Evaluación de un etiquetador morfosintáctico basado en bigramas especializados para el castellano

Ferran Pla, Antonio Molina y **Natividad Prieto**
Departament de Sistemes Informàtics i Computació
Universitat Politècnica de València
Camí de Vera s/n
46020 València
{fpla,amolina,nprieto}@dsic.upv.es

Resumen En este artículo se describe un sistema de etiquetado morfosintáctico basado en bigramas especializados que incorporan información de ciertas palabras del vocabulario en determinados contextos. La técnica propuesta para la construcción de estos modelos se basa en el aprendizaje a partir de datos, por lo que se puede aplicar de forma directa a cualquier lenguaje y a diferentes tareas de desambiguación.

El sistema de etiquetado morfosintáctico se ha evaluado sobre el corpus en castellano *LexEsp*, obteniéndose un 97.4% de precisión global, y un 93.5% de precisión sobre el conjunto de las palabras ambiguas. Este resultado y su comparación con los obtenidos utilizando otros sistemas de etiquetado propuestos en la literatura, confirman la viabilidad de la aproximación que aquí se presenta.

1 *Introducción*

En los últimos años se ha observado un interés creciente por la utilización de técnicas inductivas o basadas en corpus en prácticamente todos los campos de estudio del área de procesamiento del lenguaje natural. El principal atractivo de estas técnicas es que permiten alcanzar resultados muy satisfactorios sin que sea necesaria la intervención humana en el proceso de resolución. Se han aplicado con éxito en la resolución de problemas como el etiquetado morfosintáctico, el análisis sintáctico superficial, la detección de constituyentes de la oración, la resolución del ligamiento preposicional, etc.

Las características comunes de los métodos basados en corpus son que generalmente dan lugar a modelos estadísticos sencillos, cuyos parámetros se estiman a partir de datos y que suelen permitir un alto grado de modularidad y portabilidad; de esta forma, su uso para la resolución de diferentes tareas sobre diferentes lenguajes suele realizarse sin requerir modificaciones importantes. También comparten como característica que permiten incorporar ciertas reglas o restricciones que son bien conocidas por los expertos humanos y que mejoran substancialmente las prestaciones de los sistemas.

En concreto, para el problema del etiquetado morfosintáctico de textos no restringidos, las técnicas que mejores resultados proporcionan son las basadas en modelos de Markov [5][2], las basadas en reglas de transformación [3], los métodos basados en memoria [6], y los basados en el principio de máxima entropía [17].

En este artículo se presenta un sistema de etiquetado morfosintáctico que utiliza lo que hemos denominado *modelos contextuales especializados*. Éstos son modelos de bigramas de categorías léxicas enriquecidos con información de ciertas palabras del vocabulario en determinados contextos. Estos modelos permiten reflejar dependencias léxico-contextuales, que en muchos casos ayudan de manera notable a resolver ciertas ambigüedades estructurales.

El sistema propuesto se ha evaluado experimentalmente sobre el corpus en castellano *LexEsp*, estableciendo comparaciones con aproximaciones desarrolladas por otros investigadores bajo las mismas condiciones de experimentación.

2 *Etiquetado morfosintáctico utilizando modelos de Markov*

Desde el punto de vista estadístico, el etiquetado morfosintáctico de textos (*POS tagging*) se puede definir como un problema de maximización. Sea $\mathcal{C} = \{c_1, c_2, ..., c_N\}$ un conjunto de categorías léxicas y $\mathcal{V} = \{w_1, w_2, ..., w_m\}$ el vocabulario de la aplicación. Dada una frase de entrada $w =$

$w_1, \ldots, w_T$ de longitud T, se trata de encontrar la secuencia de categorías léxicas de máxima probabilidad en el modelo, es decir:

$$\begin{aligned} \widehat{c} &= \arg\max_{c \in \mathcal{C}^T} P(c|w) \\ &= \arg\max_{c \in \mathcal{C}^T} \left(\frac{P(c)P(w|c)}{P(w)} \right) \end{aligned} \quad (1)$$

Dado que la maximización presentada en la ecuación anterior es independiente de la frase en entrada, w, es suficiente con maximizar el numerador de la misma, donde $P(c)$ representa las probabilidades contextuales o modelo de lenguaje y $P(w|c)$ las probabilidades léxicas.

