

El orden de palabras en español. Descripción del fenómeno y propuesta de detección automática de alteraciones en el orden regular

Automatic analysis of ambiguities in Spanish: the 'noun' and 'verb' categories

Walter Koza

Grupo INFOSUR-UNR-Becario de CONICET

Rosario, Argentina

walter_koza@yahoo.com.ar

Abstract

The present article discusses word order in Spanish and presents an automatic detection method of alterations in regular order, specifically for those elements shifted to the beginning of the clause.

The distribution of words and syntagms does not occur crosslinguistically the same way. By no means does this imply an arbitrary but a definable pattern of distribution. Out of the works consulted, the contributions made in this field by Greenberg [1], Koktova [2] and Holan et al [3] are presented here.

Next the relationship between word order and distribution of information is observed, with the aim of displaying some specificities of word order in Spanish later.

Finally, an automatic recognition method of regular word order alterations through the use of Smorph and Module Post Smorph (MPS) software packages is presented. Through the formalization of certain linguistic structures, detection rules were created and tested in a corpus composed of journalistic texts. The results obtained show 97.2% precision and 86% coverage.

Keywords: word order – Topic – Focus – Alteration of the regular order – comma

Resumen

En el presente artículo, se indaga la cuestión del orden de palabras en español y se presenta un método de detección automática de alteraciones en el orden regular, específicamente para aquellos elementos desplazados al inicio de la cláusula.

La distribución de palabras y sintagmas no se hace del mismo modo en las lenguas del mundo. Esto no implica que se trate de una distribución arbitraria, sino que, por el contrario, la cuestión del orden responde a patrones que pueden ser definidos. De los trabajos consultados, aquí se presentan los aportes realizados en este terreno por Greenberg [1], Koktova [2] y Holan et al [3].

Posteriormente se observa la relación entre el orden de palabras y la distribución de la información, para luego exponer algunas particularidades sobre el orden de palabras en el español.

Por último, se presenta un método de reconocimiento automático con los softwares Smorph y Módulo Post Smorph (MPS) de alteraciones del orden regular. Mediante la formalización de ciertas estructuras lingüísticas, se crearon reglas para la detección y se pusieron a prueba en un corpus

conformado por textos periodísticos. Los resultados obtenidos arrojan un 97,2% de precisión y un 86% de cobertura.

Palabras Claves: Orden de Palabras – Tópico – Foco – Alteración del orden regular – coma.

1. INTRODUCCIÓN

Se pretende abordar la cuestión del orden de palabras en español y presentar un método de detección de alteraciones en el orden regular por medio de herramientas informáticas.

Una particularidad del orden de palabras es que su alteración se indica a través de la coma. A tales efectos, aquí se retoman los lineamientos generales de la tesis doctoral que estoy realizando sobre el análisis de las funciones de este signo de puntuación dentro del marco de la lingüística computacional, dirigido por la doctora Zulema Solana y financiado por una beca de CONICET.

En lo que atañe al trabajo informático, se recurrió a los programas Smorph [4] y al Módulo Post Smorph (MPS) [5]. 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, dando como resultado las formas correspondientes a un lema con los valores pertinentes. Este programa es una herramienta declarativa, y la información utilizada está separada de la maquinaria algorítmica, lo que 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. MPS (Módulo Post-Smorph) ha sido especificado en el GRIL por Caroline Hagège, José Rodrigo, Gabriel Bès y Faizza Abacci, e implantado en C++ en un contexto de Windows por Faizza Abacci [5]. Posteriormente, fue extendido en Pasmó, 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 las entradas de la sintaxis estándar, como ser fechas, cantidades, cuestiones relativas a la sufijación y prefijación, el tratamiento de los clíticos y de las contracciones. 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.

El artículo se organiza de la siguiente manera:

En primer lugar, se presentan los estudios sobre el orden de palabras en general, llevados a cabo por Greenberg [1], Koktova [2] y Holan [3]. En segundo lugar, se observa la relación que se da entre el orden de palabras y el modo en que se presenta la información. Posteriormente, se exponen algunas consideraciones específicas referentes al estudio del orden de palabras en español, tomando como base el estudio de Fernández Soriano [6]. Por último, se presenta el análisis de este fenómeno y la implantación en máquina realizada para la detección del orden regular con los softwares mencionados.

