

TESIS

FUNDAMENTOS Y LÍMITES DE LOS SISTEMAS DE VERIFICACIÓN AUTOMÁTICA DE LA SINTAXIS Y EL ESTILO

Javier Gómez Guinovart

Universidad de Vigo

uvifejgg@cesga.es

Tesis doctoral presentada el 28 de junio de 1996 en la Facultad de Filología de la Universidad de Santiago de Compostela. Director: Dr. Juan Carlos Ruiz Antón. Ponente: Dr. Guillermo Rojo Sánchez. Tribunal: Dr. Joseba Abaitua Odriozola, Dra. Irene Castellón Masalles, Dr. José María García-Miguel Gallego, Dr. Leonardo Gómez Torrego, Dr. Kepa Sarasola Gabiola.

Resumen:

En esta tesis se presentan los fundamentos y límites lingüísticos e informáticos de la verificación automática de la sintaxis y el estilo, prestando una atención especial a los problemas de corrección lingüística de los textos redactados en lengua castellana. El objetivo de la tesis consiste en establecer las bases teóricas, el ámbito de aplicación y las diversas estrategias de implementación informática de los sistemas de verificación lingüística desarrollados en el área del procesamiento del lenguaje natural.

La tesis consta de tres partes principales. La primera parte está dedicada a la caracterización del error gramatical y a su procesamiento informático. En

ella, se revisan los fundamentos lingüísticos teóricos de la distinción entre gramaticalidad y agramaticalidad, se examinan diversos criterios de caracterización y clasificación de los errores gramaticales, y se analizan las diferentes estrategias informáticas para su tratamiento automático, centrándose en los procedimientos elaborados por la lingüística de corpus para la detección de los errores sintácticos a partir del etiquetado de los textos, en las técnicas de atenuación de condiciones gramaticales propias del análisis gramatical automático resistente, y en los métodos de verificación sintáctica mediante reconocimiento de patrones.

La segunda parte de la tesis está dedicada a caracterizar la impropiedad del estilo y a investigar las posibilidades de su tratamiento informático. En ella, tras analizar las principales aproximaciones teóricas al estudio del estilo, se establece una caracterización de la impropiedad estilística ajustada a las posibilidades de su procesamiento automático, y se examinan las diversas estrategias informáticas aptas para esta aplicación, centrándose en los métodos lingüístico-cuantitativos empleados para valorar el nivel de dificultad de comprensión del sentido de un texto, en las estrategias de verificación fundadas en tipologías textuales estilísticas, y en las técnicas de detección basadas en la gradación de los rasgos estilísticos y en la determinación de sus umbrales de impropiedad.

Por último, en la tercera parte de la tesis, se presenta una aportación a la metodología de evaluación de los componentes de reconocimiento y corrección de errores de los sistemas de verificación de la sintaxis y el estilo.

**Aportaciones a la resolución de la elipsis en lenguaje natural
utilizando técnicas incrementales**

Tesis doctoral. Universidad Politécnica de Valencia. Septiembre 1996

Autor: Manuel Palomar (e-mail: mpalomar@dlsi.ua.es)

Directora: Lidia Moreno

El propósito principal de esta Tesis ha sido el tratamiento y resolución de la elipsis que se produce en la coordinación, como uno de los problemas principales destacados en el análisis de sentencias para la obtención de la representación de su significado.

Mucho se ha debatido en los últimos años sobre el apropiado nivel de procesamiento del lenguaje que trate la resolución de la elipsis. Y se ha debatido desde dos puntos de vista; la aproximación sintáctica y la semántica. La aproximación sintáctica, Lappin, McCord, Haik, Fiengo y May, consiste en la reconstrucción sintáctica del constituyente incompleto a partir de su antecedente. Mientras que la aproximación semántica, Dalrymple, Kehler, Gawron y Peters, consiste en identificar un significado para el constituyente incompleto y asignárselo directamente; es decir, la información se recupera en la representación semántica.

Para esta resolución, nosotros hemos adoptado el punto de vista semántico, donde se toma la forma lógica del antecedente para la reconstrucción del constituyente incompleto o elidido. Consideramos que no es necesario una reconstrucción sintáctica de los elementos elididos, ya que la reconstrucción semántica proporciona los mecanismos adecuados para una vez identificado el antecedente del constituyente elidido asignarle su forma lógica, por lo que resuelve de forma más eficiente el problema.

