

LA LEY DE ZIPF EN EL CASTELLANO Y HERRAMIENTAS PARA SU COMPUTACIÓN

García Olmedo, F.; García Sánchez, P.A.; Martínez Sevilla, A.

Dpto. de Álgebra. Universidad de Granada.

La presente demostración tiene por objeto presentar una herramienta informática para la obtención de datos reales sobre un idioma dado y su posterior comparación con los datos teóricos que se derivan de la ley de Zipf, una vez calculada la constante que lleva su nombre, así como otras leyes.

Aunque la herramienta es válida para cualquier lengua, los ejemplos se han elaborado a partir de textos en Inglés y Castellano. Se analizan diversos ficheros obteniendo para cada uno una tabla de frecuencias de aparición de palabras en el texto introducido, una ordenación decreciente en probabilidad por frecuencia relativa de las mismas (listada en una tabla separada) y la consiguiente nube de puntos. A dicha nube de puntos se le ajusta una recta mínimo-cuadrática y la recta proporcionada por la ley de Zipf. Las consiguientes representaciones se hacen en escala logarítmica, en la cual curiosamente la recta de mínimos cuadrados se aproxima a la de Zipf, poniéndose de manifiesto cierta coherencia teórica entre ambas construcciones. Se añaden tablas que ilustran los resultados aportados por el índice de Waring-Herdan.

Los programas fueron implementados en C. El almacenamiento y posterior ordenación de los datos por frecuencia se hace en disco mediante el uso de una tabla Hash.

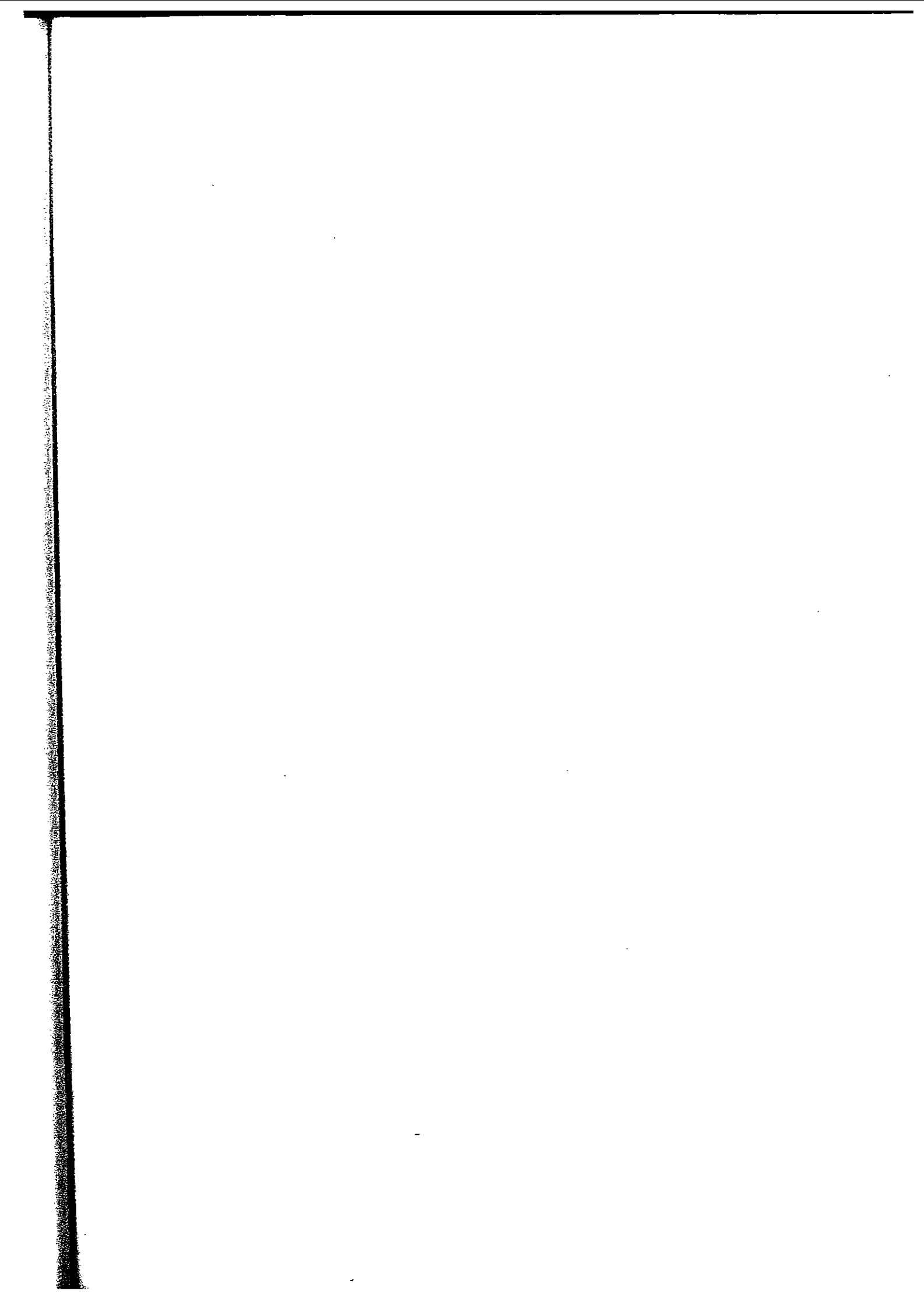
La presente herramienta pretende ser la primera de un conjunto más amplio que permita más detalle en el análisis de un lenguaje así como en la comparación de los datos empíricos con los teóricos obtenidos a la luz de los distintos modelos propuestos. En el futuro podrán ser implementadas las correcciones a la ley de Zipf propuestas por Mandelbrot mediante los parámetros V y D (dimensión fractal), ley de Zipf para los resultados del conteo de raíces y lexemas, etc.

