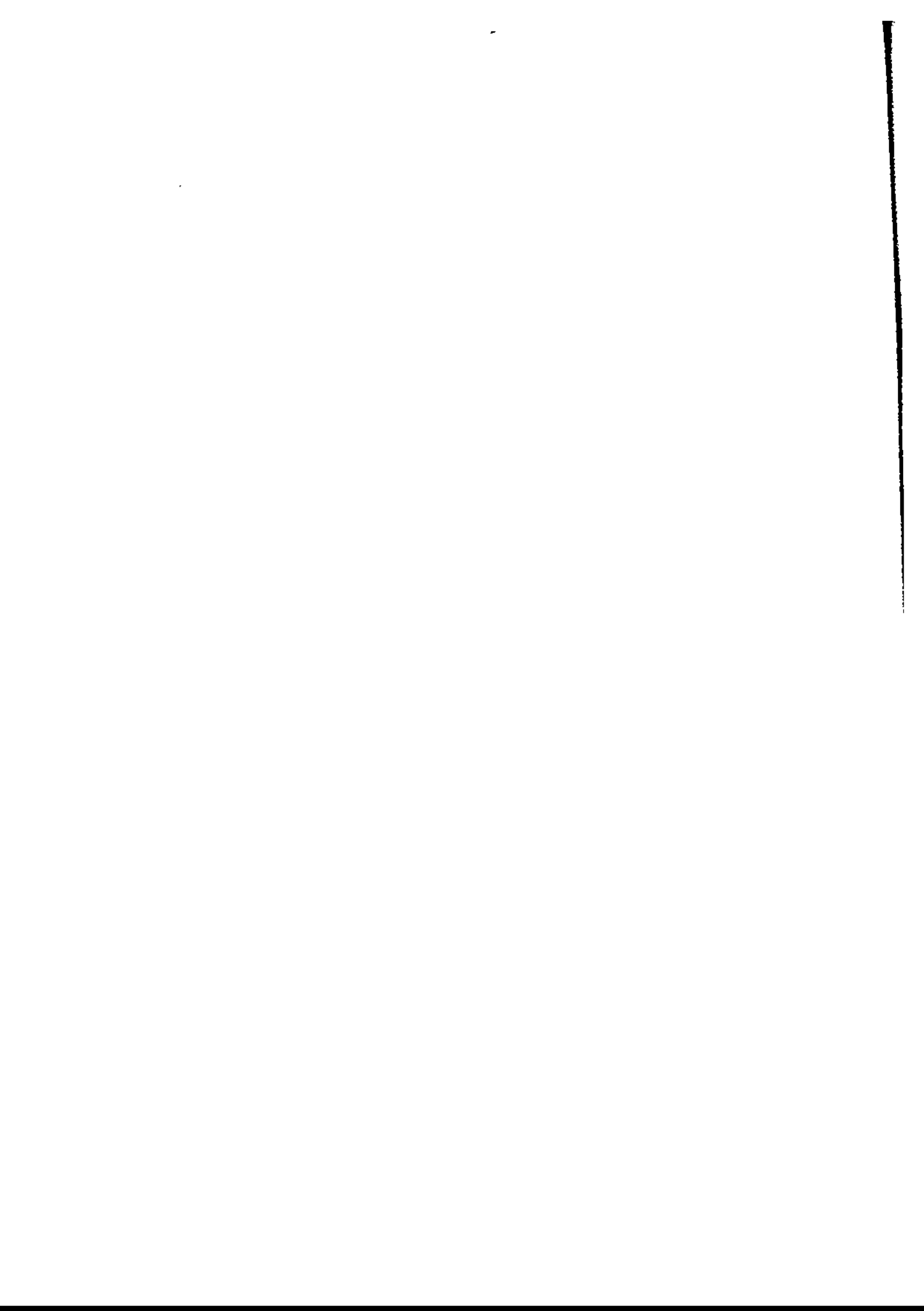


Demostraciones



Web WordNet Interface.

(Un sistema para el desarrollo de WordNets a través de la Web)

Laura Benítez, Sergi Cervell, Gerard Escudero
Mónica López, German Rigau y Mariona Taulé
Grup de Tractament de Llenguatge Natural.
Universitat Politècnica de Catalunya.
Universitat de Barcelona.
{laurab, escudero, g.rigau}@lsi.upc.es
{cervell, lopez, mtaule}@lingua.fil.ub.es

Resumen

Este sistema se ha construido dentro del marco del desarrollo de las wordnets (bases de conocimiento léxicas) del castellano y del catalán¹.