El problema de la resolución de la elipsis lo hemos dividido en dos tareas: la primera será determinar la estructura paralela entre el constituyente incompleto y su antecedente, y por otro lado la formación de

la relación implícita. La división del problema de la elipsis en dos partes permite una mejor descripción de la formación de la relación y una mejor caracterización del problema de determinación del paralelismo. Por lo que desde nuestro punto de vista, para determinar el paralelismo es necesario criterios sintácticos semánticos y posteriormente para la formación de la relación se pueden adoptar criterios semánticos. Con ello, aportamos un nuevo método de resolución que incorpora una adaptación de recientes trabajos sobre la elipsis, donde resalta la idea de estructuras paralelas, pero mientras estas aproximaciones acentúan el paralelismo semántico, nosotros usamos tanto el paralelismo sintáctico como el semántico.

Este nuevo método está basado en las técnicas incrementales de evaluación de programas Datalog como herramienta para la definición e implementación de gramáticas de contexto libre; Y está desarrollado mediante un nuevo tipo de Gramáticas de Cláusulas Definidas (DCGs), las Gramáticas Datalog (DLG), basadas en la teoría de bases de datos.

El nuevo formalismo que presentamos es una extensión de las Gramáticas Datalog, presentadas por V.Dahl, P.Tarau, L.Moreno y M.Palomar, el cual mejora la eficiencia de las tradicionales Gramáticas de Cláusulas Definidas (DCG) mediante apropiados mecanismos de compilación (como por ejemplo la resolución-OLDT). Esta extensión mejora el tratamiento y resolución de las estructuras coordinadas y plantea por primera vez una implementación en Prolog de tal gramática, utilizando el algoritmo de evaluación semi-naive, que nos permite "congelar" el proceso de análisis de las oraciones para poderlo tratar metagramaticalmente.

Por último, hemos adaptado el mecanismo de comprobación semántica de tipos (IRSAS) a nuestro método, que permite resolver gran número de casos de coordinación y de elipsis.

Reconocimiento y generación integrada de la morfología del español: Una aplicación a la gestión de un diccionario de sinónimos y antónimos

TESIS DOCTORAL

Autor: José R. Pérez Aguiar (jperez@dis.ulpgc.es)

Director: Octavio Santana Suárez (osantana@dis.ulpgc.es)

Facultad de Informática. Universidad de Las Palmas de Gran Canaria

Fecha de presentación: 6 de Julio de 1996

Calificación: APTO CUM LAUDE POR UNANIMIDAD

<http://protos.dis.ulpgc>

Esta tesis aporta una solución al problema del reconocimiento y la generación de la morfología del español; aprovecha la estructura flexiva de la lengua para orientar su realización a datos más que a reglas —mejora la velocidad. Un pormenorizado estudio de las características morfológicas da lugar a un preprocesador que actúa sobre un diccionario de formas básicas y genera el universo de palabras flexionadas y derivadas que se estructura en forma de raíces y terminaciones —minimiza la ocupación. Se construye un procesador que es el encargado de reconocer y generar.

El análisis de una palabra se lleva a cabo mediante: a) segmentación en posibles pares raíz-terminación, b) localización de cada una de las raíces, c) decodificación de la terminación y d) construcción de la forma principal de la que proviene; el resultado también aporta su categoría gramatical. La

generación de una forma derivada o flexionada de una forma principal se logra gracias a: a) la detección de la raíz y b) la búsqueda de la terminación que corresponda a la forma que se desee conseguir.

Las flexiones consideradas son género y número para sustantivos y adjetivos, grado superlativo para los adjetivos y conjugación para los verbos que incluye pronombres enclíticos. Como derivación de sustantivos y adjetivos se consideran los sufijos apreciativos —aumentativos, diminutivos y despectivos. Se dispone además de un conjunto de relaciones morfológicas entre diferentes formas principales que faculta la obtención de formas derivadas que conllevan cambio de categoría gramatical —sustantivación, adjetivación, verbalización y adverbialización. El trabajo incluye un módulo de prefijación que permite incorporar prefijos a las palabras.

Se desarrolla una aplicación a la gestión de un diccionario de sinónimos y antónimos mediante una interfaz amigable que permite formas de exploración imaginativas y versátiles además de ofrecer la respuesta en forma concordante con la entrada.