Requerimientos Hardware

1. Ordenador personal compatible IBM (se aconseja el uso de un ordenador con procesador i386 ó i486).
2. Disco duro.
3. VGA color.

Bibliografía básica

- [1] Müller. Estadística Lingüística. Gredos 201. 1965.
- [2] Welsh, D. Codes and Cryptography. Clarendon Press, 1988.
- [3] Yavuz, D. Zipf's Law and Entropy. IEEE Transactions for Information Theory, 20,p. 650. 1974.



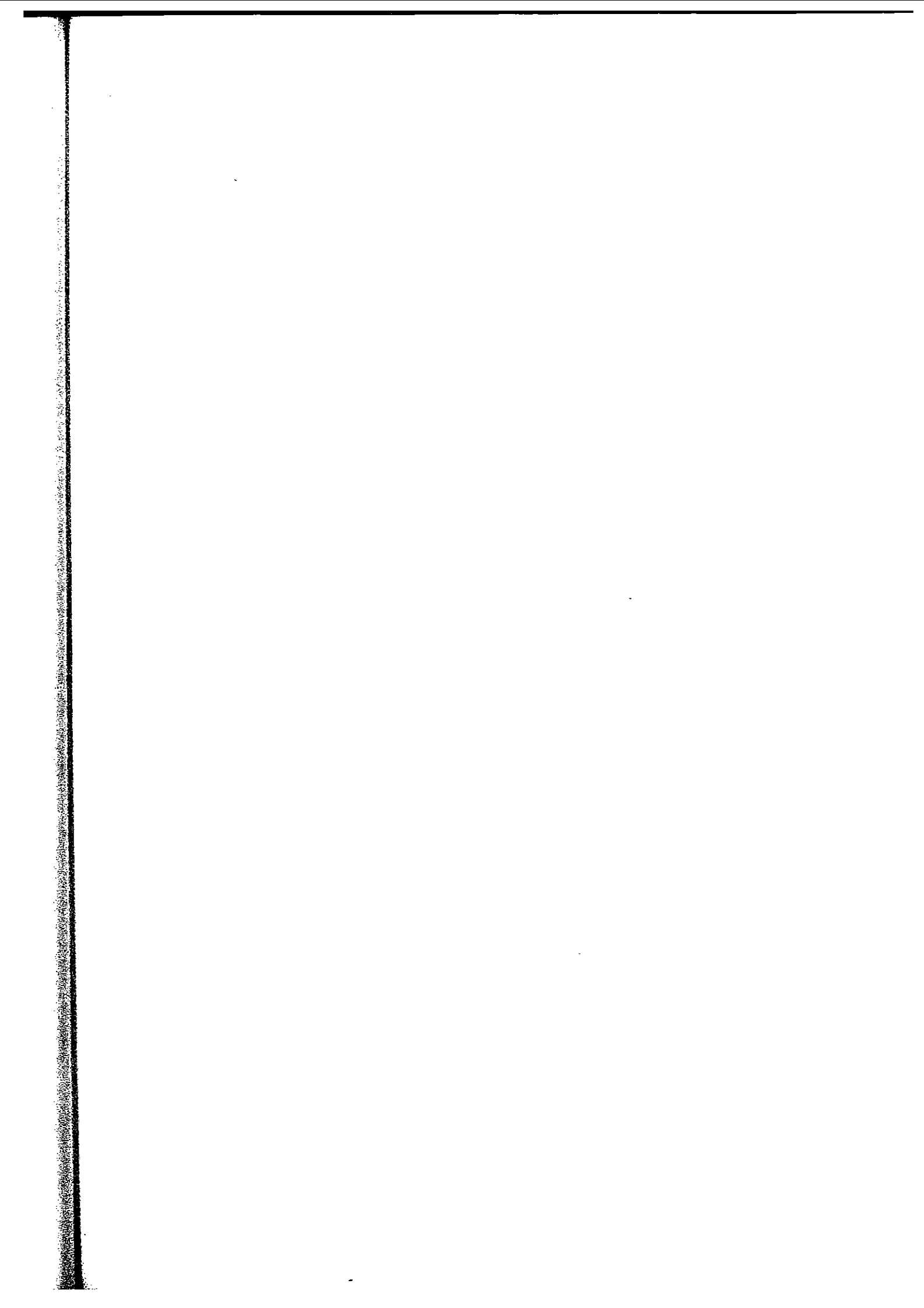
DEMOSTRACIÓN DEL CONVERTOR TEXTO-VOZ AMIGO Y DEL MÓDULO DE ANÁLISIS SINTÁCTICO

*Miguel Ángel Rodríguez Crespo, J. Gregorio Escalada,
A. Macarron Larumbe y L. Monzon Serrano*

Telefónica, I+D
Emilio Vargas nº6
28043 MADRID

RESUMEN

Se realizará una demostración del funcionamiento del convertor texto-voz AMIGO funcionando en tiempo real sobre una estación de trabajo.



CONJUGACIONES VERBALES

Santana O., Hernández Z. J., Rodríguez G.

Edificio Departamental de Informática y Matemáticas
Campus Universitario de Tafira
Las Palmas de Gran Canaria, 35017, Apdo. 322

Resumen

El presente trabajo se enmarca dentro de un proyecto mayor, concebido como un sistema de ayuda en la elaboración de documentos escritos. En concreto, se desarrolló una herramienta capaz de resolver cómo se conjuga un tiempo determinado de cualquier verbo. Se ha desechado la solución trivial, que sería una voluminosa base de información en la que se encontrarían la totalidad de los verbos conjugados.

