

# PREDICCIÓN DE PALABRAS EN CASTELLANO

Sira E. PALAZUELOS CAGIGAS (2), José Lázaro RODRIGO MATEOS (1,3),

Juan I. GODINO (1) Francisco ALIAGA (3), José L. MARTÍN SÁNCHEZ Santiago AGUILERA (1)

{rodrigo, godino, aguilera}@die.upm.es, {rodrigo, aliaga}@eucmax.sim.ucm.es,  
{sira, jlmartin}@depeca.alcala.es

(1) Laboratorio de Tecnología de Rehabilitación  
ETSI de Telecomunicación. Ciudad Universitaria s/n 28040 MADRID

(2) Dpto. de Electrónica. Escuela Politécnica, Univ. De Alcalá.  
Campus Universitario. Crta. Madrid-Barcelona, Km 33,600 28871 Alcalá de Henares (Madrid)

(3) Dpto. de Filología Española-I. Univ. Complutense . Ciudad Universitaria s/n 28040 MADRID

## Resumen

La predicción de palabras consiste en intentar adivinar la o las palabras que un usuario está escribiendo, basándose en las letras que están ya escritas, y en una información adicional proveniente del texto creado previamente. Las palabras predichas se muestran al usuario para que seleccione entre ellas, permitiéndole así introducir de forma rápida la palabra deseada con muy poco esfuerzo. Con el sistema de predicción de palabras se pretende ayudar a usuarios con problemas fundamentalmente físicos, (por ejemplo discapacidad motriz, esclerosis lateral amiotrófica, parálisis cerebral, parapléjicos, accidentados de coche, etc) pero también lingüísticos, a la escritura de textos de toda índole. Por ello la predicción se incluye en distintos tipos de programas: editores de texto, comunicadores, programas para la enseñanza de español, etc.

Describiremos aquí los diferentes métodos de predicción de palabras en castellano que se están utilizando en una aplicación para discapacitados motrices, empleando un modelo híbrido que conjuga métodos estadísticos con reglas declarativas y consiguiendo incrementar la velocidad en la escritura desde 2-3 palabras por minuto a 8-10.

## Agradecimientos

Este trabajo se enmarca en el proyecto CAM 07T / 0010 / 1997

## Abstract

The word prediction consists in trying to guess the word or words which a user is writing, based on added information coming from the previously written text. The predicted words are shown to the user in order to select a word, allowing the user to write the wanted word faster and with very little effort.

The word prediction system intends to help users with fundamentally physical problems, as well as linguistic problems, to write all kind of texts, from a simple sentence in order to communicate, up to writing an important text. Therefore the prediction is included in different kinds of programs: text editors, communicators, programs for the teaching of Spanish as second language etc.

Immediately afterwards we will describe the different methods of word prediction in Spanish which are being used in an application for users with motorical disabilities increasing the typing rate from 2-3 words up to 8-10 words per minute.

## Acknowledges

This work is part of the project CAM 07T / 0010 / 1997.

## 1 *Introducción*

La predicción de palabras es uno de los métodos más comúnmente utilizados para ayudar, tanto en la comunicación personal como en la escritura, a personas con discapacidades motrices. Las ventajas que se obtienen de su uso dependen del grado de discapacidad, de su capacidad lingüística, etc., y fundamentalmente se mide en la aceleración en la escritura, y en la disminución del esfuerzo necesario para escribir el texto. Las personas con problemas lingüísticos también pueden beneficiarse del uso de la predicción, ya que pueden irse guiando con las palabras predichas en su proceso de escritura, obteniendo un apoyo gramatical. A las personas con dislexia se les facilita la escritura, ya que serán capaces de reconocer las palabras correctamente escritas cuando aparezcan en la pantalla, incrementándose así la calidad de producción de los textos escritos.

A continuación describiremos en primer lugar, dos de las aplicaciones en las cuales está incluida la predicción de palabras. En las secciones inmediatamente siguientes explicaremos los métodos de predicción utilizados, y por último, presentaremos las conclusiones y perspectivas de nuestro trabajo.

