

Problemas de la traducción automática: la morfología

*Ulrike Tallowitz
C.E.L.E.-U.N.A.M.*

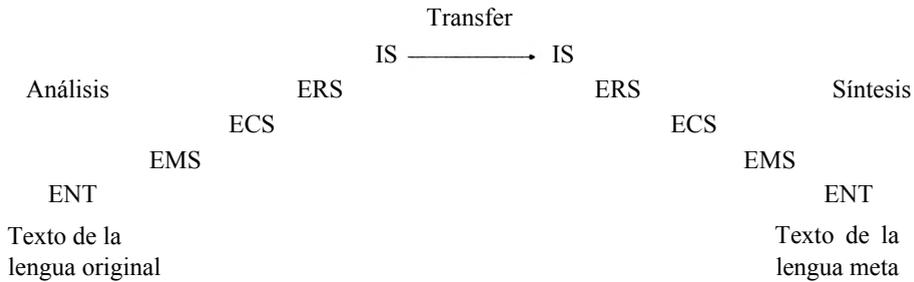
1. Introducción

En la traducción automática (TA) por computadora quedan todavía muchos problemas por resolver; uno de ellos es el manejo de la morfología. Para ejemplificar los problemas que se presentan en este contexto, se describirá el análisis morfológico en un sistema de traducción automática específico: el de EUROTRA, que está en proceso de desarrollo en la Comunidad Europea. Se presentará tanto el funcionamiento actual como algunos aspectos que todavía presentan problemas, proponiendo pasos para su solución.

2. El proceso de traducción

Se distinguen dos enfoques: el uso de un “lenguaje pivote”, y las “técnicas de transferencia”.

El primero trata de crear un lenguaje pivote que represente la estructura profunda de todas las lenguas como enlace de una lengua a la otra. Es como un “*tertium comparationis*”. Es decir, las estructuras de la lengua fuente se presentan en la forma de un lenguaje pivote, en un nivel muy abstracto, y de allí se generan las formas estructurales de la lengua meta. Este principio es muy ambicioso; en el fondo se trataría de describir los “universales” del lenguaje, y tal lenguaje “interlingual” tendría que comprender todos los fenómenos de todas las lenguas. En la práctica se usa más bien la técnica de “transferencia” de una lengua fuente a una lengua meta para concentrarse mejor en los matices sutiles de dos lenguas específicas. La hipótesis básica de esta técnica es que el proceso de traducción se puede describir como una sucesión de proyecciones, de manera que un texto pasa por un cierto número de representaciones gramaticales determinadas por criterios lingüísticos. Hay un nivel morfológico, un nivel de sintaxis de superficie, uno de sintaxis profunda, y por fin uno semántico-sintáctico. El proceso mismo se compone de tres fases: Análisis, Transferencia y Síntesis.



- ENT = EUROTRA Normalized Text
- EMS = EUROTRA Morphological Structure
- ECS = EUROTRA Constituent Structure
- ERS = EUROTRA Relational Structure
- IS = Interface Structure

(cf. Tallowitz 1990)

En un nivel de gramática profunda (IS), la transferencia de una lengua a otra se hace estructuralmente. Las estructuras básicas sintácticas de la lengua fuente se transfieren, con reglas complejas, a las estructuras de la lengua meta. Cada nivel de representación está definido por un **diccionario** y una **gramática**: un generador que consiste en un número finito de reglas estructurales. Los diferentes niveles están relacionados entre sí por **traductores**: un número finito de reglas que transforman un nivel en otro. El nivel IS utiliza descripciones de la gramática relacional (Perlmutter 1983;1984). El verbo está en el centro de la oración, de él dependen los complementos que se describen por su función en la oración. Los sistemas de traducción hoy en día funcionan con base en la oración. No toman en cuenta ni factores textuales ni sociolingüísticos, los que, según se sabe, determinan la comprensión del lenguaje en gran medida. Para considerar estos factores, sería necesario incorporar “conocimiento del mundo” formalizado, lo que no ha podido resolverse hasta la fecha.

3. La morfología

En lenguas como el alemán, el ruso o el japonés, la posición de las partes de la oración no informan sobre su función en ella. Mientras que en inglés la posición fija de los componentes de la oración nos indica si se trata de un sujeto o de complementos, en alemán un complemento directo, por ejemplo, puede igualmente colocarse al principio, al medio o al final de una oración. En este grupo de lenguas,

“con orden libre de palabras”, las terminaciones morfológicas son necesarias para reconocer la función de cada parte de la oración y el análisis correcto de ésta.

Hasta hace un año, el sistema EUROTRA no tenía un componente morfológico: el parsing (descomponer la oración en sus elementos) se hacía a partir de los límites de palabras, las cuales se buscaban posteriormente en el diccionario para su identificación. Por consiguiente, todas las formas posibles de cada palabra tenían que estar enlistadas en el diccionario: en el caso de los adjetivos alemanes una entrada podía comprender 216 formas por lexema (4 casos, 3 géneros, 3 tipos de declinación, 2 números, 3 formas de comparación). La situación era parecida en cuanto a los verbos (personas, número, tiempos verbales, modos, etc.) y los sustantivos (casos, géneros, número). Se trataba de diccionarios de “formas completas”. Por supuesto, esto no es económico en cuanto a espacio y tiempo, sin mencionar la falta de generalización lingüística.