La conjugación de los verbos regulares no plantea ningún tipo de problema, puesto que es absolutamente sistemática. En consecuencia, el quid de la cuestión lo constituyen los verbos irregulares. La solución adoptada ha consistido en la elaboración de un sistema que conjuga cualquier tiempo verbal a partir de un infinitivo —previa determinación de su regularidad o irregularidad—, utilizando para ello el conocimiento de una serie de reglas que siguen los diferentes verbos.

Requerimientos software y hardware:

Sistema operativo:	MS-DOS 5.0
Memoria en disco:	3 Mbytes
Memoria Ram:	3 Mbytes XMS

Conjugaciones Verbales

Introducción

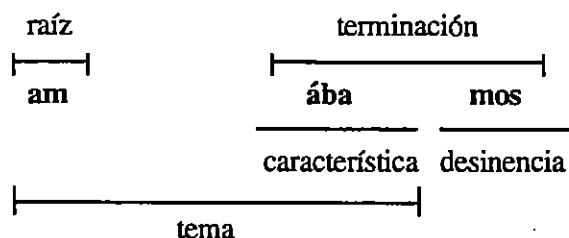
El presente trabajo se enmarca dentro de un proyecto mayor, concebido como un sistema de ayuda en la elaboración de documentos escritos. En concreto, se desarrolló una herramienta capaz de resolver cómo se conjuga un tiempo determinado de cualquier verbo. Se ha desechado la solución trivial, que sería una voluminosa base de información en la que se encontrarían la totalidad de los verbos conjugados.