2. TRES ENFOQUES TEÓRICOS SOBRE EL ORDEN DE PALABRAS. LOS APORTES DE GREENBERG, KOKTOVA Y HOLAN

Se ha observado que la distribución de palabras y sintagmas no se hace del mismo modo en las lenguas del mundo. Esto no implica que se trate de una distribución arbitraria, sino que, por el contrario, la cuestión del orden responde a patrones que pueden ser definidos [6]. De los trabajos

consultados, aquí se presentan los aportes realizados en este terreno por Greenberg [1], Koktova [2] y Holan et al [3].

2.1. Greenberg

Greenberg [1] es el primero en establecer una tipología a partir del orden básico en el que se disponen los tres elementos más significativos de la oración: sujeto (S), verbo (V) y objeto directo (O), denominados “universales del lenguaje”. Las lenguas se clasifican de acuerdo con las permutaciones de los tres constituyentes principales de la cláusula. Los diferentes grupos de lenguas se constituyen mediante el orden básico para la oración simple aseverativa. Estos serían los siguientes:

- SVO
- SOV
- VSO
- VOS
- OVS
- OSV

Los dos primeros órdenes son los más frecuentes y el quinto y el último casi inexistentes.

2.2. Koktova

De una sólida formación praguense y generativista [7], el aporte de Koktova [2] radica en la presentación de una nueva teoría gramatical basada en el orden de las palabras (Word-Order Based Grammar). En ella, el orden de palabras es el principio gramatical primario.

Koktova sostiene que la línea temporal de la cláusula es la principal conductora del significado oracional. No obstante, y simultáneamente, esa linealidad no deja de ser una desventaja para la comunicación verbal a causa de su unidimensionalidad, es decir, la información que se transporta (de manera verbal) fluye fuera del tiempo irrecuperablemente. A tales efectos, las lenguas naturales poseen tres medios principales para modificar la línea temporal:

- La segmentación o partición múltiple de la estructura comunicativo-informativa de la cláusula (la segmentación de la oración en el nivel profundo);
- La alternancia de elementos más o menos importantes comunicativamente (la pulsación de la oración);
- La segmentación en el nivel de superficie de la cláusula por medio de varios segmentadores en segmentos comunicativamente relevantes.

Vale aclarar que los dos últimos, en realidad, se solapan.

Además de los medios de alteración de la línea temporal, Koktova propone dos niveles de órdenes de palabras, el Orden de Palabras Profundo (OPP) y el Orden de Palabras de Superficie (OPS). Aquí, el concepto de Orden Profundo es solamente una abstracción útil que le permite a la autora explicar ciertos fenómenos del lenguaje referentes al orden de palabras. Se plantean cinco tipos de órdenes posibles:

- 1) Orden de Palabras Profundo Fijo;
- 2) Orden de Palabras Profundo Libre;

- 3) Orden de palabras superficial fijo;
- 4) Orden de palabras superficial libre que se corresponde con el orden de palabras profundo;
- 5) Orden de palabras superficial libre que no se corresponde con el orden de palabras profundo.

Las lenguas naturales deben compaginar la tendencia a la regularidad (la coincidencia entre las propuestas del Orden Profundo y el Orden de Superficie) con la tendencia de carácter psicolingüístico a la irregularidad (ruptura entre las dos estructuras), que tiene que ver con la necesidad de romper la línea temporal de la comunicación verbal [7].

2.3. Holan et al

Holan y sus seguidores [3] presentan un sistema de formalización del orden de palabras basados en postulados matemáticos. Parten de una visión crítica de la Functional Generative Description, como la “proyectividad” (projectivity), para intentar establecer una serie descriptiva de los cambios del orden en la oración.

Su teoría está sustentada en el concepto de “hole” (hueco, agujero), que indica una relación matemática entre el número de palabras de una oración y el número posible (máximo y mínimo) de huecos estructurales previos. Dichos huecos, previstos ya en la cobertura del nodo, pueden dar lugar a cambios del orden a partir de la intercalación de elementos que no pertenecen a ella y que se situarían en el hueco.