Los datos de estas wordnets están siendo modificados constantemente por los diferentes miembros del grupo al mismo tiempo y además desde diferentes ubicaciones. Esto ha creado la necesidad de garantizar la integridad de estos datos y su consistencia. Para facilitar el trabajo de los usuarios se ha construido una interficie a esta base de datos. Está permite buscar y manipular los datos de manera fácil y efectiva, realizar consultas multilingües que dan una visión intuitiva de las jerarquías y, por último, consultar ciertas fuentes léxicas (diccionarios monolingües y bilingües).

1. Base de Datos

Para minimizar el coste se ha escogido una base de datos de libre distribución.

Por el momento, la base de datos contiene todos los datos de WordNet 1.5., y palabras y glosas asociadas a conceptos para las demás wordnets. Las relaciones semánticas de todas la wordnets son las de la wordnet inglesa. Las wordnets cargadas son fragmentos de las del castellano, vasco y catalán.

Es fácil integrar nuevas wordnets en el sistema y la estructura de la base de datos permite desarrollar software de exportación / importación fácilmente. De hecho, tenemos un programa de exportación e importación de

wordnets monolingües a las herramientas del proyecto EuroWordNet² (Novell, 1996-1998).

2. Interficie

Para construir las wordnets castellana y catalana se ha hecho necesaria la creación de una interficie de acceso a la base de datos y las diferentes fuentes léxicas. Dada la distancia física que separa a los diferentes miembros del proyecto que modifican los datos y para permitir a cualquier usuario consultar los datos abiertamente, se ha desarrollado la interficie para que funcione a través de la Web.

Las operaciones que permite la interficie son las siguientes:

- **Consult Multilingual WordNet.** Su objetivo principal es ayudar en la evaluación de errores y omisiones en las jerarquías de las diferentes wordnets. Para hacer posible esto, la interficie puede mostrar la jerarquía semántica de una de las wordnets, para cualquier relación semántica, cogiendo como punto de partida un concepto, sentido o palabra de una de las lenguas disponibles.
- **Edit Multilingual WordNet.** Su objetivo básico es la edición de las palabras y glosas de todas las wordnet, excepto de la inglesa.
- **Consult Lexical Resources.** Permite consultar los diferentes diccionarios monolingües y bilingües disponibles.
- **Edit Statistics.** Esta opción muestra una lista del histórico de cambios realizados con la interficie.

3. Líneas futuras

Las ampliaciones que se están realizando ahora o que se seguirán en un futuro próximo vienen listadas a continuación:

¹ El sistema presentado en este documento está enmarcado en el contexto de los proyectos ITEM (TIC96-1243-C03-03), EuroWordNet (LE2-4003) y CREL (uno de los centros de referencia de la *Generalitat de Catalunya* para la investigación y el desarrollo).

² Polaris 1.3. de Novell Inc.

- Ampliación de la base de datos y de la interficie para que acepte todos los datos especificados por el proyecto EuroWordNet.
- Adición de la consulta de corpus a la consulta de fuentes léxicas.
- Inclusión de los datos de la clase Levin a los verbos de las wordnets.
- Generador de informes con estadísticas del estado de los datos y de su evolución en el tiempo.

