

Proyecto Eurotra

J.C.Ruiz Anton

J.Vidal

PROYECTO EUROTRA

J.C. Ruiz Anton / J. Vidal

1. Aspectos políticos y organizativos.

Es sabido que a la vez que un inestimable bien cultural, la riqueza lingüística de la CEE es un obstáculo bastante importante para la eficacia administrativa de la Comunidad. Los gastos de traducción oscilan entre el 35% y el 65% en las diferentes instituciones europeas. Los retrasos en las traducciones obligan con mucha frecuencia a aplazar urgentes medidas políticas, porque sólo entran en vigor al publicarse en el Diario oficial en todas las lenguas simultáneamente. No es extraño, por tanto, que la CEE haya mostrado creciente interés por la traducción automática prácticamente desde sus comienzos (1).

Intentando hallar una solución al problema, la CEE sufraga (en cooperación con los gobiernos de los estados miembros) el programa EUROTRA, cuyo resultado ha de ser un prototipo de traducción automática entre las nueve lenguas oficiales de la Comunidad (portugués, español, francés, inglés, holandés, alemán, italiano, danés y griego moderno). Este prototipo cubrirá textos de tipo administrativo, técnico y científico, con unas 20,000 entradas léxicas por cada lengua.

Por decisión del Consejo de la CEE en Noviembre de 1982 (modificada en 1986, a consecuencia de la entrada de España y Portugal), se fija al programa una duración de siete años, con una dotación presupuestaria de algo más de 30 millones de ECUs.

El proyecto EUROTRA pretende asimismo difundir el conocimiento y las técnicas de la Lingüística Computacional en toda la Europa Comunitaria, creando infraestructuras en aquellos países (como España) con poca tradición en el campo. Para ello, el proyecto tiene una estructura descentralizada, con equipos de investigación en cada estado. Un equipo central de expertos diseña la arquitectura general del sistema, y suministra las especificaciones lingüísticas y del software usado por todos los grupos nacionales. Un comité de enlace se ocupa de la coordinación de los estos trabajos.

En España funcionan dos grupos de investigación, situados en Madrid y Barcelona. El director general del proyecto es el Profesor R. Cerdà, de la Universidad Central de Barcelona. El director del equipo de Madrid es el Profesor Marcos Marín, de la Universidad Autónoma.

2. Arquitectura del sistema.

EUROTRA es un modelo transferencial: el proceso de traducción se divide en tres fases: el análisis y la generación son puramente monolingües; además, entre cada par de lenguas opera un módulo de transferencia que se ocupa de traducir el léxico y aquellos fenómenos sintácticos irreductibles a alguno de los otros componentes. Un aspecto

fundamental es que en la traducción de un texto entre dos lenguas (digamos L y M), el análisis y la síntesis en la lengua L son por completo independientes de la lengua M.

El formalismo de EUROTRA es de una gran simplicidad y elegancia, y está inspirado en los modelos gramaticales de unificación (2). La hipótesis básica es que el proceso de traducción puede descomponerse en una serie de proyecciones entre niveles: morfología, sintaxis de constituyentes, sintaxis de relaciones, semántica.

De acuerdo con la filosofía del modelo, un nivel de representación está definido por un generador, que consiste en una serie finita de reglas de buena formación (constructores). El paso de un nivel de proyección a otro tiene que poder ser descrito por una serie de reglas (traductores).

Los generadores y los traductores se escriben en un lenguaje de representación especial (lenguaje de usuario), que recuerda el formalismo de la Gramática Transformacional clásica. Así, los constructores adoptan la forma de reglas de estructura de frase (si bien con anotaciones que las hacen sensibles al contexto); los traductores son muy similares a transformaciones.

Todo el software de EUROTRA está escrito en C-PROLOG, y corre sobre el sistema operativo UNIX. Actualmente, el sistema es razonablemente rápido (aunque ésta no es una idea obsesiva: no se olvide que se trata de una investigación pre-industrial), y hay planes de aumentar la velocidad de procesamiento del lenguaje de usuario implementando algunas de sus partes directamente en C (por ejemplo, el analizador sintáctico tipo Earley).

NOTAS:

(1) A principios de los 60, el Centro Europeo para el Tratamiento de la Información Científica (CETIS) llevó a cabo algunas investigaciones en este campo, y sufragó algunas fases del proyecto Georgetown (aplicadas a la traducción de textos científicos del ruso al inglés). Más recientemente, la CEE ha adquirido el sistema SYSTRAN para sus servicios de traducción, dedicando bastantes esfuerzos a mejorarlo y ampliarlo a otras lenguas no previstas originalmente.

(2) Como ilustración, la división entre sintaxis de constituyentes y sintaxis relacional es reminiscente de la Gramática Léxica Funcional (LFG), de J. Bresnan y R. Kaplan. Igualmente, la teoría de rasgos de EUROTRA está inspirada en la de la Gramática Sintagmática Generalizada (GPSG), de Gazdar y colaboradores.

Mesa redonda sobre proyectos de L.N. en España

Ponentes:

A. Benedí (Facultad de Informática, U.P.Valencia)
M.Castillo (Universidad de Malaga)
J. Llísterri (Facultat de Lletres, UAB)
E.Moragas (Bibliograf)
J.M. Pardo (Escuela Superior Ingenieros Telecominación,UPM)
J.M. Sopena (Facultad de Psicología,UB)
M.F. Verdejo (Facultat d'Informatica, UPC)

Moderador:

F. Ros (FUNDESCO)

Guion:

- Breve presentación de los trabajos realizados por su equipo: enfoque metodológico, resultados , recursos humanos y equipamiento utilizado.
- Situación comparativa, en términos cualitativos, de la situación en España con respecto a otros países: resultados, cobertura linguistica del idioma, métodos,...
- Situación comparativa en términos cuantitativos: cuántos profesionales/investigadores, realizaciones, oferta comercial,...
- Qué acciones de fomento serian interesantes emprender:
 - difusión de información
 - formación de investigadores: qué perfil y tipo de formación
 - coordinación de proyectos de investigación
 - colaboración industria-equipos de investigación