## 2 *Aplicaciones*

### 2.1 Editor predictivo para personas con discapacidades motrices

Es un editor de texto/comunicador que puede ser utilizado con un solo pulsador. Como intenta ilustrar a continuación la figura 1, todas las opciones de los menús son barridas secuencialmente de modo que el usuario deberá presionar el pulsador cuando el barrido pase por la opción deseada.

(Figura 1)



De esta forma el usuario puede escribir un texto, imprimirlo, e incluso transmitirlo de forma oral mediante síntesis de voz. La predicción de palabras permite escribir palabras completas, con el mismo esfuerzo que conllevaría escribir una letra, hecho este especialmente interesante para personas con problemas de movilidad dado que posibilita la aceleración de la escritura.

El contacto con los usuarios nos ha permitido a su vez comprobar que la predicción de palabras les ayudaba en muchos más aspectos que el de la aceleración de la escritura, como por ejemplo:

- Ayuda a los niños disléxicos, ya que al presentarles de antemano la palabra correctamente escrita, la reconocen, con lo cual la eligen y queda escrita de forma correcta en el texto.
- Apoyo en la lectura del texto, ya que el sistema puede hablar.
- Mejora en el nivel de escritura por la disminución de las faltas de ortografía.
- Aumento de la riqueza del vocabulario, ya que muestra palabras y el niño las va reconociendo.
- Ayuda en la construcción de la frase, desde que se introdujo la información gramatical en la predicción.

### 2.2 Sistema de enseñanza del español como segunda lengua

El hecho de presentar ante el usuario la palabra que desea escribir, abre la puerta de la aplicación del motor predictivo a la enseñanza de lenguas como un pre-corrector ortográfico. El sistema puede ir pre-corrigiendo las opciones del alumno según va escribiendo y así por ejemplo simplemente resolverle la duda de si una palabra se escribe con g o j, b o v, s o z, etc.

Este módulo de español como lengua extranjera (MELE) puede ser útil tanto para personas discapacitadas como para el público en general.

El motor predictivo se utiliza en varias líneas de trabajo a la hora de desarrollar plantillas de ejercicios. En ejercicios de redacción, o creación de texto escrito, cuando un alumno se enfrenta a la escritura de una palabra con un problema determinado, fundamentalmente ortográfico. Al

seleccionar la palabra propuesta, se presenta al usuario la opción de tener una explicación de la regla ortográfica con sus correspondientes ejercicios. Así podemos tratar, por ejemplo, las formas flexivas de verbos irregulares.

En ejercicios de creación de texto, guiando al alumno con un determinado conjunto de palabras o estructuras, se le presenta un seguimiento de las reglas que la escritura del enunciado va satisfaciendo y las categorías que se esperan en un determinado momento, es decir, se le presenta un análisis. Así podemos tratar, por ejemplo, ciertos casos de los usos de subjuntivo, preposiciones etc.

MELE no está concebido como un método más para aprender español, multimedia o interactivo, sino que a partir de la experiencia multidisciplinar de nuestro equipo, tanto en ingeniería como en lingüística y enseñanza de ELE, preparamos una herramienta abierta a las necesidades concretas de cada profesor en cada situación particular, de tal manera que el profesor puede preparar sus propias clases interactivas con el alumnado, sin por ello tener que saber informática.

MELE consta de una serie de temas, ordenados por niveles, en los que se encuentra una explicación teórica seguida de una serie de ejercicios predeterminados, tal y como los puede utilizar un alumno.

### 3 Predicción de palabras

#### 3.1 Métodos estadísticos de predicción de palabras

El sistema de predicción de palabras que hay implementado actualmente incluye varios métodos de predicción independientes, cada uno de los cuales está basado en distintas fuentes de información, y que proporciona distintas ventajas a los usuarios.

- **Diccionario general:** es un diccionario de 40000 palabras, ordenado por frecuencias, que incluye información gramatical. Este diccionario no puede ser personalizado, es decir, no puede ser modificado por el usuario.

Hay matrices de bigramas y trigramas de categorías, que para cada par de categorías nos presenta las categorías que pueden seguirlas, y con qué probabilidad.