Bibliografía

- Aronson, A. & Lowery, J. (1997). *HTML 3.2. Manual of Style*. - Emeryville: Ziff Davis Press, 1997. - ISBN: 1-56276-529-9.
- Asterias, J.; Climent, S.; Ferreras, J.; Rigau, G. & Rodríguez, H. (1997). "Combining Multiple Methods for the Automatic Construction of Multilingual WordNets". In proceedings of Conference on Recent Advances on NLP. (RANLP'97). Tzgov Chark, Bulgaria 1997.
- Benítez, L.; Cervell, S.; Escudero, G.; López, M.; Rigau, G.; Taulé, M. (1998a). "Methods and tools for building the Catalan WordNet". In workshop on Language Resources for European Minority Languages at Conference on Language Resources and Evaluation (LREC). Granada, Spain, 1998.
- Benítez, L.; Escudero, G.; Ferreras, J. & Rigau, G. (1998b). "Making Tractable the *Enciclopèdia Catalana* Bilingual Dictionary". Technical Report LSI-98-3-T. LSI Department, *Universitat Politècnica de Catalunya*, 1998.
- Benítez, L.; Escudero, G.; Ferreras, J. & Rigau, G. (1998c). "The Application of some Automatic Methods for the Construction of Multilingual WordNets". Technical Report LSI-98-4-T. LSI Department, *Universitat Politècnica de Catalunya*, 1998.
- Benítez, L.; Escudero, G.; Ferreras, J. & Rigau, G. (1998d). "WWI: A Multilingual WordNet Interface using the Web. Technical Manual". Technical Report LSI-98-5-T. LSI Department, *Universitat Politècnica de Catalunya*, 1998.
- Benítez, L.; Escudero, G.; Ferreras, J. & Rigau, G. (1998e). "WWI: A Multilingual WordNet Interface using the Web. User Manual". Technical Report LSI-98-6-T. LSI Department, *Universitat Politècnica de Catalunya*, 1998.
- CGI. The Common Gateway Interface Specification. [Web Page] "<http://hoohoo.ncsa.uiuc.edu/cgi/interface.html>".
- Cookies. Client Side State - HTTP Cookies. [Web Page] "<http://home.netscape.com/newsref/std/cookiespec.html>".
- Goodman, D. (1996). *Danny Goodman's JavaScript Handbook*. - Foster City: IDG Books Worldwide, 1996. (tr: *Programación en JavaScript*. - Madrid: Anaya Multimedia, 1997. - ISBN: 84-415-0080-0).
- Miller, G. (1990). "Five papers on WordNet". In Special Issue of International Journal of Lexicography: actas 1990.
- Mini SQL. Hughes Technologies. Mini SQL 2.0 User Guide. [Web page] "<http://www.Hughes.com.au/>".
- Muelver, J. (1996). *Creating CoolT M Web Pages with Perl*. - Foster City: IDG Books Worldwide, 1996. (tr: *Creación de páginas Web con PERL*. - Madrid: Anaya Multimedia, 1997. - ISBN: 84-415-0076-2).
- Novell (1996-1998). The EWN Database Tool. Version 1.3.3. Copyright, Novell Inc. 1996-1998.
- Rigau, G.; Asterias, J. & Agirre, E. (1997). "Combining Unsupervised Lexical Knowledge Methods for Word Sense Disambiguation". In proceedings of the 34th Annual Meeting of the Association for Computational Linguistics (ACL'97). Madrid, Spain, 1997.
- Wall, L.; Christiansen, T. & Schwartz, R. L. (1996). *Programming Perl*. - Second Edition. - Sebastopol: O'Reilly & Associates, 1996. - ISBN: 1-56592-149-6.

SAPLEN (Sistema Automático de Pedidos en Lenguaje Natural)

Ramón López-Cózar Delgado
Dpto. Electrónica y Tecnología de Computadores
Universidad de Granada
18071 Granada, España
e-mail: gas0@elvira.ugr.es

Resumen

Nuestra demo sería similar a la realizada en el anterior Congreso de la SEPLN, celebrado en Madrid. La demo consistiría en realizar una exhibición del sistema de diálogo que estamos desarrollando, denominado SAPLEN (Sistema Automático de Pedidos en Lenguaje Natural). La finalidad de este sistema es atender mediante el habla las consultas y pedidos de los clientes de los restaurantes de comida rápida, si bien, en la versión actual, la comunicación se realiza de forma escrita. El sistema SAPLEN ha sido mejorado substancialmente desde el anterior Congreso.

Actualmente, tiene presentes posibles errores de reconocimiento de las palabras pronunciadas por los usuarios, para lo cual usa valores de confianza asociados a las palabras, y un umbral de confianza que determina la corrección en el reconocimiento de las mismas. La estrategia del diálogo ha sido mejorada notablemente, por ejemplo, el cliente debe proporcionar su número de teléfono y la dirección a la cual deben ser enviados los productos encargados.