Para la resolución de esta ecuación se suelen introducir algunas aproximaciones (asunciones de Markov), que aunque no siempre proporcionan la solución exacta, permiten obtener resultados bastante precisos con costes computacionales aceptables. Para modelos de Markov de primer orden (bigramas), el problema se reduce a resolver la siguiente ecuación:

$$\arg\max_{c_1 \ldots c_T} \left(\prod_{1 \ldots T} P(c_i|c_{i-1}) P(w_i|c_i) \right) \quad (2)$$

Los parámetros de esta ecuación se pueden representar como un modelo de Markov en el que los estados están etiquetados con una categoría léxica, las probabilidades de contexto se corresponden con las probabilidades de transición entre estados y las probabilidades léxicas con las probabilidades de emisión de símbolos en cada estado. El proceso de etiquetado se puede llevar a cabo de manera eficiente mediante programación dinámica utilizando el algoritmo de Viterbi.

3 *Modelos de Markov parcialmente lexicalizados*

La forma más directa de introducir información que amplíe el contexto considerado en los modelos de bigramas, es el uso de formalismos de modelos de Markov de ordenes superiores (típicamente trigramas); esta ampliación, al considerar una historia mayor, define un número de parámetros muy elevado y presenta inconvenientes en su estimación. Para soslayar este inconveniente se suelen seguir básicamente dos tendencias 1) considerar dependencias de mayor longitud sólo en ciertos casos preestablecidos, con el fin de mantener tamaños de modelos aceptables y 2) hacer intervenir las palabras en el modelo contextual para establecer nuevas restricciones estructurales que describan relaciones entre las palabras y sus categorías.

Así por ejemplo, en [1] se considera un modelo de Markov de primer orden que es ampliado, duplicando ciertos estados y fusionando otros, con el fin de poder tener en cuenta mayor información de sus predecesores. Otros autores hacen intervenir las palabras en los modelos contextuales, como en [9] (sólo para un cierto número de palabras) o en [10] utilizando modelos de Markov totalmente lexicalizados, con el consiguiente aumento del tamaño del modelo y la aparición de nuevos problemas asociados a la estimación de sus parámetros.

Dentro del campo de la inferencia gramatical encontramos algunas aproximaciones que adoptan ciertos criterios similares para establecer restricciones estructurales más complejas. La metodología de inferencia gramatical mediante generadores mórficos [7], también utilizada en [18], define una función de etiquetado, de acuerdo a un cierto criterio definido por un experto, que produce un reetiquetado del conjunto de muestras de aprendizaje y en consecuencia, un cambio en el alfabeto. Con esta función se permite distinguir ciertos símbolos, dependiendo por ejemplo de la posición en la que se encuentren en una determinada cadena, con lo que se consiguen modelizaciones contextuales (bigramas) más complejas.

En este trabajo se presenta una técnica de especializacion del modelo contextual subyacente en un modelo de Markov (y en general, en cualquier modelo regular), con la incorporación de ciertas palabras al mismo, además de las categorías léxicas, y así poder establecer ciertas restricciones de contexto ligadas al léxico.

La especialización de ciertas palabras según sus categorías léxicas, elegidas teniendo en cuenta criterios lingüísticos o de manera automática a partir del conjunto de entrenamiento, redunda en una mejor modelización contextual como se mostrará en los resultados que se presentan en este y otros artículos [14]. Algunos de los criterios que se pueden definir para elegir el conjunto de palabras objeto de la especialización son los siguientes: las palabras más frecuentes, las

palabras con mayor error de etiquetado y las palabras pertenecientes a categorías cerradas. La especialización de determinadas palabras permite, en cierto modo, incorporar conocimiento lingüístico a los modelos, y así posibilitar la representación de restricciones estructurales que no se podrían establecer considerando solamente las categorías léxicas.