Así el sistema nos mostrará las palabras más probables que pertenezcan a las categorías que pueden seguir a las categorías de las palabras anteriores. Tendrá en cuenta la probabilidad de la palabra, y la probabilidad que tiene la categoría de seguir a las anteriores. Si se queda sin palabras gramaticalmente correctas, mostrará las más frecuentes, sin tener en cuenta su categoría.

- **Diccionarios temáticos:** permiten personalizar los diccionarios. Procesan un determinado texto, obteniendo las frecuencias de todas las palabras, y los pares de palabras consecutivas que aparecen en él, (incluso tiene en cuenta tríos de palabras para considerar, por ejemplo, el caso de algunas preposiciones). Categoriza todas las palabras que conoce el diccionario general.

Una vez que se ha creado un diccionario temático, el usuario lo cargará en el sistema cada vez que quiera producir texto de ese área temática. De esta forma, se contempla desde el primer momento vocabulario específico del área temática. Al escribir cada palabra se mostrarán las que el usuario había escrito tras ella cuando había producido textos dentro de esa área temática. En el caso de las preposiciones se considerará igualmente el uso dado en los textos creados en ese área la palabra anterior.

En caso de no encontrar más palabras, el sistema predecirá utilizando el diccionario general, teniendo en cuenta la información gramatical que tenga.

- **Diccionario del texto en curso** es un diccionario que se va generando a medida que se va escribiendo. Su utilidad radica en lograr que las palabras nuevas tengan que escribirse solamente una vez, y que las palabras más recientemente escritas sean mostradas antes que otras de mayor frecuencia en el diccionario general. El diccionario del texto en curso también guarda información de los pares de palabras escritos.
- **Filtro de las palabras rechazadas:** Si una palabra aparece varias veces en el menú y no es seleccionada por el usuario, asumiremos que no es la deseada. La palabra entonces ya no será mostrada más veces aunque al escribir nuevas letras siga siendo correcta.

Estos tres diccionarios son utilizados conjuntamente, de forma que las palabras que se predicen incluyen información probabilística, gramatical, "temporal", y de "estilo del usuario":

- Probabilística: información de la frecuencia de la palabra (considerando tanto la frecuencia en el diccionario general como la particular del usuario).
- Gramatical: las palabras predichas son coherentes gramaticalmente con la (o las) anteriores.
- "Temporal": ya que prevalecen las últimas palabras que el usuario ha escrito anteriormente (el uso de una palabra hace que sea más probable volver a utilizarla en el mismo texto). Tienen así prioridad en la predicción las más recientemente utilizadas.
- Estilo del usuario: como consecuencia de dar prioridad a las palabras usadas más recientemente, el sistema predictor va "adaptándose" al estilo del usuario.

Si incluimos más información a la predicción, tendremos más posibilidades de obtener las palabras más adecuadas, y la palabra deseada aparecerá antes en el menú.

### 3.2 Chart parser extendido

Hasta ahora, los métodos de predicción utilizados están descritos en los artículos publicados, en particular en "[Pala et alii 96,97 y 98]". En estos métodos, la información gramatical que utilizamos para predecir es simplemente la de las dos palabras anteriores, y se pierde información contextual.

Buscando tener en cuenta toda la información gramatical de la frase en curso, se decide utilizar un parser que analice el texto escrito hasta el momento de la predicción, y según este análisis, prediga las categorías posibles gramaticalmente correctas en el punto de la frase que nos encontremos.

El analizador elegido es un chart parser, que, al funcionar de izquierda a derecha, de forma natural analiza lo que se lleva escrito hasta el momento, y nos da información de la situación en la que se ha quedado y cual serían las categorías posibles de la palabra siguiente para que fueran correctas (es decir, que satisfagan las reglas). Una vez tengamos las categorías,

buscaremos en los diccionarios las palabras etiquetadas con tales categorías.