Análisis parcial de corpus con la gramática AVALÓN

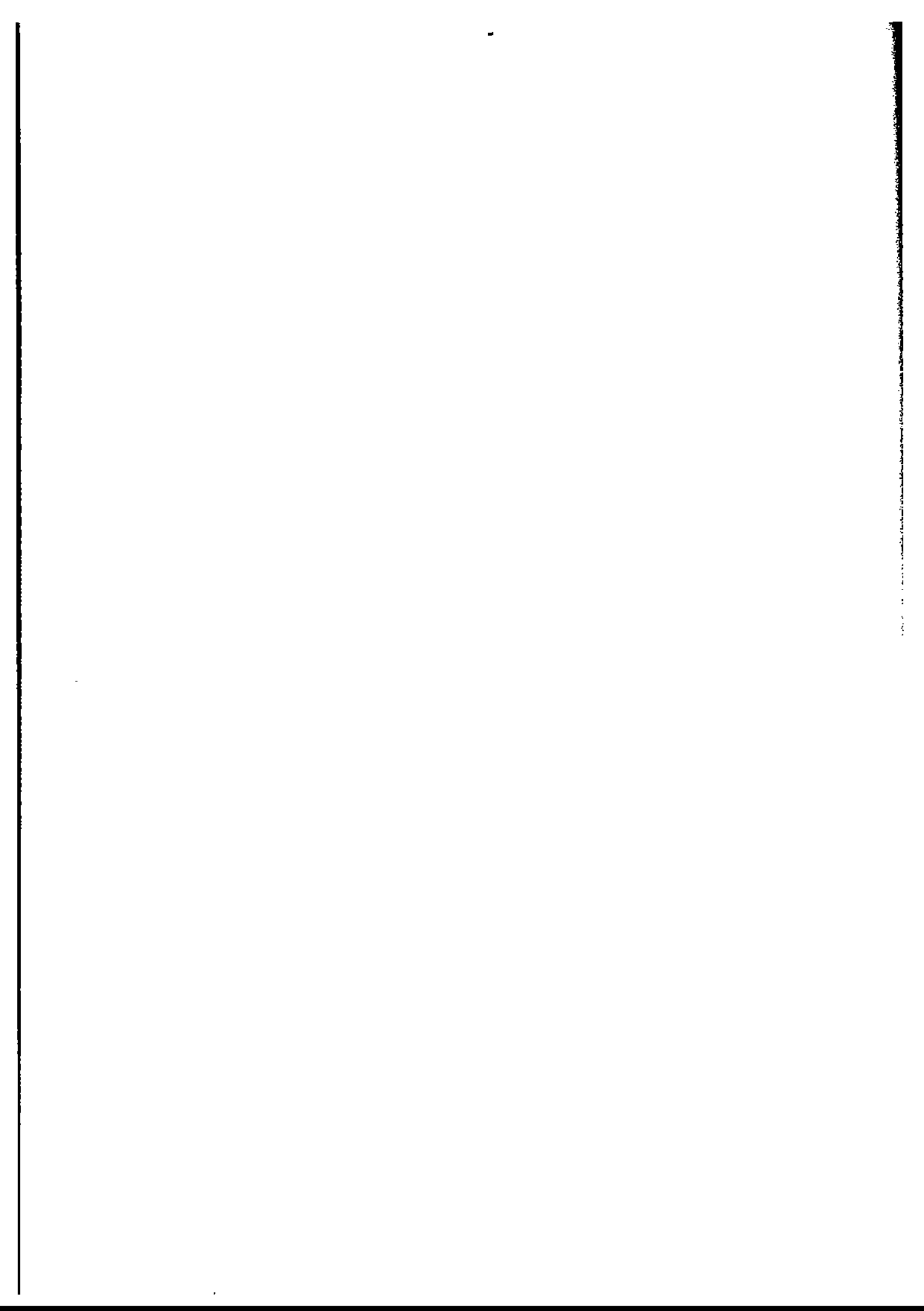
Conchita Álvarez Lebrede
Pilar Alvariño Alvariño
Adelaida Gil Martínez
Teresa Romero Quintáns
Paula Santalla del Río
Susana Sotelo Docío

fesdocio@usc.es

Resumen

En relación con la comunicación aceptada con el título *AVALÓN, una gramática formal basada en corpus*, proponemos una demostración del uso del *parser* generado a partir de una adaptación de la gramática para llevar a cabo análisis parcial de corpus no restringido. La demostración constará de las partes siguientes:

- Proceso de generación del lexicon para la gramática. Tendrá dos partes: en primer lugar, se mostrará la producción de categorías nominales en sus diversas fases. En segundo lugar, tendrá lugar la producción de categorías verbales y la inserción de restricciones semánticas en estas categorías. Finalmente se mostrará un interfaz que integra todo el proceso.
- En la segunda parte de la demostración presentaremos el preprocesado de un texto: esto implica segmentación, conversión a minúsculas de las mayúsculas iniciales y la sustitución de abreviaturas donde corresponda.
- En la parte final, tendrá lugar el análisis parcial del texto preprocesado previamente; por análisis parcial entendemos la segmentación de los enunciados identificados en el preprocesado en sucesiones de frases (nominales, adjetivas, adverbiales o verbales, sin tener en cuenta la coordinación). En la medida de lo posible intentamos que esta segmentación fraseal reproduzca una estructura de constituyentes sintácticos en tanto que funcionales.