4 *Formulación del proceso de especialización*

En este apartado se describe el proceso que se realiza sobre el conjunto de muestras de aprendizaje con el fin de poder obtener modelos contextuales especializados.

Dado el conjunto de categorías morfosintácticas $\mathcal{C}$, el vocabulario de la aplicación $\mathcal{V}$ y un conjunto de entrenamiento $E \subset (\mathcal{V} \times \mathcal{C})^*$ formado por pares de palabra y etiqueta $(\langle w_1, c_1\rangle, ..., \langle w_M, c_M\rangle)$, el objetivo es conseguir un nuevo conjunto de entrenamiento con información léxico-contextual.

Para ello se define un conjunto, $\mathcal{W}_e \subset \mathcal{V}$, formado por las palabras que se considerarán en el modelo contextual y una función de especialización f_e que, a partir del conjunto de entrenamiento original (E), proporciona un nuevo conjunto de entrenamiento $(\hat{E})$ sobre el que se aprenderá el modelo especializado. Esta especialización hace que, el conjunto de categorías original $\mathcal{C}$, se incremente con las palabras consideradas en $\mathcal{W}_e$, especializadas en sus diferentes categorías morfosintácticas. El nuevo conjunto $\hat{\mathcal{C}} \subset ((\mathcal{W}_e \cup \lambda) \times \mathcal{C})$ queda determinado aplicando una función f_e, definida de la siguiente manera:

$$f_e : E \subset (\mathcal{V} \times \mathcal{C})^* \rightarrow \hat{E} \subset (\mathcal{V} \times \hat{\mathcal{C}})^*$$

$$f_e(\langle w_i, C_i\rangle) = \begin{cases} \langle w_i, (w_i, c_i)\rangle & \text{si } w_i \in \mathcal{W}_e \\ \langle w_i, (\lambda, c_i)\rangle & \text{si } w_i \notin \mathcal{W}_e \end{cases}$$

Como se puede observar, la función f_e realiza un reetiquetado del conjunto de entrenamiento original (E) consistente en sustituir la etiqueta C_i de la palabra w_i, por la nueva etiqueta (w_i, C_i) si w_i pertenece al conjunto de palabras a especializar, o bien, conservar la etiqueta existente si no pertenece a dicho conjunto (λ, C_i), donde en este caso λ representa la cadena vacía.

5 *Descripción del sistema*

En el sistema de etiquetado que se propone se pueden distinguir dos fases: aprendizaje y etiquetado.

5.1 Fase de aprendizaje

El aprendizaje de los modelos se realiza utilizando corpora etiquetados con categorías morfosintácticas. A partir de la secuencia de categorías morfosintácticas se aprenden los modelos contextuales: bigramas o bigramas especializados. En este último caso se utilizará un corpus de entrenamiento especializado tal y como se ha presentado en la sección anterior. En cualquier caso, los modelos construidos se suavizan utilizando una técnica de *back-off* [8].

Las probabilidades léxicas se estiman por máxima verosimilitud a partir de las frecuencias de las palabras, categorías, y de cada palabra en cada categoría. La talla del vocabulario (del orden de catorce mil palabras) y el uso de un analizador morfológico, que proporciona el conjunto de categorías para cada palabra, garantiza la fiabilidad de la estimación. Se asume que las palabras desconocidas para el analizador morfológico pueden pertenecer a cualquier categoría abierta y su probabilidad se aproxima mediante la probabilidad de la categoría correspondiente.

A continuación se presenta un ejemplo en el que se pretende modelizar de forma separada la aparición de la palabra *que* como pronombre relativo (PR), de su aparición como conjunción subordinante (CS). Para ello, se incluye esta palabra en el conjunto de palabras a especializar $\mathcal{W}_e$ (ver figura 1).