En la última parte de su artículo, Holan y sus seguidores señalan las posibles aplicaciones de las investigaciones realizadas, en la lingüística computacional. Establecen que existe una diferencia profunda entre la complejidad de la tarea de *parsing* y la tarea de *checking-grammar*, entendido lo primero, o bien como la indagación sobre cuál es la estructura más probable o bien como el sistema de producción del conjunto de todas las oraciones posibles; y, entendiendo lo segundo, como la verificación de si una estructura pertenece o no a un determinado conjunto de cadenas de estructuras. Todo esto ocurre principalmente si se tiene en cuenta los *adjuntos libres* de la oración.

3. SOBRE EL ORDEN DE PALABRAS Y LA DISTRIBUCIÓN DE LA INFORMACIÓN

En la comunicación, se introducen elementos que proporcionan información nueva al oyente, a la vez que se mencionan otros que se suponen conocidos. Al respecto, se ha coincidido en dividir las oraciones, desde una perspectiva funcional, en dos componentes diferenciados: tema y foco [8]. En el orden no marcado (también denominado “orden objetivo”), el tema precede al foco. No obstante, ambos pueden alterar su ordenación por medio de mecanismos de movimiento, lo que implica hablar de órdenes “subjativos” [6].

La colocación de las unidades en la cadena lingüística es un recurso de marcación. La marcación se utiliza para resaltar aquellos aspectos lingüísticos que el hablante considera más relevantes y que puede manifestarse a partir de procedimientos morfológicos, sintácticos o semánticos, dependiendo del tipo de elemento que se desea instrumentalizar como marcador. A su vez, dentro de este fenómeno, deben considerarse también las funciones informativas referentes al ‘tema’ y al ‘foco’, que son de gran importancia en la descripción de ciertos órdenes de palabras. El tema es aquello de lo que trata la oración, lo que se dice sobre el tema se denomina ‘comentario’. Zubizarreta [9]

distingue dos tipos de temas: el ‘tema discursivo’ y el ‘tema oracional’. Así por ejemplo, dada una cláusula como (1):

(1) [María es una cocinera muy buena, pero deja la cocina hecha un desastre.]

pueden desprenderse, entre otros, dos temas posibles:

- María;
- La habilidad de María como cocinera.

El primero puede considerarse tema de la oración (aunque quizá también pueda funcionar como tema del discurso); el segundo, en cambio, solo puede identificarse como tema discursivo. El tema discursivo únicamente brinda información sobre el tema, es decir, comenta sobre el tema; el tema oracional, por otro lado, es el sujeto de un predicado lógico. El predicado del tema se corresponde con el ‘comentario’ [9].

Más allá de que el tema oracional pueda asociarse a distintas posiciones dentro de la cláusula, en varias lenguas, entre ellas el español, hay ciertas posiciones que pueden funcionar exclusivamente como tema; tal es el caso de la posición periférica de la oración. Según Zubizarreta, se distinguen dos tipos de construcciones con temas ubicados en el extremo izquierdo de la cláusula. Una de ellas es la conocida como *Hanging Topic*, traducida por la autora como ‘Tema Vinculante’ y la otra es la ‘Dislocación a la izquierda’. El tema vinculante se distingue de la dislocación a la izquierda, desde el punto de vista discursivo, en que el primero tiene como función cambiar de tema en un discurso dado, por ello puede estar precedido facultativamente por expresiones del tipo ‘en cuanto a’, ‘con respecto a’, etcétera. Por ejemplo:

(2) [(**En cuanto a**) **Juan**, parece que el barrio habla de **él** con sorna.]

Zubizarreta sintetiza las propiedades del tema vinculante y la dislocación a la izquierda con los siguientes ítems.

Tema vinculante

1. Introduce un cambio de tema discursivo (3).

(3) [(**En cuanto a**) **Pedro**, todo el mundo desconfía de **él**.]

2. Aparece exclusivamente en la periferia de la cláusula matriz (4), (5).

(4) [**China**, por supuesto, es completamente sabido que todo el mundo comenta el desarrollo económico de **esa** nación.]

(5) *[Es completamente sabido que, **China**, todo el mundo comenta el desarrollo económico de **esa** nación.]

3. La relación entre el tema y una cierta posición dentro de la oración es una relación de correferencia; no existe relación de dependencia gramatical (ver cláusulas (4) y (5)).