La conjugación de los verbos regulares no plantea ningún tipo de problema, puesto que es absolutamente sistemática. En consecuencia, el quid de la cuestión lo constituyen los verbos irregulares. La solución adoptada ha consistido en la elaboración de un sistema que conjuga cualquier tiempo verbal a partir de un infinitivo —previa determinación de su regularidad o irregularidad—, utilizando para ello el conocimiento de una serie de reglas que siguen los diferentes verbos.

Morfología del verbo

El verbo, por sus características formales, es aquella parte de la oración que contiene *morfemas* flexivos de número, persona, tiempo y modo. Suele aplicarse la denominación de *desinencias* a los morfemas de número y persona, el de *características* a los de modo y tiempo. Si se elimina de una forma verbal su desinencia y característica, lo que queda es la *raíz* o *lexema*. La unión de la raíz con la

característica se denomina *tema*. Cualquier forma verbal se construye básicamente con una raíz y una *terminación* constituida por los morfemas.



Los verbos se encuentran clasificados en tres tipos: Los de la 1ª conjugación -los terminados en ar- los de la segunda conjugación -los terminados en er y los de la tercera conjugación -los terminados en ir- La mayor parte de los verbos son regulares y mantienen su raíz de forma invariable a lo largo de su conjugación —si exceptuamos la posición del acento de intensidad. Existen también en cantidad apreciable verbos que presentan *raíces irregulares* debido a variaciones, ya sean vocálicas y/o consonánticas, además de otros tipos de anomalías; estos son los conocidos verbos irregulares.

LA IRREGULARIDAD VOCÁLICA Consiste en el cambio de una vocal por otras:
e por i. **concebir ⇒ concibo**

LA IRREGULARIDAD CONSONÁNTICA tiene lugar si se reemplaza una consonante por otra, o se añade una consonante a la consonante de la raíz del infinitivo:
c por qu delante de e. **caber ⇒ quepo**

LA IRREGULARIDAD MIXTA aparece cuando se sustituye una vocal y una consonante por otra vocal y otra consonante, o en la adición del grupo ig a la última vocal de la raíz:
ab por ep. **saber ⇒ sepa**

Existen, además, irregularidades que afectan al tema (di, haz, doy, voy, etc.), y otras de más difícil sistematización:

- contracciones.
- verbos con más de una raíz.
- participios y gerundios irregulares.

No se consideran propiamente irregularidades los simples cambios ortográficos:

- g por gu delante de e o i **sigo ⇒ sigue**

En todos estos casos no ha variado el *fonema*, sino el caracter que lo representa. Algunas otras aparentes anomalías obedecen a principios generales del sistema fonológico español y tampoco se consideran irregularidades [leer/leyendo].

Grupos de tiempos irregulares

Siempre que en un verbo aparezca un tiempo irregular, esta irregularidad afecta a otros tiempos. Se reúnen en tres grupos los tiempos irregulares que comparten la misma irregularidad:

- Si es irregular el presente de indicativo, también lo son los otros presentes (presente de subjuntivo e imperativo).
- Si es irregular el pretérito indefinido, poseen la misma irregularidad el pretérito imperfecto de subjuntivo y el futuro imperfecto de subjuntivo.
- Si es irregular el futuro imperfecto de indicativo, tiene la misma irregularidad el condicional simple.

El pretérito imperfecto de indicativo carece de irregularidades, salvo raras excepciones que se reducen casi exclusivamente a los imperfectos heredados del latín *era* e *iba*, de los verbos de *raíz múltiple* *ser* e *ir*.

Respecto a los tiempos compuestos, no presentan más irregularidades que las del auxiliar *haber* y algunos participios.

La Conjugación

El primer paso al afrontar la solución es determinar si el verbo pertenece al conjunto de los verbos regulares o al de los irregulares, ya que ello conlleva una estrategia de tratamiento diferente.