Una plataforma para soporte a la investigación lingüística: Análisis morfológico de la lengua ugarítica

Ana García-Serrano, Jesús Contreras
ISYS-Hemeneumática
DIA-Facultad de Informática
Universidad Politécnica de Madrid
{agarcia,jcontrer}@isys.dia.fi.upm.es

Resumen

Herramienta para ayuda a la investigación

Se presenta una herramienta para ayuda a la investigación lingüística en el área de la morfología. Ha sido desarrollada para cubrir las necesidades de filólogos del grupo de investigación del laboratorio de Hermeneumática del CSIC, en su labor de estudio de la lengua ugarítica. El principal objetivo es ofrecer soporte automático para el planteamiento de hipótesis y revisión de las conjeturas durante el trabajo del filólogo en el estudio de una lengua desconocida. Debe permitir al filólogo la gestión y actualización de los diferentes tipos de informaciones que el filólogo maneja durante las labores propias de su investigación. El diseño y el desarrollo de la aplicación se han orientado al usuario final bajo la aproximación de un sistema basado en el conocimiento con interfaces amigables para todos los componentes de información así como de los procesos y métodos de su gestión.

Lengua ugarítica

La lengua estudiada proviene del reino de Ugarit cuyos restos se hallan en el territorio de la actual Siria y se datan del segundo milenio Antes de Cristo. Es uno de los primeros alfabetos fonéticos y la lengua ugarítica pertenece a la familia de las semíticas. El corpus disponible se halla en tablillas de cerámica escrito con letra cuneiforme, y contiene unas 60.000 palabras completas, el resto son palabras incompletas debido al deterioro del material cerámico. Tanto el léxico como todo el conocimiento sobre la morfología ugarítica se ha tomado del laboratorio de hermeneumática del centro de estudio del antiguo oriente del CSIC, que ha sido utilizado en el AMU un analizador morfológico desarrollado sobre una plataforma Macintosh [Cunchillos & Cervigón].

Metodología KSM

Se ha utilizado durante el diseño del modelo de la morfología ugarítica la metodología KSM [Cuenca & Molina 96] para estructurar la morfología en diferentes áreas de conocimiento. A partir de esta estructuración se identificó el formalismo específico de cada área concreta y descomponer las tareas en sub tareas utilizando diferentes paradigmas. Siguiendo los

pasos de la metodología KSM se identificaron las áreas de conocimiento (jerarquía), definiendo para cada una su parte declarativa (la base de conocimiento) y su parte funcional (tareas). Las unidades primarias en esta jerarquía contienen el conocimiento de dominio que es gestionado por el experto lingüista: añadiendo nuevos elementos, eliminando o revisando la validez de otros o simplemente modificando los existentes.

Las áreas primarias identificadas para el modelo de morfología son cinco. La primera contiene el léxico con la relación de lexemas atestiguados de la lengua. Se ha utilizado una tabla con las tareas de consulta, inserción y borrado asociadas. La segunda contiene las reglas de formación que permiten construir palabras morfológicamente correctas de la lengua a partir de secuencias válidas de morfemas y lexemas. Otra área es la que contiene la relación de todos los morfemas conocidos, con un formalismo similar al área del léxico. El área de clasificación gramatical utiliza un formalismo de jerarquía de marcos para clasificar la palabra de entrada según sus componentes, identificado su categoría y calculando el valor de los atributos gramaticales. El último área incorpora el conocimiento sobre los fenómenos lingüísticos que puede sufrir una palabra, como la asimilación de componentes, su transformación o pérdida en determinadas condiciones. También permite definir la forma de cálculo de los atributos gramaticales a partir de los componentes identificados.

Una vez diseñado el modelo, se instancia con el conocimiento del dominio de la lengua ugarítica en las bases de conocimiento de cada área primaria. La aplicación final utiliza el modelo diseñado, ofreciendo un interfaz amigable a cada base de conocimiento, haciendo transparente al usuario la codificación interna del conocimiento y permitiendo su fácil gestión.

Funcionalidades

El sistema permite analizar morfológicamente una palabra dada ofreciendo todas las soluciones posibles, según el conocimiento de la aplicación. Este conocimiento es gestionado por los lingüistas, de manera que lo pueden alterar, borrar o

incrementar. Cada modificación tiene efectos inmediatos en el resultado del siguiente análisis, de manera que la aplicación se puede utilizar para validar conjeturas, observando el comportamiento y los resultados. En cada análisis el sistema presenta una explicación que contiene los elementos de las bases de conocimiento que ha utilizado, justificando el resultado y el análisis propuesto. Además del análisis morfológico de una palabra, es posible definir parcialmente la entrada al sistema, dejando que éste genere las diferentes alternativas de solución. De esta forma pueden obtenerse todas las palabras que cumplen unas restricciones morfosintácticas (de género, número, etc..) o es posible especificar parcialmente los componentes de la palabra de entrada, al definirse solamente una parte y siendo el sistema el que genera todas las restituciones morfológicamente válidas.