En este caso, la aplicación de la función de especialización f_e únicamente remplaza en el conjunto E, el par *que/PR* por el par *que/(que,PR)* y el par *que/CS* por *que/(que,CS)*.

En la parte izquierda de la figura 1 se puede observar el modelo de Markov obtenido a partir del conjunto de entrenamiento E y en la parte derecha las variaciones que sobre éste se producen al considerar el nuevo conjunto de entrenamiento $\hat{E}$ cuando se especializa la palabra *que*.

Se observa cómo para la palabra *que* aparecen dos nuevos estados especializados, cuyos símbolos asociados son *(que,PR)* y *(que,CS)*, que distinguen los distintos contextos (NC–VMS y NC–VMI) en los que esta palabra ha aparecido en el conjunto de entrenamiento. Obsérvese, que estos estados especializados sólo pueden emitir la palabra *que* por lo que su probabilidad léxica debe ser igual a uno.

Conjunto de Entrenamiento (E) a especializar con $\mathcal{W}_e$= {que}:

El/TD profesor/NC que/PR vino/VMI dijo/VMI al/SP alumno/NC que/CS callase/VMS ./Fp

La/TD casa/NC donde/PR vivo/VMI es/VMI de/SP madera/NC ./Fp

Llevo/VMI paraguas/NC cuando/CS llueve/VMI ./Fp

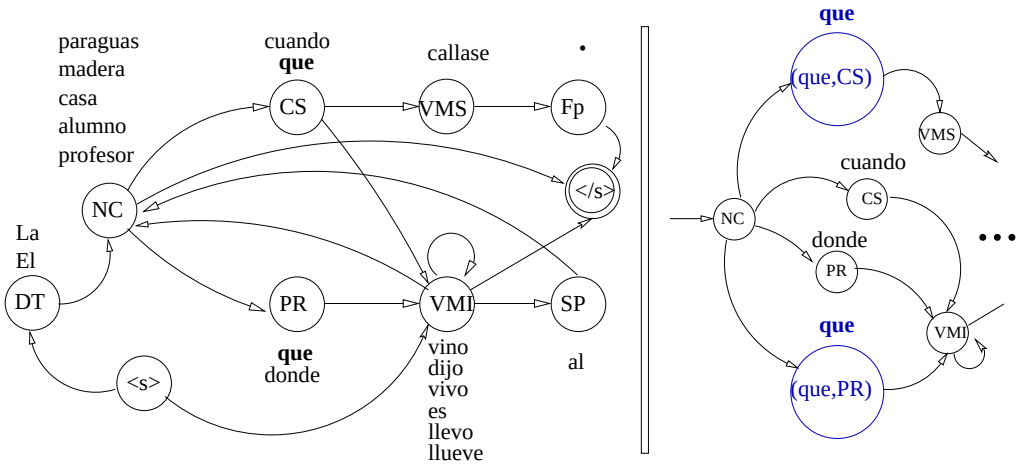

Figura 1: Ejemplo de especialización de algunos estados en un modelo de Markov.

5.2 Fase de etiquetado o desambiguación léxica

El sistema de etiquetado toma como entrada un texto no restringido el cual es convenientemente separado en unidades o *tokens*. Esta segmentación se realiza mediante el analizador morfológico MACO [4] el cual, además, proporciona las posibles categorías léxicas para cada *token* detectado. El proceso de etiquetado se realiza mediante el algoritmo de Viterbi. Se trata de encontrar la secuencia de estados de mayor probabilidad en el modelo contextual que sea compatible con el texto de entrada, teniendo en cuenta las probabilidades léxicas. Una vez obtenida dicha secuencia, como cada estado tiene asociada una única categoría léxica, se dispone de la mejor secuencia de categorías (etiquetado léxico) para la secuencia de palabras (*tokens*) de entrada.