4. La relación no está restringida sintácticamente: el tema vinculante puede entrar en relación con cualquier posición dentro de la oración (6), (7), (8), (9), (10).

(6) [**En cuanto al** **perro**, parece que los padres **lo** miman demasiado.]

- (7) [**En cuanto al perro**, parece que **el desgraciado** no quiere comer carne cruda.]
- (8) [(En cuanto a) Juan, conocemos a la mujer que lo abandonó.]
- (9) [(En cuanto a) Juan, vamos a asegurarnos bien antes de llamarlo para contarle.]
- (10) [(En cuanto a) Juan, que María lo haya engañado sorprendió a todo el mundo.]

Dislocación a la izquierda

1. El tema puede aparecer en la periferia izquierda de la cláusula matriz o en la periferia de la cláusula subordinada (11), (12).

- (11) [**A sus hijos**, María **los** llevó al cine.]
- (12) [Estoy seguro de que **a sus hijos**, María **los** llevó al cine.]

2. Existe una dependencia gramatical entre el tema y la posición dentro de la cláusula con la cual se relaciona, como lo indica la presencia de la preposición adyacente al tema nominal en ciertos casos (13), (14).

- (13) [Es seguro que **de Juan**, todo el mundo habla.]
- (14) [Es seguro que **a Juan**, María **le** dice cosas que no son verdad.]

3. La relación está restringida sintácticamente: el tema no puede entrar en relación con una posición dentro de una cláusula relativa, de una cláusula adverbial o de una cláusula sujeto (15), (16), (17).

- (15) *[Es seguro de que **a Juan**, conocemos a la mujer que **lo** abandonó.]
- (16) *[Creo que es más conveniente que **a Juan**, nos cercioremos de la verdad antes de llamarlo.]
- (17) *[Es seguro que **a Juan**, que María **lo** haya engañado sorprendió a todos.]

El foco es la parte no presupuesta de la oración. Un mismo sintagma no puede funcionar simultáneamente como tema y como foco y, mientras que el tema es el sujeto lógico de la predicación, el foco es parte del predicado, de la información que se predica del tema. Zubizarreta presenta un análisis de este término enfocado en la acentuación de la cláusula para luego proponer una serie de posibles órdenes de palabra.

Ni el tema ni el foco pueden considerarse como unidad atómica. En una cláusula, puede haber más de un tema o ninguno y ni este ni el foco se relacionan necesariamente con la división en constituyentes, por más que sea cierto que en muchas ocasiones tema y sujeto coinciden. [6].

Con respecto a las dislocaciones, Francesconi [10], en su estudio sobre la dislocación en el español y el italiano, observa que si bien se ha escrito mucho sobre la dislocación a la izquierda, no ha ocurrido lo mismo con la dislocación a la derecha. Una probable causa de ello es que, en el español, esta construcción no es demasiado frecuente. Ejemplo de dislocación a la derecha sería:

- (18) [No lo soporto más, **a tu hermano**.]

Lo interesante del trabajo de Francesconi es la diferenciación entre la dislocación y la “conjugación objetiva”. En español, el fenómeno de la reduplicación de clíticos interfiere con las dislocaciones, pero, si el clítico del objeto directo parece más a un nombre recapitulativo, en los casos de objeto indirecto, los clíticos revelan su naturaleza de marcadores de concordancia (conjugación objetiva). A tales efectos, para poder distinguir en español los derivados de la conjugación objetiva de las dislocaciones, hay que valerse de un “ulterior elemento diagnóstico”, esto es, se necesita que aparezca una pausa señalada por una coma que separe el constituyente del núcleo frasal. [10]. A continuación, ejemplos tomados del artículo del autor:

Conjugación objetiva

(19) [¿Qué le habéis hecho al Ignacio?]

(20) [A Juan lo vi ayer.]

Dislocación a la izquierda

(21) [A Lucía, la he visto que comía.]

Dislocación a la derecha (también llamada Ripensamento)

(22) [¿Qué le habéis hecho, al Ignacio?]