Al presentar al filólogo todas las soluciones que sean correctas morfológicamente, permite la validación de las hipótesis introducidas o efectuar posteriormente el análisis sintáctico automático exhaustivo al disponer de todas las alternativas morfológicas posibles.

Pasos de la demostración

En primer lugar se mostrará la funcionalidad básica del sistema, introduciendo una palabra de entrada y efectuando un análisis. Se mostrarán todas las soluciones de descomposición con sus posibles análisis según el estado del conocimiento incorporado. El sistema dispone de una base de datos donde almacena las soluciones inferidas anteriormente, de manera que el usuario puede decidir si recuperar un análisis hecho o hacer una nueva inferencia. Según las explicaciones que se ofrecen para la solución, es posible editar todos los elementos de las bases de conocimiento que hayan sido utilizadas a lo largo del proceso del análisis.

En este punto de la demostración se pueden introducir modificaciones para observar que efectos tienen sobre los resultados obtenidos. La parte menos atestiguada de la morfología ugarítica esta comprendida en el área de fenómenos lingüísticos, siendo interesante formular algunas hipótesis y poder verificar las consecuencias que produce. También es posible añadir nuevos elementos descriptivos que permitan clasificar palabras de entrada, añadir borrar o modificar lemas de composición o modificar lexemas o morfemas. Una modificación interesante consiste en añadir al léxico lexemas de otras lenguas similares (hebreo antiguo, fenicio, etc..) y ver como se comporta bajo las restricciones de la morfología ugarítica.

Dado el estado de las tablas, el filólogo se ve obligado a hacer restituciones de palabras incompletas. El sistema permite definir palabras incompletas en su entrada, indicando que algunos componentes tienen forma indefinida e inferir así todas las posibles soluciones que satisfagan esta definición parcial. Para evitar que el sistema haga una generación intratable de miles de palabras se pueden añadir restricciones de tipo gramatical, si se conocen. Se puede limitar la categoría (verbo), o algún atributo (número, género, etc..) utilizando algunas reglas de tipo sintáctico atestiguadas. La complejidad del proceso de inferencia es lineal respecto al número de soluciones que se ofrecen, siendo recomendable, por tanto, introducir todas las restricciones conocidas en la entrada.

Ficha técnica

El diseño y la implementación del prototipo se realizó con la herramienta KSM operativa sobre Solaris OS. Esta plataforma básica se enriqueció con interfaces en Tcl/Tk.

La aplicación actual se ha realizado en Windows NT 4.0, utilizando Ciao Prolog [Hermenegildo 94] para el desarrollo de la parte de gestión de conocimiento así como de las tareas propias del sistema. El Ciao Prolog es un lenguaje en desarrollo del grupo CLIP del Departamento de Inteligencia Artificial de la UPM que permitirá el desarrollo del modelo distribuido de esta aplicación. Los interfaces de las bases de conocimiento así como la aplicación principal se han implementado utilizando Microsoft Visual Basic 5.0.

Referencias

- [Cuenca & Molina 96] J. Cuenca and M. Molina. KSM: An environment for Design of Structured Knowledge Models. Chapter of the book "Knowledge-Based Systems-Advanced Concepts, Techniques and Applications". De. Spyros G. Tzafestas. World Scientific Publishing Company.
- [Cunchillos & Cervigón 96] J.L. Cunchillos and R. Cervigón (1996) "SIAMTU: El Analizador Morfológico Ugarítico en el Entorno 4D, Technical Report.
- [Cunchillos & Zamora 95] J. L. Cunchillos and J. A. Zamora (1995): Gramática Ugarítica Elemental, Ediciones Clásicas.
- [García-Serrano, Contreras, 98] A. García Serrano, J. Contreras "A Computational Platform for Ugaritic Morphological Analysis", LREC 98.
- [Hermenegildo 94] .M. Hermenegildo, the CLIP group. Some Methodological Issues in the Design of CIAO - A Generic, Parallel, Concurrent Constraint System. Principles and Practice of Constraint Programming, pages 123-133, Springer-Verlag, LNCS, 874, May 1994.