Este sistema necesita "ser alimentado" con reglas gramaticales con información probabilística, que luego utilizará para dar preferencia a unas reglas sobre otras, a la hora de predecir. Representaremos las reglas válidas como ramas de un árbol, y a medida que avanza el análisis de la frase en curso averiguaremos cuáles son las ramas válidas del árbol, y asignamos a cada una la probabilidad correspondiente.

En esta probabilidad entran en juego varias probabilidades, que se averiguan partiendo de las siguientes frecuencias:

- Frecuencia absoluta de la palabra W: número de veces que aparece dicha palabra en el corpus, independientemente de la categoría a la que pertenezca.

$$\text{Frec}(W) = \# W = \text{total\_frec\_pal}$$

- Frecuencia absoluta del par palabra-categoría: número de veces que cada palabra W apareció en el corpus categorizado perteneciendo a cada categoría gramatical  $C_i$ .

$$\text{Frec}(W \cap C_i) = \#(W \cap C_i)$$

$$= \text{frec\_pal\_y\_categ}[i]$$

- Probabilidad de que una palabra W pertenezca a la categoría  $C_i$ :

$$\text{Pr}(C_i/W) = \frac{\text{Frec}(W \cap C_i)}{\text{Frec}(W)} =$$

$$\frac{\text{frec\_pal\_y\_categ}[i]}{\text{total\_frec\_pal}}$$

- Frecuencia absoluta de cada categoría gramatical  $C_i$ : Número de veces que aparece cada categoría en el corpus. Suma de las frecuencias absolutas de todas las palabras pertenecientes a dicha categoría. (Acumulado en un campo de la estructura Lista\_Categorías llamado FrecCateg)

$$\text{Frec}(C_i) = \# C_i =$$

$$\text{Lista\_Categorías}[C_i] \text{FrecCateg}$$

- "Lexical probability": probabilidad de una determinada palabra W fijada una categoría  $C_i$ :

$$\Pr(W/C_i) = \frac{\text{Frec}(W \cap C_i)}{\text{Frec}(C_i)}$$

$$\frac{\text{frec\_pal\_y\_categ}[i]}{\text{Lista\_Categorias}[C_i].\text{Frec\_Categ}}$$

- Probabilidad de cada regla: probabilidad de que cada regla derive su símbolo no terminal correspondiente. Si la regla  $R_i$  deriva el símbolo no terminal  $S_k$ , la probabilidad de dicha regla será:

$$\Pr(R_i) = \frac{\#R_i}{\#S_k} = \text{yprprob}[R_i]$$

Basándonos en estos datos, pretendemos predecir la categoría de la siguiente/actual palabra que el usuario escriba basándonos en la secuencia de palabras anteriores del texto.

Si cada palabra tuviera solamente una categoría posible, las ramas activas del analizador serían aquellas cuya secuencia de categorías hasta el momento coincida con la secuencia de categorías de las palabras escritas por el usuario. Esa secuencia sería única, y la probabilidad de cada rama sería el producto de las probabilidades de las reglas implicadas en su generación. Según esto, la probabilidad de que la siguiente palabra pertenezca a una determinada categoría será la suma de las probabilidades de todas las ramas que den lugar a dicha categoría. Por ello, cada regla lleva asociada información del producto de las probabilidades de las reglas que han dado lugar a ella, en el campo `prob_regla`.

$$\Pr_i(C_i) = \sum_{\text{Ramas que esperan } C_i} \left( \prod_{\text{Reglas implicadas en esa rama}} \Pr(R_i) \right)$$

$$= \sum_{\text{Ramas que esperan } C_i} (R_i.\text{prob\_regla})$$

Como las palabras pueden tener varios significados y pertenecer a distintas categorías gramaticales, la probabilidad de seguir por cada rama debe ser ponderada por la probabilidad de que cada palabra pertenezca a la categoría necesaria para seguir por esa rama.

$$\Pr(C_i) = \sum_{\text{Ramas que esperan } C_i} \left( \prod_{\text{Reglas implicadas en esa rama}} \Pr(R_i) \prod_{\text{Categorías implicadas en la regla } R_i} \Pr(C_i / W_j) \right)$$