4. ALGUNAS PARTICULARIDADES SOBRE EL ORDEN DE PALABRAS EN ESPAÑOL

A diferencia de otras lenguas con estructura básica lineal, el español tiene una flexibilidad mayor. Se trata de una lengua con la capacidad de alterar la construcción sintáctica prototípica y hacer prevalecer otros factores de carácter expresivo. A partir de este planteo, pueden distinguirse dos tipos de órdenes, un orden más fijo y un orden libre. Con respecto al segundo, dicha libertad para alterar el orden básico se ha relacionado con la riqueza de las marcas flexivas nominales y con la existencia de un sistema de preposiciones [6].

En el español coexisten distintos órdenes relativos posibles. Tal así, que, por ejemplo, el sujeto de una oración con verbo transitivo puede estar en la posición inicial (23), inmediatamente detrás del verbo (24) y detrás del complejo formado por el verbo y su objeto directo (25).

(23) [Juan donó su casa.]

(24) [Donó Juan su casa.]

(25) [Donó su casa Juan.]

No obstante, hay determinados contextos en los que la posición del sujeto es obligatoriamente posverbal. Esto puede deberse a el tipo de construcción, por ejemplo en estructuras interrogativas (principales y subordinadas); las propiedades del sujeto, generalmente los SN sin determinantes no son sujetos preverbiales; el tipo de verbo, la ergatividad o la inacusatividad; etcétera.

En el caso de los adverbios modificadores del sintagma verbal, estos pueden aparecer antes o después del verbo, a diferencias de otras lenguas. La excepción la da el adverbio de negación ‘no’, que únicamente admite la posición preverbal.

Los clíticos se ubican a la izquierda del verbo, inmediatos a este, o bien pegados al final. La manera en que estos pronombres se combinan con el verbo y entre sí se aborda en el análisis del sintagma verbal núcleo flexionado del español, de Bès y Solana [11].

Por último, los adjuntos o circunstanciales pueden colocarse en varias posiciones: delante del OD, entre el OD y el OI o al final de la oración, y no parece haber restricciones con respecto al arreglo que establecen los distintos circunstanciales entre sí.

5. ANÁLISIS E IMPLANTACIÓN EN MÁQUINA

De acuerdo con lo expuesto hasta aquí, planteo lo siguiente: Es posible establecer un orden básico en el español determinado por la estructura Sujeto – Verbo – Objeto Directo – Adjuntos y, sobre la base de esa estructura, órdenes “derivados”, los que se deben a cuestiones gramaticales, como así también, a la distribución de la información en la cláusula. Tanto en el primer caso, como en el segundo, se pueden alterar dichos órdenes mediante el uso de coma.

Los elementos desplazados pueden ubicarse, como se ha mencionado, en la periferia izquierda o en la derecha de la oración. Asimismo, también es posible hallar alteraciones no ubicadas en los extremos de la cláusula, sino en una posición más intermedia (26).

(26) [El niño, **momentos antes del comienzo de la función**, le pidió a su madre permiso para ir al baño.]

En este caso, se está ante una dislocación a la izquierda, pero en la que el elemento que se corrió no llega al inicio de la oración. Tal fenómeno es considerado por algunos autores como “inciso reubicable”, pues, se trata de un elemento que se movió de su lugar habitual para ubicarse en una posición incidental.

El trabajo de implantación en máquina está focalizado en la detección de los elementos desplazados a la izquierda (se traten de construcciones de tema vinculante o dislocaciones a la izquierda propiamente dichas). Por el contrario, no se tratarán las dislocaciones a la derecha. Se focalizó en aquellos complementos circunstanciales desplazados que se ubican al principio de la oración y que están conformados por un sintagma preposicional (27) o adverbial (28) y los objetos directos encabezados por la preposición “a” (29).

(27) [**De mala manera**, Juan increpó a su esposa.]

(28) [**Como siempre**, Pedro llegó tarde.]

(29) [**A su esposa**, Juan la vio con otro hombre.]

Para establecer reglas de reconocimiento de estas construcciones, debió tenerse en cuenta el punto de la oración anterior a la analizada. Es decir, si luego de un punto de fin de oración aparecía un sintagma preposicional o adverbial y una coma, entonces había alteración del orden. A modo de ejemplo:

(...) cuándo le darán el alta. **Más allá de la broma**, el médico Luis Buonomo señaló que (...)