Análisis morfosintáctico orientado a corpus del español

J. Atserias, J. Carmona, I. Castellón, S. Cervell, M. Civit, Ll. Márquez,
M.A. Martí, Ll. Padró, H. Rodríguez, M. Taulé, J. Turmo

UPC-UB

Resumen

Esta demostración presenta el proceso de análisis morfosintáctico de corpus del español mediante herramientas y recursos lingüísticos desarrollados en los proyectos ITEM¹ y LEXESP².

El proceso que presentamos consiste en tres etapas:

1. - análisis morfológico
2. - desambiguación morfológica
3. - análisis sintáctico

Estos tres procesos han sido integrados en GATE (General Architecture for Text Encoding), un entorno gráfico desarrollado en Sheffield para la integración de diferentes herramientas de Ingeniería lingüística [Cunningham 96].

1. A. Morfológico: MACO

El conocimiento lingüístico se ha organizado en clases de raíces y en modelos flexivos asociados a las raíces. Actualmente Maco es capaz de reconocer un total del 800.000 formas que corresponden a unos 90.000 lemas [Carmona98]. La arquitectura de Maco es modular y cada uno de los módulos puede ser activado o desactivado en cada aplicación concreta. Los módulos desarrollados

¹ ITEM TIC96 - 1243 - C03 - 02 (<http://sensei.ieec.uned.es/item>) tiene como objetivo principal la integración de herramientas y recursos de PLN en una única plataforma con el fin de facilitar la construcción de sistemas de extracción y recuperación multilingüe. Las lenguas implicadas en el proyecto son el español, el catalán y el euskera.

² Lexesp (Lexesp II Acción especial APC96-0125), proyecto iniciado en el Departamento de Psicología de la Universidad de Oviedo, tenía como objetivo la creación de una base de datos textual. El corpus desarrollado tiene 5,5 millones de palabras y incluye noticias de diversos ámbitos, textos literarios, artículos científicos, etc.

son los siguientes: reconocedor de fechas, abreviaturas, nombres propios, compuestos, números, signos de puntuación, palabras y clíticos. El analizador real es el que se encarga del reconocimiento de palabras.

La implementación de MACO se ha realizado en perl -5.0 sobre plataforma UNIX.

El análisis se ha aplicado sobre una muestra de corpus de Lexesp (100.000 palabras) proporcionando para cada palabra todos las posibles interpretaciones (palabra, lema y categoría Parole). La velocidad aproximada de análisis es de 600 palabras por segundo en una SUN Ultra Sparc y de 200 palabras por segundo en Linux en un Pentium-120. El análisis sobre la globalidad del corpus (5.5 millones de palabras) ha dado como resultado una cobertura de un 99.5%. El corpus presenta un 39.26% de palabras ambiguas y una media de 2.63 de etiquetas por palabra ambigua y un 1.64 por palabra del corpus. El índice de acierto es de 99.3% (palabras que tienen la etiqueta correcta como resultado).

2. Desambiguación Morfológica

La desambiguación morfológica se ha realizado mediante dos etiquetadores.

1) Un etiquetador basado en árboles de decisión [Márquez & Rodríguez 98] que adquiere el conocimiento a partir de un corpus anotado (aprendizaje supervisado) que también aprende reglas específicas para el tratamiento de palabras desconocidas (palabras que no aparecen el corpus modelo).

2) Un etiquetador basado en la relajación de etiquetas [Padró 98], que puede utilizar información de diversas fuentes, desarrollado en un formato de restricciones contextuales.

Los resultados obtenidos por ambos etiquetadores indican un índice superior a un 97%. Mediante la combinación de ambos se puede mejorar la precisión hasta 97'82%

3. Análisis Sintáctico

El análisis sintáctico [Castellón et al. 98] se ha llevado a cabo mediante el analizador TACAT y una gramática independiente de contexto GRAMESP.

TACAT [Atserias, J. & H.Rodríguez 1998] es un analizador desarrollado en C++ que sigue una estrategia ascendente y utiliza reglas de contexto libre. Una de las funcionalidades que proporciona este analizador es el filtrado de la estructura de salida del análisis, evitando la aparición de determinadas categorías no léxicas. Tacat permite la aplicación incremental de las gramáticas ya que el formato del texto de entrada puede estar anotado morfológicamente (una o más categorías) o bien un texto analizado sintácticamente.