Al igual que antes, en cada momento podemos disponer de la información del producto de las probabilidades de que cada palabra pertenezca a cada categoría, a lo largo de todas las ramas del árbol que han dado lugar al punto donde nos encontremos. Esto está almacenado en el campo `prodlexicprob`, por lo cual, la probabilidad de que con esa cadena de palabras estemos en un punto determinado del árbol será:

$$\text{Prob}(\text{punto del árbol}) = \text{prob\_regla} * \text{prodlexicprob}$$

y la probabilidad de que con dicho árbol se prediga la categoría  $C_i$  se obtendrá con una expresión similar a la anterior:

$$\Pr(C_i) = \sum_{\text{Ramas que esperan } C_i} R_i.\text{prob\_regla} * R_i.\text{prodlexicprob}$$

Una vez tengamos la probabilidad de que cada categoría sea la siguiente, predicaremos las palabras, dependiendo de su categoría:

$$\begin{aligned} \Pr(W) &= \sum_{\text{Para todas las categorías a las que pertenece } C_i} (\Pr(W/C_i) * \Pr(C_i)) \\ &= \sum_{\text{Para todas las categorías a las que pertenece } C_i} \frac{\text{frec\_pal\_y\_categ}[W \cap C_i]}{\text{Lista\_Categorias}[C_i].\text{Frec\_Categ}} \\ &\quad * \sum_{\text{Ramas que esperan } C_i} R_i.\text{prob\_regla} * R_i.\text{prodlexicprob} \end{aligned}$$

Como hay una parte de dicho sumatorio independiente de la palabra, podemos acumularlo, haciendo una variable nueva, que almacene ese valor para cada categoría:

$$\begin{aligned} \_ \text{frec\_cat\_egs\_prefs\_bison}[C_i] &= \\ &\sum_{\text{Ramas que esperan } C_i} \frac{R_i.\text{prob\_regla} * R_i.\text{prodlexicprob}}{\text{Lista\_Categorias}[C_i].\text{Frec\_Categ}} \end{aligned}$$

Con lo cual, la probabilidad de que cada palabra sea la siguiente es:

$$\begin{aligned} \Pr(W) &= \sum_{\text{Para todas las categorías a las que pertenece } w} \text{frec\_pal\_y\_categ}[W \cap C_i] \\ &\quad * \_ \text{frec\_cat\_egs\_prefs\_bison}[C_i] \end{aligned}$$

### 3.2.2 Información gramatical

Como se ha dicho anteriormente (cfr. Diccionario general) el sistema consta de un diccionario de unas 40000 palabras en su versión

actual, cuyas entradas estaban etiquetadas utilizando las siguientes categorías:

CALIF	POSES	INDEF
DEMOS	INTERR	PRON
CARDIN	ORDIN	PREP
ARTDET	ARTIND	SUST
ADJET	ADVERB	CONJ
DETER	VERBO	SIGNO

Estas categorías posibilitan la creación de bigramas y trigramas de categorías, que permiten al sistema dar unos resultados aceptables según dos criterios:

- Pulsaciones: el porcentaje de pulsaciones necesario para escribir un texto tecleando todas las letras en relación con las realmente efectuadas por el usuario utilizando la predicción de palabras que es de un 40 %.
- Elección/éxito en la predicción. El 84% de las palabras del texto son predichas correctamente. XXX

Así, por ejemplo, predeciría sucesivamente las palabras componentes de los enunciados siguientes:

(1) *la casa de los espíritus*

(2) *sobre libre para sellar*

Sin embargo un análisis detallado de las predicciones nos muestra que en (1) el sistema ha utilizado un léxico categorizado como sigue

la	ARTDET(artículo determinado)
casa	SUST (sustantivo)
de	PREP (preposición)
los	ARTDET
espíritus	SUST

y en (2)

sobre	VERBO
libre	VERBO
para	VERBO
sellar	VERBO

Es decir, el sistema predecía con éxito, dado que existen bigramas que predicen a partir de un pronombre un verbo, a partir de un verbo otro verbo y a partir de un verbo un adjetivo. Los elementos léxicos cumplen con las exigencias de poseer las categorías pronombre, verbo y adjetivo, aun cuando, la información morfológica de cada elemento léxico en la

secuencia (2) no corresponde con la que se esperaría de un análisis morfológico desde un punto de vista lingüístico.