Si se ha utilizado un modelo especializado, el proceso de etiquetado es básicamente igual, sólo es necesario aplicar una función inversa que deshaga la especialización definida. Esta función, que denotaremos por f_d, deberá proporcionar, a partir de las etiquetas consideradas en el modelo especializado, el conjunto de categorías léxicas inicial ($\mathcal{C}$) que serán asignadas finalmente a las palabras de la frase de entrada.

$$f_d : \hat{\mathcal{C}} \rightarrow \mathcal{C}$$

$$f_d\left(\langle w_i, c_i\rangle\right) = c_i \quad \text{donde} \quad w_i \in (\mathcal{W}_e \cup \lambda)$$

6 *Experimentos y resultados*

La evaluación del sistema de etiquetado morfosintáctico se ha realizado sobre el corpus en castellano *LexEsp*, utilizando el conjunto de etiquetas PAROLE (65 etiquetas). Se ha definido un conjunto de entrenamiento formado por 65,000 palabras (3,510 frases aprox.) y uno de prueba de unas 25,000 palabras (1,300 frases aprox.) obtenido del conjunto de datos ya etiquetado y supervisado manualmente. En este conjunto de frases el 39.8% de la palabras son ambiguas, presentando una ambigüedad media de 2.6 categorías/palabra (1.6 categorías/palabra sobre el total). De ellas el 0.4% son desconocidas para el analizador morfológico.

6.1 Resultados de etiquetado

El criterio de especialización elegido conduce a tomar las palabras pertenecientes a categorías cerradas. De éstas se eliminan los signos de puntuación, los nombres propios y las cifras, y se incorporan algunas palabras de frecuencia muy alta en el conjunto de entrenamiento [14]. Básicamente sólo se han

considerado preposiciones, pronombres, relativos y determinantes. Con este criterio se especializan 45 palabras que son las siguientes:

$\mathcal{W}_e$ ={*al, ambos, aunque, bajo, bien, como, contra, cuando, de, del, desde, durante, entonces, entre, eran, incluso, la, las, le, les, lo, los, mediante, misma, mismo, más, nada, ni, no, para, pero, poco, por, que, se, seguro, ser, siempre, sin, sobre, sus, sí, también, todas, todo, total, último, único*}

	Precisión	
Método	global	ambiguas
BIG	96.97%	92.39%
BIG_{esp}	97.42%	93.52%

Tabla 1: Comparación de la precisión de etiquetado léxico sobre el corpus LexEsp entre modelos BIG y BIGesp.

Con el criterio de especialización propuesto se obtienen los resultados que se muestran en la tabla 1. Se ha pasado de un 96.97% de precisión global con bigramas (BIG) a un 97.42% con bigramas especializados (BIG_{esp}). En términos absolutos supone una disminución de 114 errores (de 762 errores con *BIG* a 648 con BIG_{esp}), que considerando un intervalo de confianza del 95%, permite concluir que la utilización de bigramas especializados en el sistema de etiquetado conduce a una mejora significativa en la precisión (97.42%±0.21% frente a 96.97%±0.21%). Los porcentajes de precisión de etiquetado sobre el conjunto de palabras ambiguas son del 92.4% cuando se usan bigramas y 93.5% si se usan bigramas especializados. La especialización utilizada no afecta a la complejidad de los modelos (supone un incremento de 68 a 199 estados) ni a la eficiencia del algoritmo de etiquetado.

Se observa que la utilización de modelos lexicalizados no sólo mejora la precisión global de etiquetado, sino que también contribuye a incrementar la precisión sobre las palabras que se han especializado en el modelo contextual. En la tabla 2 se muestra un estudio comparativo de la precisión de etiquetado del sistema utilizando un modelo de bigramas y otro de bigramas lexicalizado sobre algunas

	Frecuencia		# errores	
Palabra	Train	Test	BIG	BIG_{esp}
que	1680	709	109	83
lo	268	140	53	13
los	1188	371	17	11
la	2530	869	12	7
las	743	270	4	3

Tabla 2: Frecuencia de aparición de ciertas palabras especializadas y variación en el número de errores en su etiquetado al utilizar bigramas especializados.

de las palabras ambiguas del conjunto $\mathcal{W}_e$ definido, en concreto, las palabras *que, lo, los, la, las.* Estas palabras son las que han experimentado mejoras más significativas con la utilización de modelos lexicalizados. No obstante, su especialización contribuye favorablemente a mejorar las tasas globales de precisión de etiquetado.