Además, también se consideró la posibilidad de que hubiese entre el punto y el constituyente desplazado, un marcador discursivo (a), una conjunción copulativa (b) o adversativa (c), o una conjunción seguida de un marcador discursivo (d):

- (a) . **Sin embargo, con tenacidad**, la siguió buscando.
- (b) . **Y con tenacidad**, la siguió buscando.
- (c) . **Pero con tenacidad**, la siguió buscando.
- (d) . **Y sin embargo con tenacidad**, la siguió buscando.

Fue posible también reconocer a más de un constituyente desplazado. Por ejemplo:

(30) [**Ayer, en casa de María**, sonó el timbre.]

En este caso, se cuenta, después de la primera coma, un nuevo sp o sadv más una segunda coma.

5.1. Análisis morfológico y reconocimiento de signos de puntuación

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. Aquí, la coma debe declararse en el archivo de 'entradas' que es el diccionario fuente que utiliza el programa para proceder al análisis morfológico. A modo de ejemplo, el análisis del siguiente fragmento:

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

Smorph da como 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'].
 ['tensor', 'EMS', 'v', 'MODOV', 'ind', 'PERS', '1a', 'NUM', 'sg', 'TPO', 'pres', 'TR', 'r', 'TC', 'c1'].
 ','.
 ['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', 'c1'].

'a'.

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

'declarar'.

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

,

'.

['linsig', 'EMS', 'pun']. [12]

El output generado por Smorph va a ser el input con el que trabajará MPS. Una vez obtenidos los resultados de Smorph, se constituyeron las reglas de reconocimiento de los diversos sintagmas. A partir de ellos, es posible establecer luego, las reglas para la detección de los elementos desplazados al inicio de la cláusula, indicados por comas.

5.2. Reconocimiento de alteraciones con MPS

En este estadio de la detección automática con MPS, en primer lugar se declararon las reglas para el reconocimiento de sintagmas nominales –de nombres comunes (SN) y de nombres propios (snnpr)–, adjetivales (sadj), verbales (sv), preposicionales (sp) y adverbiales (sadv). En el caso de los sp, se consideró que estuvieran conformados por una preposición más un SN, que a su vez podía contener un sp en su interior (ejemplo: ‘en la casa *de mi suegra*’); o un snnpr (‘en Rosario). Para los sintagmas adverbiales, se determinó que tuvieran un adverbio actuando como núcleo del sintagma y los siguientes elementos optativos:

- Un especificador a la izquierda del núcleo dado por un cuantificador (adverbio que implica una noción de cantidad; ejemplo: ‘más allá’) y adyacentes (‘siempre alegre’);
- Un sp actuando como complemento (‘más allá de la broma’).

No se tuvieron en cuenta a las subordinadas adverbiales, quedando estas para trabajos futuros.

Se crearon reglas para el reconocimiento en Mps a partir de los elementos mencionados. Tanto para uno como otro programa, se crearon reglas que respondían a las siguientes modelizaciones:

- Punto + [(sp o sadv) + coma] ≥ 1 [13]
- Punto + conj. + [(sp o sadv) + coma] ≥ 1
- Punto + marc. disc. + coma + [(sp o sadv) + coma] ≥ 1
- Punto + conj. + marc. disc.+ coma+ [(sp o sadv) + coma] ≥ 1

Como ejemplo, se presenta la regla correspondiente al primer ítem y con un único elemento desplazado:

%punto+sp+coma da alter%

S1 [L1, 'EMS', 'pun']

S2 [L2, 'EMS','sp']

S3 [L3, 'EMS','coma']

--> S1+S2+S3 [L1+L2+L3, 'EMS', 'ALTER-1'].

%. En el acto,%

He aqui algunos ejemplos de la detección lograda:

(...) le darán el alta

' . Más allá de la broma ,'

['pfp más allá de el broma cc', 'EMS', 'ALTER-1'].

el médico Luis Buonomo señaló que (...)

(...) la bailaora Sara Baras

' . De esta última ,'

['pf de esta último cc', 'EMS', 'ALTER-1'].

Mariana Pineda”, de Federico García Lorca, en versión para ballet flamenco con dirección de Luis Pasqual, fue asimismo un montaje muy celebrado. (...)