En el ejemplo siguiente (3) tras la escritura de la palabra *he*, el sistema podría predecir *manzana* dado existe un bigrama que predice un nombre tras un verbo

(3) *he manzana*

cuando en realidad el usuario esperaba escribir el enunciado siguiente

(4) *he mandado*

Es decir, en los casos de los dos tipos (2)) y (3) hay una desviación, que si bien en el primer caso (2) no afecta a una predicción satisfactoria, en el segundo (3) presenta en principio ciertos inconvenientes.

Una rápida observación de la gramática del castellano nos indica que entre un verbo auxiliar como *haber* en forma personal y un participio no ser puede introducir nada. Esta información debe introducirse en el sistema.

Al lado de estos casos siempre tendremos además problemas de ambigüedad del lenguaje como en (1) dado que las palabras componentes del ejemplo antes citado admiten diferentes etiquetas morfológicas:

la	(artículo det)	(pronombre)
casa	(sustantivo)	(verbo)
de	(preposición)	(verbo)
los	(artículo)	(pronombre)

o bien problemas de tipo semántico como los que ilustran los ejemplos siguientes

(4) *has manzanas*

(5) *come la revista*

En (4) en una primera lectura se puede pensar que es agramatical, pero sería necesario precisar si se debe admitir o no en castellano ese valor de *haber* con el sentido de *tener*.

En (5) sería necesario disponer de una información semántica (o sintáctico-semántica) en la entrada léxica del diccionario.

Frente a estas observaciones, se impone la adopción de un sistema híbrido (cfr. chart parser extendido) que conjugue la información estadística, la introducción de rasgos y un

conjunto de reglas declarativas con los siguientes criterios:

- a) No consideración del aspecto semántico en un primer momento
- b) Adaptación de las categorías usadas inicialmente en el diccionario a nuestras necesidades
- c) Adaptación de las entradas léxicas según las nuevas categorías
- d) Definición de un conjunto de reglas declarativas

a) No consideración del aspecto semántico.

La importancia de la información de tipo semántico e incluso pragmático parece clara en los métodos de procesamiento del lenguaje natural. Sin embargo, en nuestra aplicación no es una prioridad dado que el usuario tipo (persona con discapacidad motriz) posee una competencia de su lengua y está escribiendo en tiempo real, seleccionando sin dificultad entre las palabras propuestas.

b) Adaptación de categorías.

Partiendo de la base que no existe un buen o mal etiquetado, sino en función de los objetivos previstos, nos planteamos la adaptación de las categorías inicialmente existentes según un criterio claro: mientras que no se manifieste que una palabra o conjunto de palabras tienen un comportamiento sintáctico diferente, en una lectura de izquierda a derecha, que el resto de las palabras del mismo paradigma, no se justifica una nueva categoría (por muy tradicional que sea ésta); y al contrario, cuando se manifieste un comportamiento diferente, está justificada la adopción de una nueva categoría.

Así por ejemplo, existe un conjunto reducido de palabras que se encuentran en varias categorías tradicionales en lingüística (Determinantes, artículos, pronombres, etc.). Tras una observación detenida del comportamiento sintáctico del léxico incluido en esas categorías, adoptamos una categorización de acuerdo con los criterios expuestos que

justifica la siguiente presentación:

ya que el comportamiento de un gran conjunto del léxico incluido en las anteriores categorías de artículos determinados, indeterminados, posesivos, indeterminados o demostrativos es el mismo:

*los dos amigos*

*mis dos amigos*

Sin embargo, hay un comportamiento particular de elementos léxicos debido a la concordancia. El género y el número se tratan no por reglas categoriales sino mediante rasgos. Así por ejemplo para predecir *niño* en el ejemplo siguiente

*El niño*

no sólo se utiliza una regla cuya información descriptiva coincidiría con un bigrama, sino que además hay unos rasgos de *el* que permiten predecir *niño*.

c) Adaptación de las entradas léxicas según las nuevas categorías

En el momento actual nos encontramos en una fase de recategorización del diccionario.