6.2 Comparación experimental con otras aproximaciones

En la tabla 3 se presenta un estudio comparativo de los resultados obtenidos con nuestro sistema frente a los que se obtienen usando algunas aproximaciones propuestas por otros investigadores. La evaluación de los distintos sistemas se hace sobre el mismo corpus y bajo las mismas condiciones de experimentación, es decir, se consideran los mismos conjuntos de aprendizaje y prueba y las mismas restricciones.

Método	Precisión
TT	97.00%
RB	97.30%
RT	97.20%
RBT	97.40%
BIG	96.97%
BIGesp	97.42%

Tabla 3: Comparación de las prestaciones de etiquetado léxico de nuestro sistema con otras aproximaciones sobre el corpus LexEsp.

Los sistemas con los que ha realizado la comparación son los siguientes:

- **TT**: etiquetador basado en árboles de decisión. Bajo esta aproximación el problema de etiquetado se plantea como un

problema de clasificación. El modelo de lenguaje está constituido por un conjunto de árboles de decisión estadísticos que se corresponden con ciertas clases de ambigüedad. Teniendo en cuenta estas clases, las probabilidades léxicas a priori de las palabras se recalculan dependiendo del camino seguido en el árbol [11].

- **R**: etiquetador basado en un modelo híbrido que utiliza técnicas de relajación [13] para combinar diferentes fuentes de conocimiento como: bigramas (**RB**), trigramas (**RT**), o ambos (**RBT**).

Se observa que los mejores resultados de precisión de etiquetado morfosintáctico se obtienen cuando se utilizan modelos de bigramas especializados y son comparables a los obtenidos mediante el método de relajación con bigramas y trigramas.

7 Conclusiones

En este artículo se ha presentado un método para construir modelos de Markov lexicalizados y se han incorporado en un sistema de etiquetado morfosintáctico de textos no restringidos. Se ha definido un criterio de lexicalización consistente en incorporar en el modelo de lenguaje información contextual sobre aquellas palabras del conjunto de entrenamiento pertenecientes a categorías cerradas. Este criterio de lexicalización puede ser aplicado automáticamente de manera sencilla a cualquier conjunto de entrenamiento independientemente de la lengua y del conjunto de etiquetas morfosintácticas utilizado.

La evaluación del sistema de etiquetado morfosintáctico sobre el castellano, en concreto sobre el corpus *LexEsp*, demuestra la viabilidad de la propuesta realizada. La utilización de bigramas especializados conduce a mejoras significativas respecto al uso de modelos no especializados, y con respecto a las aproximaciones contrastadas. Sería interesante realizar un estudio cualitativo de la relación existente entre las palabras especializadas en determinados contextos sintácticos y su efecto sobre el etiquetado.

Al tratarse de una aproximación basada en corpus, su aplicación a otras lenguas o tareas de desambiguación se puede realizar de forma directa. En [15] se presentan resultados de etiquetado sobre el corpus en inglés Penn Treebank. También para el inglés, se han utilizado modelos contextuales lexicalizados o especializados en el análisis sintáctico superficial [16] y para la segmentación de frases en cláusulas [12].

8 Agradecimientos

Este trabajo se ha desarrollado en el marco del proyecto *TUSIR* subencionado por la CICYT con número TIC2000-0664-C02-01. Los autores agradecen las sugerencias recibidas de los revisores de este trabajo.

Referencias

[1] T. Brants. Estimating Markov model structures. In *Proceedings of 4th International Conference on Spoken Language Processing*, 1996.

[2] T. Brants. TnT – a statistical part-of-speech tagger. In *Proceedings of the Sixth Applied Natural Language Processing (ANLP-2000)*, Seattle, WA, 2000.