(...) terminaron aburriendo

' . Pero independientemente de la calidad ,'

['pfp pero independientemente de el calidad cc', 'EMS', 'ALTER-1'].

la gente llenó las salas en todas las funciones pagas (entre 8 y 65 dólares) y se apiñó en las plazas y espacios al aire libre, con espectáculos gratis. (...)

(...) cultural y moral

' . Al mismo tiempo ,'

['pf al mismo tiempo cc', 'EMS', 'ALTER-1'].

se impone lanzar una fuerte desgravación impositiva para las empresas (...)

Sobre un total de 72 alteraciones de este tipo, se lograron reconocer 62 y se marcó erróneamente 2, lo que implica un 97,22% de precisión y un 86% de cobertura.

6. CONSIDERACIONES FINALES

Se presentaron los enfoques teóricos de Greenberg, Koktova y Holan sobre el orden de palabras y

posteriormente, la relación entre el orden de las palabras y la presentación de la información. En tercer lugar, se trajeron a colación algunas cuestiones referentes al orden de palabras en español.

En el trabajo de implantación en máquina, se propuso un método de detección automática de aquellos elementos desplazados de su orden habitual y ubicados al inicio de la cláusula. Sobre un total de 72 casos, se detectaron 62 alteraciones y se marcaron 2 de manera errónea, lo que implica un 97,2% de precisión y un 86% de cobertura.

Referencias

- [1] Greenberg, J. *Universals of Language*. MIT Press, Cambridge, 1963.
- [2] Koktova, E. *Word-order Based Grammar*. Mouton de Gruyter, Berlín, 1999.
- [3] Holan T. et Al. "On Complexity of Word Order". ÚFAL TECHNICAL Report, Universitas Carolina Praguensis, Praga, 2000.
- [4] Ait-Mokhtar, S. *L'analyse présyntaxique en une seule étape*. Tesis doctoral. Universidad Blaise-Pascal/Grilll, Clermont-Ferrand, 1998.
- [5] Abbaci, F. *Développement du Module Post-Smorph*. Memória del DEA de Linguistique et Informatique. Universidad Blaise-Pascal/GRIL. Clermont-Fd.
- [6] Fernández Soriano, O. *Sobre el orden de palabras en español*. Dicenda, Cuadernos de Filología Hispánica, N° 11, Edit. Complutense, Madrid, 1993.
- [7] Padilla García, X. *El orden de palabras en el español coloquial*. Tesis doctoral. Universitat de València, Facultad de Filología, Departamento de Filología Española. Valencia, 2001.
- [8] En el presente trabajo no se va a entrar en la discusión terminológica respecto de los términos 'dislocación (a la izquierda o a la derecha)', 'tematización', 'topicalización', 'rematización', etcétera y se mantendrán las denominaciones propuestas por Zubizarreta.
- [9] Zubizarreta, M. "Las funciones informativas: Tema y Foco", en Bosque I. y Demonte V. (Dir.), *Gramática descriptiva de la lengua española*, Tomo III, Espasa Calpe, Madrid, 1999.
- [10] Francesconi, A. *La dislocación en la sintaxis italiana y española*. AISPI, Actas XXIII, Centro Virtual Cervantes, 2005.
- [11] Bès, G. y Solana, Z. *Sintagma verbal núcleo flexionado en español*. Revista Infosur, N° 1, 2007.
- [12] Referencias: 'EMS', Estructura Morfosintáctica; 'prep', preposición; 'det', determinante; 'nom', nombre; 'GEN', género; 'masc', masculino; 'NUM', Número; 'sg', singular; 'adj', adjetivo; 'v', verbo; 'MODOV', Modo Verbal; 'ind', indicativo; 'PERS', Persona, 'TPO', Tiempo; 'pres', presente; 'TR', Tipo de Regularidad; 'r', regular; 'TC', Tipo de Conjugación; 'c1', primera conjugación; 'npr', nombre propio; 'cl', clítico; 'prets', pretérito perfecto simple; 'ir', irregular; 'infin', infinitivo; 'linsg', línea siguiente, 'pun', punto.
- [13] Con ≥ 1 se indica que, por lo menos, debe haber uno de los elementos mencionados.