Gramesp [Civit et al 98] identifica con un alto nivel de calidad grupos sintagmáticos (sn,sp,sa,sadv), formas verbales complejas (gv) y la coordinación de grupos léxicos. Con un porcentaje más bajo Gramesp propone grupos superiores como sv y oraciones.

La demostración se realizará en dos entornos de trabajo, por un lado mediante GATE. Y por otro en UNIX para mostrar la velocidad del sistema en un proceso masivo en el que no se requiere un modo de visualización. También está disponible esta demostración en la dirección <http://nipadio.lsi.upc.es/cgi-bin/demo.pl>.

Referencias

- Acebo,S.; Ageno,A.; Climent,S.; Farreres,X.; Padró,L.; Ribas,F.; Rodríguez,H. & Soler,O. (1994). "MACO: Morphological Analyzer Corpus-Oriented." ESPRIT BRA-7315 Aquilex II, Working Paper \#31.
- Atserias J y H. Rodríguez (1998) "TACAT: TAGged Corpus Text Analyzer" Technical Report LSI-UPC RT-2-98.
- Castellón, I. ; Atserias, J. & Civit, M. (1998). Syntactic Parsing of Unrestricted Spanish Text.
In *Proceedings of 1st International Conference on Language Resources and Evaluation, LREC'98*. Granada, Spain.
- Carmona,J.; Cervell,S.; Márquez, L.; Martí,M.A.; Padró,L.; PlacerR.; Rodríguez,H.; Taulé,M. & Turmo,J. (1998). "An Environment for Morphosyntactic Processing of Unrestricted Spanish Text." In *Proceedings of 1st International Conference on Language Resources and Evaluation, LREC'98*. Granada, Spain.
- Civit y I. Castellón (1998) "Gramesp: una gramática de corpus para el español" RESLA 1998 (en prensa)
- Cunningham,H.; Wilks,Y. & Gaizauskas,R. (1996). GATE - a General Architecture for Text Engineering. In *Proceedings of 16th International Conference on Computational Linguistics, COLING '96*. Copenhagen, Denmark.
- Márquez,L. & Rodríguez,H. (1998). "Part-of-Speech Tagging Using Decision Trees."
In *Proceedings of the 10th European Conference on Machine Learning, ECML'98*. Chemnitz, Germany.
- Padró,L. (1996). "POS Tagging Using Relaxation Labelling." In *Proceedings of 16th International Conference on Computational Linguistics, COLING '96*. Copenhagen, Denmark.
- Padró,L. (1998) *A Hybrid Environment for Syntax-Semantic Tagging*. PhD Thesis. Dept. Llenguatges i Sistemes Informatics. Universitat Politècnica de Catalunya. Barcelona

CON-TEXT. Un corrector gramatical de bajo nivel

Flora Ramírez Bustamante
Fernando Sánchez León
Laboratorio de Lingüística Informática
Facultad de Filosofía y Letras
Universidad Autónoma de Madrid
E-28049 Madrid (Spain)
{flora;fernandog@maria.llif.uam.es}

Thierry Declerck
DFKI GmbH
(German Research Center
for Artificial Intelligence)
Stuhlsatzenhausweg 3
D-66123 Saarbrücken (Germany)
declerck@dfki.uni-sb.de

Resumen

En este trabajo se presentan los resultados del proyecto Con-Text, un corrector gramatical para el español basado en el uso de técnicas de bajo nivel, tales como la segmentación y análisis morfológico de textos. Asimismo, se describen las herramientas de análisis lingüístico del nivel morfológico sobre las que se ha construido este sistema y se presenta la doble arquitectura sobre la que se ha implementado.

Esta demostración está basada en el artículo que con el mismo título se presenta en el XIV Congreso de la SEPLN.



Conversor texto-voz multilingüe para español, catalán, gallego y euskera

M.A. Rodríguez, J.G. Escalada, D. Torre
Telefónica I+D

Resumen

Se demostrará el funcionamiento del conversor texto-voz multilingüe de Telefónica I+D para los idiomas español, catalán, gallego y euskera. Este sistema es el resultado de un proyecto de investigación; no es un producto comercial.

La demostración servirá para mostrar de una forma práctica los contenidos que se presentarán en la comunicación "Conversor texto-voz multilingüe para español, catalán, gallego y euskera", que ya ha sido aceptada para su presentación en el congreso.

El conversor texto-voz multilingüe es un programa que funciona en tiempo real. Será posible introducir textos para comprobar la síntesis de voz proporcionada por el sistema en los casos que se quieran comprobar. Se podrá modificar el idioma de funcionamiento y variar distintas características de la voz sintética, como la velocidad y el tono.