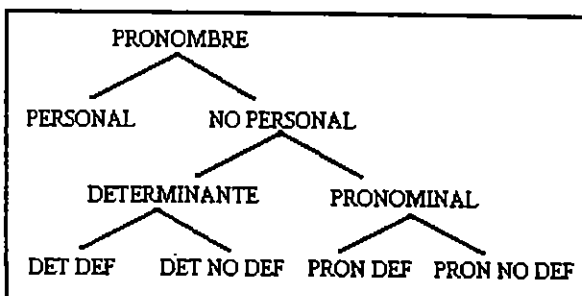
d) Definición de un conjunto de reglas en función de las nuevas categorías.

Las reglas declarativas se están definiendo de modo progresivo, dando cuenta del mayor número de observaciones lingüísticas posibles que sean relevantes sintácticamente, e introduciendo asimismo información de rasgos.

En el ejemplo anterior mostrábamos cómo se utilizan rasgos para predecir *niño* a partir de *el*. La información transmitida mediante rasgos se utiliza también para la declaración de reglas. Así por ejemplo no solo se utilizará una información categorial que nos permita saber que a un artículo como *el* le puede seguir un verbo como comer, sino que utilizaremos rasgos para dar cuenta de la observación lingüística siguiente. A un artículo masculino singular le puede seguir un verbo en infinitivo, pero no es posible ni con un artículo femenino, ni con un verbo en forma flexiva.

El (detdef) masc Verbo(Infinitivo)

El tratamiento del verbo lo hacemos utilizando fundamentalmente rasgos, lo que nos permite tratar los verbos compuestos:



Vaux Vpart (masc)

Las reglas declarativas se van definiendo de modo progresivo siguiendo las observaciones del léxico más frecuente.

#### 4 Conclusiones y perspectivas.

En el presente artículo mostramos el estado de trabajo de un sistema de predicción de palabras en el que se adopta un modelo híbrido que conjuga métodos estadísticos con reglas declarativas.

Como perspectivas inmediatas se plantean el incremento del número de reglas, acabar de recategorizar el diccionario, aumentar la potencia del analizador y realizar la evaluación con textos de gran tamaño.

Se plantea igualmente un sistema de evaluación doble, una evaluación automática a partir de textos y otra, realizada por usuarios discapacitados.

#### 5 Referencias

[1999] Palazuelos-Cagigas Sira E. Aguilera-Navarro Santiago Rodrigo Mateos José L. Godino-Llorente. Juan I Martín-Sánchez José L. *Considerations on the Automatic Evaluation of word prediction systems*. En preparación

[1998] Palazuelos, S., Aguilera S., Rodrigo J.L., Godino J. "Grammatical and statistical word prediction system for Spanish integrated in an aid for people with disabilities": Poster Congreso: ICSLP, Sydney, 30 Nov - 4 Dic 1998

[1997] Palazuelos Cagigas, S. E., Godino Llorente, J.I., Aguilera-Navarro, S.. Comparison Between Adaptive and Non-Adaptive Word Prediction Methods in a Word Processor for Motorically Handicapped Non Vocal Users. Proceedings de AAATE Conference 1997 (pp. 158-162). Thesalonica, Greece.

[1997] Carlberger A., Carlberger J., Magnuson T., Palazuelos-Cagigas S.E., Hunnicutt M.S. & Aguilera, S, *Profet, a New Generation of Word Prediction: An Evaluation Study*. Proceedings of the Workshop on NLP for Communication Aids, ACL/EACL'97 (pp. 23-28). Ed.

"Association for Computational Linguistics", Madrid

[1996] Palazuelos Cagigas, S. E., Aguilera Navarro, S: Report on Word Prediction for Spanish. Deliverable WP7T3D.2IR of "VAESS: Voices, Attitudes and Emotions in Speech Synthesis". TIDE N. 1174 August 1996.

[1995] Allen, J, "Natural Language Understanding", Second Ed. Benjamin / Cummings, CA.