La Conjugación Regular

Cuando el verbo a conjugar es regular, el proceso a seguir para obtener un tiempo determinado es relativamente simple: a) se extrae la raíz del verbo separándola de su terminación, b) a partir de la terminación se determina la conjugación a la que pertenece el verbo (1ª, 2ª, 3ª), y c) en función de la misma y del tiempo deseado, se obtienen las terminaciones que hay que añadir a la raíz para construir el tiempo en todas sus personas.

Los problemas que pueden darse son de dos tipos:

- 1) Los cambios ortográficos que aparecen en la unión de la raíz con la terminación. Se resuelven mediante un conjunto de *Reglas de Unión* —comunes con los verbos irregulares.
- 2) Lo plantean aquellos verbos cuya única irregularidad consiste en la acentuación de algunas personas en determinados tiempos. Esta irregularidad responde a reglas bien definidas.

La Conjugación irregular

Los verbos irregulares lo son porque en determinados tiempos y personas se apartan de la norma general, introducen variaciones en sus raíces y en algunos casos afectan a la terminación. Afortunadamente todos estos cambios presentan un alto grado de sistematización —no carente de excepciones— que permite el desarrollo de una normativa para conjugar casi cualquier verbo irregular —naturalmente, las rarezas necesitan reglas específicas. Lo propuesto se basa en dos pilares fundamentales:

- 1) El establecimiento de las reglas que siguen los verbos en sus variaciones; pueden ser de tres tipos: *Cambio de Raíz*, *Excepción* y *Postexcepción*. Además existe el conjunto de *Reglas de Unión* —comunes con los verbos regulares, y una clasificación de terminaciones verbales que contempla las especificidades propias de los verbos irregulares.
- 2) La determinación de un grupo reducido de verbos —modelos para los demás— que representan todas las posibilidades de combinación de los tres tipos de reglas anteriores.

Reglas de Cambio de Raíz

Son un conjunto de cambios que pueden experimentar las raíces de los verbos irregulares para conjugar los distintos tiempos. Básicamente se corresponden con las irregularidades vocálicas, consonánticas y mixtas. Como norma general, un verbo irregular utiliza hasta tres reglas de cambio de raíz, una por cada grupo de tiempos irregulares (GTI).

Consisten, por tanto, en simples cambios de una secuencia de caracteres de la raíz por otra diferente. Por ejemplo, para el verbo **almorzar** los tiempos del primer GTI se forman cambiando la *o* por *ue* y la raíz que queda es **almuerz**.

Reglas de Excepción

En casi todos los verbos irregulares existen algunos tiempos y personas que no siguen las reglas de cambio de raíz. Por ejemplo, en el verbo **almorzar**, el presente de indicativo se forma con la raíz **almuerz** —salvo las 1ª y 2ª personas del plural que son **almorzamos** y **almorzáis** respectivamente. La excepción descrita se reproduce en los otros dos tiempos del GTI: el presente de subjuntivo y el imperativo —la situación se repite para varios verbos modelo. En otros casos es excepción todo el presente de indicativo menos la 1ª persona del singular (**asir** — **asgo**, **ases**, **ase**, **asimos**,...) y esa situación suele ir acompañada de la excepción en las 2ª persona del singular y 3ª del plural del imperativo (**ase**, **asid**).

Estas posibilidades de normalización han permitido instrumentar un prudente número de reglas que modelizan las excepciones.

Reglas de Unión

Fundamentalmente se trata de reglas ortográficas y anomalías debidas al sistema fonológico español, y son en su mayoría reglas de aplicación general a los verbos, regulares e irregulares.

Las reglas de unión indican qué variaciones se deben realizar al concatenar una raíz verbal —cambiada o no— con la terminación correspondiente. Por ejemplo, si la raíz termina en *ñ* y la terminación comienza por *i* seguida de otra vocal, la *i* desaparece.

ciñ+iendo ⇒ ciñendo.

Reglas de Postexcepción

Entran en juego con posterioridad al resto de las reglas, y afectan tanto a la raíz como a la terminación después de haber realizado la unión; de ahí su denominación.