[3] E. Brill. Transformation–based Error–driven Learning and Natural Language Processing: A Case Study in Part–of–speech Tagging. *Computational Linguistics*, 21(4):543–565, 1995.

[4] J. Carmona, S. Cervell, L. Màrquez, M.A. Martí, L. Padró, R. Placer, H. Rodríguez, M. Taulé, and J. Turmo. An Environment for Morphosyntactic Processing of Unrestricted Spanish Text. In *Proceedings of the 1st International Conference on Language Resources and Evaluation, LREC*, pages 915–922, Granada, Spain, May 1998.

[5] K. W. Church. A Stochastic Parts Program and Noun Phrase Parser for Unrestricted Text. In *Proceedings of the 1st Conference on Applied Natural Language Processing, ANLP*, pages 136–143. ACL, 1988.

[6] W. Daelemans, J. Zavrel, P. Berck, and S. Gillis. MBT: A Memory–Based Part–of–speech Tagger Generator. In *Proceedings of the 4th Workshop on Very Large Corpora*, pages 14–27, Copenhagen, Denmark, 1996.

[7] P. García, E. Vidal, and F. Casacuberta. Local Languages, the Sucessor Method, and a Step towards a General Methodology for the Inference of Regular Grammars. *IEEE Transactions on Pattern Analysis and Machine Intelligence, PAMI*, 9(6):841–845, 1987.

[8] Katz, S. M. (1987). Estimation of Probabilities from Sparse Data for the Language Model Component of a Speech Recognizer. *IEEE Transactions on Acoustics, Speech and Signal Processing*, 35.

[9] J. D. Kim, S. Z. Lee, and H. C. Rim. HMM Specialization with Selective Lexicalization. In *Proceedings of the join SIGDAT Conference on Empirical Methods in Natural Language Processing and Very Large Corpora(EMNLP-VLC-99)*, 1999.

[10] S. Lee, J. Tsujii, and H. Rim. Lexicalized Hidden Markov Models for Part-of-Speech Tagging. In *Proceedings of 18th International Conference on Computational Linguistics*, Saarbrucken, Germany, August 2000.

[11] L. Màrquez. *Part–of–Speech Tagging: A Machine–Learning Approach based on Decision Trees.* PhD. Thesis, Dep. Llenguatges i Sistemes Informàtics. Universitat Politècnica de Catalunya, 1999.

[12] A. Molina and F. Pla. Clause detection using HMM. In *Proceedings of the 5th Conference on Computational Natural Language Learning*, Toulouse, France, July 2001.

[13] L. Padró. *A Hybrid Environment for Syntax–Semantic Tagging.* PhD. Thesis, Dep. Llenguatges i Sistemes Informàtics. Universitat Politècnica de Catalunya, February 1998. http://www.lsi.upc.es/~padro.

[14] F. Pla. *Etiquetado Léxico y Análisis Sintáctico Superficial basado en Modelos Estadísticos.* PhD. Thesis, Dep. de Sistemes Informàtics i Computació. Universitat Politècnica de València, 2000.

[15] F. Pla, A. Molina, and N. Prieto. Tagging and Chunking with Bigrams. In *Proceedings of the COLING–2000*, Saarbrücken, Germany, August 2000.

[16] F. Pla, A. Molina, and N. Prieto. Improving Chunking by means of Lexical-Contextual Information in Statistical Language Models. In *Proceedings of the 4th Conference on Computational Natural Language Learning and the 2nd Learning Language in Logic Workshop*, Lisbon, Portugal, September 2000.

[17] A. Ratnaparkhi. A Maximum Entropy Part–of–speech Tagger. In *Proceedings of the 1st Conference on Empirical Methods in Natural Language Processing, EMNLP*, 1996.

[18] E. Segarra. *Una aproximación inductiva a la comprensión del discurso continuo.* PhD. Thesis, Departamento de Sistemas Informáticos y Computación, Universidad Politécnica de Valencia, Junio 1993.