria del DEA de Linguistique et Informatique. Universidad Blaise-Pascal/GRIL.
- Aït-Mokthar, S. (1998). *Smorph: Guide d'utilisation. Rapport technique*. Clermont-Fd.: Universidad Blaise Pascal/GRIL.
- Alcoba, S. (2000). Puntuación y melodía de la frase. En S. Alcoba (Coord.), *La expresión oral* (pp. 147-186). Madrid: Ariel Practicum.
- Bayraktar, M., Say, B. & Akman, V. (1998). An analysis of English punctuation: The special case of comma. *International Journal of Corpus Linguistics*, 3(1), 33-57.
- Bès, G. & Solana, Z. (2004). *Análisis morfológico y gramáticas locales: Introducción y una aplicación concreta*. Ponencia presentada en las I Jornadas Argentinas de Lingüística Informática: Modelización e Ingeniería, Facultad de Humanidades y Artes, Universidad Nacional de Rosario, Rosario.
- Briscoe, T. & Carroll J. (1995). Developing and evaluating a probabilistic LR Parser of Part-Of-Speech and punctuation labels. En *Proceedings of the ACL/SIGPARSE 4th International Workshop on Parsing Technologies* (pp. 48-58). Prague: ACL.
- Chian, C., Wan, Y. & Chen, S. (2012). Punctuation generation inspired linguistic features for mandarin prosodic boundary prediction. *ICASSP 2012*, 4597-4600.
- Cipriano, N. (1976). *El uso de la coma. Teoría y práctica*. Buenos Aires: Ediciones Depalma.
- Dávalos, D. & Alvarado, E. (2009). La puntuación y otros recursos en la organización de textos infantiles propios y ajenos. *Lectura y vida*, 30(1), 6-16.
- De la Fuente, M. (2005). *La coordinación en inciso y su puntuación*. *Revista de estudios literarios*. [en línea]. Disponible en: <http://www.ucm.es/info/especulo/numero29/dobles.html>
- Desinano, N. (2004). Puntuación y gramática. En N. Múgica & N. Desinano (Comps), *Estudios del lenguaje y enseñanza de la lengua* (44-88). Rosario: Ediciones Juglaría.
- Ehrlich, E. (1992). *Theory and problems of punctuation, capitalization, and spelling*. Hong Kong: McGraw-Hill.
- Favre, B., Grishman, R., Hillard, D., Hi, H., Hakkani-Tür, D. & Ostendorf, M. (2008). *Punctuating speech for information extraction*. *Spoken Language Technologies*. [en línea]. Disponible en: <http://www.cs.nyu.edu/hengji/ssie.pdf>
- Figueras, C. (1997). *La semántica procedimental de la puntuación*. *Especulo. Revista de estudios literarios* [en línea]. Disponible en: <http://www.ucm.es/info/especulo/numero12/puntuac.html>

- Figueras, C. (2001). *Pragmática de la puntuación*. Madrid: Octaedro.
- Figueras, C. (2013). *Pragmática de la puntuación y nuevas tecnologías*. Universitat de Barcelona (en prensa).
- Garat, D. (2006). *Análisis de superficie basado en puntuación*. Tesis de magíster, Universidad de la República, Montevideo, Uruguay.
- García, E. (2013). *La puntuación en la prensa digital*. Tesis de magíster, Universitat de Barcelona, Barcelona, España.
- Hall, N. & Robinson, A. (Coord.) (1996). *Learning about punctuation*. Londres: Clevedon.
- Jones, B. (1996). Towards texting the syntax of punctuation. En *Proceedings of the 16th International Conference on Computational Linguistics (COLING-96)* (pp. 363-365). Copenhagen: COLING.
- Koza, W. (2009). Análisis automático de textos: Reconocimiento de construcciones dicendi. *Infosur revista*, 3, 95-104.
- Koza, W. (2013). El inciso delimitado por comas. Análisis del fenómeno y propuesta de detección automática. *Logos*, 23(2), 169-195.
- Luque, J. (2006). *Puntos y comas. La grafía de la articulación del habla*. Granada: EUG.
- Morales, N. (2011). La topicalización, información focalizada o tematizada. En *Memorias del XII Encuentro Nacional del Estudio en Lenguas* (pp. 460-469). Tlaxcala: Universidad Autónoma de Tlaxcala.
- Nunberg, G. (1990). *The linguistics of punctuation*. Stanford: CSLI Lecture Notes. Stanford University Press.
- Prada, J. (2001). *Marcadores del discurso en español Análisis y representación*. Tesis doctoral, Universidad de la República, Facultad de Ingeniería. Montevideo, Uruguay.
- RAE (2005). *Diccionario panhispánico de dudas*. Bogotá: Real Academia Española.
- RAE (2010). *Ortografía de la lengua española*. Bogotá: Real Academia Española.
- Rojas, H. (2011). *El uso de la puntuación en estudiantes chilenos. Una aproximación descriptiva desde el análisis del discurso*. Tesis de magíster, Universidad de Chile, Santiago, Chile.
- Roselló, J. (2010). *Análisis de los signos de puntuación en textos de estudiantes de educación secundaria*. Tesis doctoral, Universitat de València, Valencia, España.
- Sánchez, A. (2004). La puntuación y las unidades textuales: Una perspectiva discursiva para el estudio de los problemas de su uso y para su enseñanza. *Revista Educación*, 28(2), 233-254.

- Saussure, F. (2007). *Curso de lingüística general*. Madrid: Losada.
- Shriberg, E., Stolcke, A., Hakkani-Tür, D. & Tür, G. (2000). Prosody-based automatic segmentation of speech into sentences and topics. *Speech Communication*, 32(1-2), 127-154.
- Simone, R. (1991). Riflessioni sulla virgola. En M. Orsolini & C. Pontecorvo (Eds.), *La costruzione del testo nei bambini* (pp. 219-231). Firenze: La Nuova Italia.
- Wallace, M. (2012). *Uso y función de marcas de puntuación en situación de producción y revisión de epígrafes*. Tesis de magíster, Universidad Nacional de La Plata, La Plata, Argentina.
- White, M. (1995). Presenting punctuation. En *Proceedings of the Fifth European Workshop on Natural Language Generation* (pp. 107-125). Leiden: The Netherlands.
- Zamudio, C. (2004). *El papel de la escritura alfabética en la construcción del dato oral*. Tesis doctoral, Colegio de México, Centro de Estudios Lingüísticos y Literarios, México DF, México.
- Zorraquino, M. & Portolés, J. (1999). Los marcadores del discurso. En I. Bosque & V. Demonte (Dirs.), *Gramática descriptiva de la lengua española* (pp. 4551-4213). Madrid: Espasa Calpe.