Las reglas de postexcepción cubren los casos atípicos, entre otros: verbos cuya raíz irregular termina en *ng*, que forman la 2ª persona del imperativo con la raíz regular desprovista de terminación (**venir** ⇒ **ven**, **venga**) o verbos cuyo pretérito imperfecto de indicativo es irregular, caso de **ser** e **ir** —en sí mismos resultan verbos altamente irregulares.

Participios Irregulares

Hay verbos que presentan un participio irregular, tanto porque cambian la raíz, como por no utilizar las terminaciones **ado**, **ido** (**poner** ⇒ **puesto**), e incluso algunos poseen dos participios, uno regular y el otro irregular (**concluir** ⇒ **concluido**, **concluso**).

Las posibilidades de sistematización de los cambios en los participios parecen ser escasas.

Acentuación

Determinados verbos presentan variaciones en su acentuación como única irregularidad. Concretamente, tales cambios consisten en la aparición de tilde ortográfica sobre *u* o *i* en determinados

tiempos y personas (*actuar* fi *actúo*). De hecho siguen una norma en la aparición de la tilde: la ruptura de un diptongo (*enraízo*). Debe extenderse el término de diptongo a aquellos casos en que existe una *h* intercalada en el grupo vocálico (*cabrahígo*). En base a ello, se puede contemplar a estos verbos ordenados por una regla de unión propia. Los verbos que no aplican dicha norma de acentuación se consideran anómalos.

Terminaciones

El estudio de las terminaciones que utilizan los verbos a la hora de formar sus tiempos ha revelado que existen 33 conjuntos de terminaciones, que son utilizadas por los verbos regulares en la conjugación de los distintos tiempos. De los 33, 11 corresponden a cada una de las tres conjugaciones verbales. Cada conjunto se compone de las seis terminaciones de un tiempo, salvo el imperativo—que carece de la primera persona del singular— y el grupo de tiempos impersonales —constituido por el infinitivo, el gerundio y el participio.

Se han determinado además otros conjuntos de terminaciones, hasta totalizar 51, que sólo usan los verbos irregulares. Asignándole a cada conjunto un número que permite identificarlo. Sin embargo, muchos verbos irregulares utilizan el mismo sistema de terminaciones que los regulares.

Realización

La solución del problema planteado se compone de tres módulos independientes: el conjugador propiamente dicho y dos preprocesadores encargados de preparar las estructuras a partir de la información que se proporciona de los verbos. La figura 1 muestra la relación entre los tres módulos.

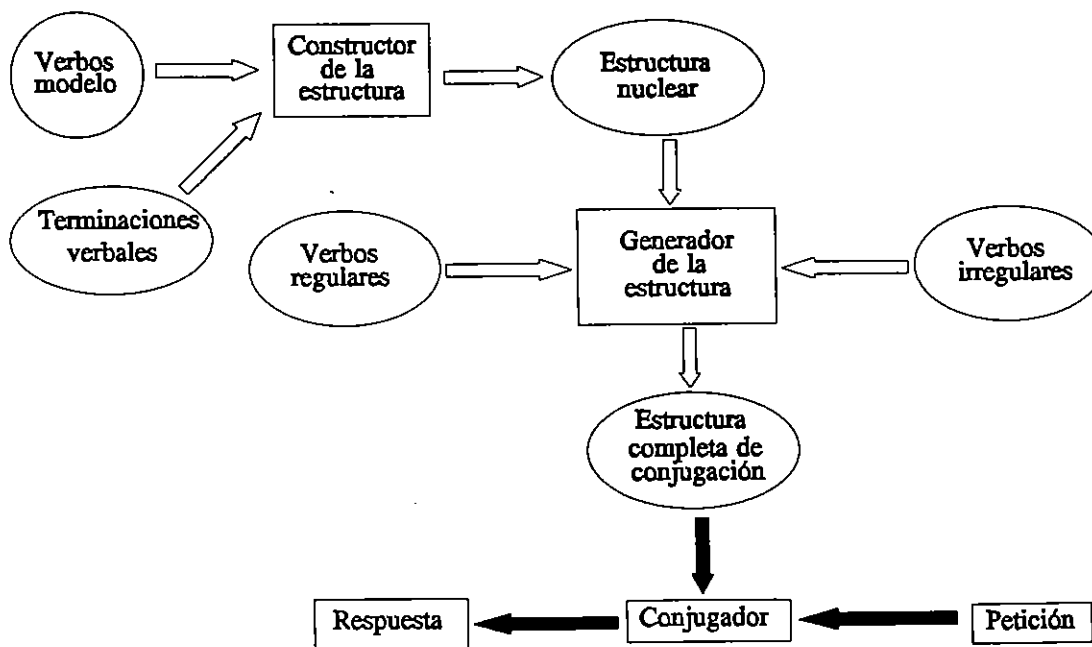


Figura 1

El primero de los preprocesadores se encarga de construir una estructura básica formada por los verbos modelo y las terminaciones verbales. Los primeros se obtienen de un fichero en el que se encuentran todos los modelos con el siguiente formato:

verbo
 código de la regla de cambio de raíz del primer GTI
 código de la regla de cambio de raíz del segundo GTI
 código de la regla de cambio de raíz del tercer GTI
 códigos de la terminación (8 líneas)
 código de la excepción

Las terminaciones se encuentran en otro fichero agrupadas de seis en seis por cada tiempo. Cuando una persona no existe (por ejemplo, la primera del imperativo), se sustituye por un asterisco (*) para indicarlo.

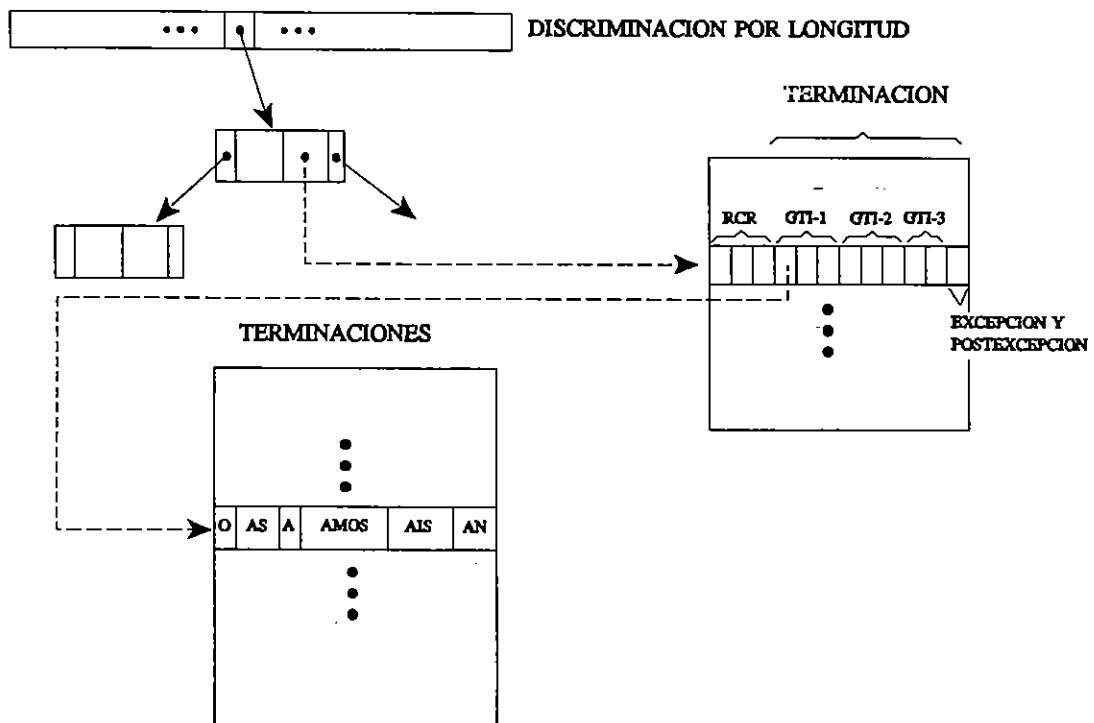


Figura 2. Estructura de Conjugación.

El segundo preprocesador toma el fichero conteniendo todos los verbos irregulares relacionados con sus modelos, de la forma:

verbo irregular:verbo modelo

y los inserta en la organización nuclear. Se genera así la estructura completa de conjugación de verbos, lista para ser usada por el módulo conjugador. Nótese que es preciso alimentar además al conjugador con un infinitivo y el código del tiempo a conjugar.

El conjugador averigua —accediendo a la estructura— si se trata de un verbo regular o irregular.

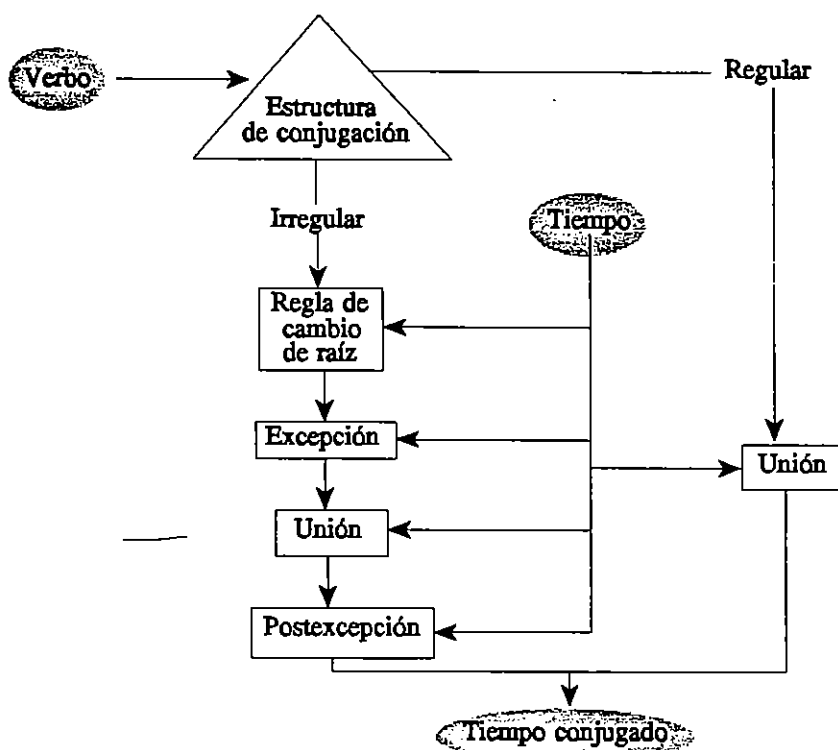


Figura 3

En caso de que sea irregular no hay más que seguir la referencia contenida en el nodo para determinar la regla a aplicar según el tiempo que se desee y el conjunto de terminaciones que se debe utilizar. Entrando entonces en juego la secuencia de acciones: a) se efectúa el cambio de raíz, b) se aplican las excepciones, c) se realiza la unión con las terminaciones y d) se llevan a cabo las postexcepciones. La figura 3 muestra el proceso seguido en la conjugación.

Si por el contrario es regular, no hay que aplicar cambio de raíces y, en consecuencia, tampoco excepciones o postexcepciones; las únicas acciones a tomar son las de realizar las uniones correctamente.

Conclusión

- Se ha querido realizar un producto capaz de conjugar cualquier tiempo verbal, tanto regular como irregular. Con estos últimos se ha afrontado el problema de sistematizar sus especificidades; creemos que se ha conseguido en gran medida, ya que son muy excepcionales los verbos con un alto grado de irregularidad.
- No se han encarado algunos aspectos como el caso de los verbos defectivos que, a primera vista, parecen poco sistematizables.
- Hay aspectos susceptibles de ser mejorados, como el tratamiento de los pasivos irregulares.

- Sería de interés dotar al sistema de la capacidad de conjugar verbos compuestos por similitud con uno conocido, sin necesidad de disponer de él en la estructura. Ello plantearía el dilema de: ¿se permite al usuario conjugar cualquier cosa que se le ocurra y tenga apariencia verbal o se le restringe a los verbos existentes?.

Bibliografía

- «Diccionario Actual de la Lengua Española». VOX, 1990.
- «Esbozo de una Nueva Gramática de la Lengua Española». R.A.E. , Espasa Calpe, 1989.
- «Diccionario de Dudas y Dificultades de la Lengua Española», 9ª Edición. Manuel Seco. Espasa Calpe, 1991.
- «Todos los Verbos Castellanos Conjugados», 17ª Edición. Ramón Alsina. Teide, 1990.