Después de un trabajo teórico de investigación y un experimento (Carulla/Maas/Sánchez, 1990), se implementó para las lenguas española y alemana el módulo morfológico EMS (EUROTRA Morphological Structure). Con la parte inflexional, este módulo no muestra problemas, pero aparecen dificultades en la parte de la derivación. Aquí, por tratarse de un cambio de formas de palabras y de un cambio simultáneo de contenido semántico, surgen los siguientes problemas:

- Las derivaciones, aunque tengan la misma forma, no son unívocas.
- No se ha logrado todavía una descripción semántica del alcance requerido para esta tarea.

4. Ejemplificación

Se presentará aquí una selección de las reglas morfológicas implementadas. La muestra consiste en un fenómeno de derivación, o sea la nominalización deverbal, sobre todo la de “acción” y la “agentiva”.

Ejemplos:

	Verbo	Sustantivo
Nominalización de acción	vertreten (sustituir) automatisieren (automatizar)	die Vertretung (la sustitución) die Automation (la automatización)
Nominalización agentiva	lesen (leer) lehren (enseñar)	der Leser (el lector) der Lehrer (el maestro)

El nivel EMS (EUROTRA Morphological Structure): En el diccionario se encuentran, por ejemplo, las siguientes entradas para describir los verbos (se presenta una selección de los rasgos más pertinentes para el tema, a veces cambiando ligeramente los nombres para mayor claridad):

{cat = verbstem, lu = rechnen, string = rechn, n_action = ung, n_agent = er, ...}.

{cat = verbstem, lu = treten, string = tret, n_action = ung, n_agent = er, ...}.

{cat = verbstem, lu = demonstrieren, string = demonstr, n_action = ation,
n_agent = ant, ...}.

Explicación de los rasgos:

cat = categoría

verbstem = raíz del verbo

lu = lexical unit

string = parte de la palabra que se analiza

n action = posibilidad de derivar un sustantivo por nominalización de acción.

El valor indica qué sufijo se usa para tal nominalización. En los ejemplos pueden formarse rechn-ung (cuenta), ver-tret-ung (sustitución), demonstr-ation (demonstración, manifestación)

n agent = posibilidad de derivar un sustantivo por nominalización agentiva, con los sufijos -er (rechn-er, ver-tret-er) o -ant (demonstrant).

Además de los verbos, los morfemas mismos tienen su entrada en este diccionario.

Por ejemplo, para la nominalización de acción:

vertret-ung (sustitución)

{cat = morph, string = ung, lu = '#ung', gender = fem, fuge = s,
source = verb, target = noun, ...}.

Nominalización agentiva: p.ej. rechn-er (calculadora)

{ cat = morph, string = er, lu = '#er', gender = mas, fuge = null,
source = verb, target = noun, ...}.

En la gramática EMS hay reglas para la estructura de palabras complejas en las que se analizan las terminaciones infleccionales y derivacionales. Una regla para derivar un verbo de un sustantivo podría presentarse así:

{cat = nounstem, structure = derived, ...}
[{cat = verbstem, n_action = UNG, n_agent = ER, target = noun, ...},
{cat = morph, n_action = UNG, n_agent = ER,
source = verb, target = noun, ...}].

La regla anterior indica que un sustantivo (automation) se compone de la raíz de un verbo (automatisier-) más un morfema derivacional del tipo ung= (derivación

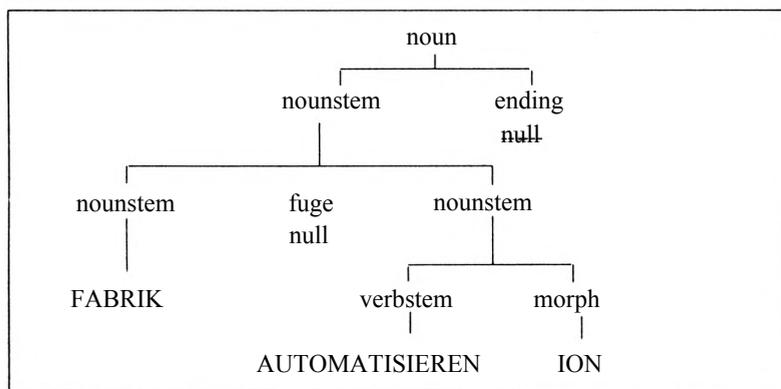
de acción), que en el ejemplo sería -ion. Esta regla posibilita el análisis de palabras como:

Schwimmer, Vertreter, etc.	(derivación agentiva)
Student, Demonstrant, etc.	(derivación agentiva)
Schaetzung, Einschätzung, etc.	(derivación de acción)
Demonstration, Restauration, etc.	(derivación de acción)

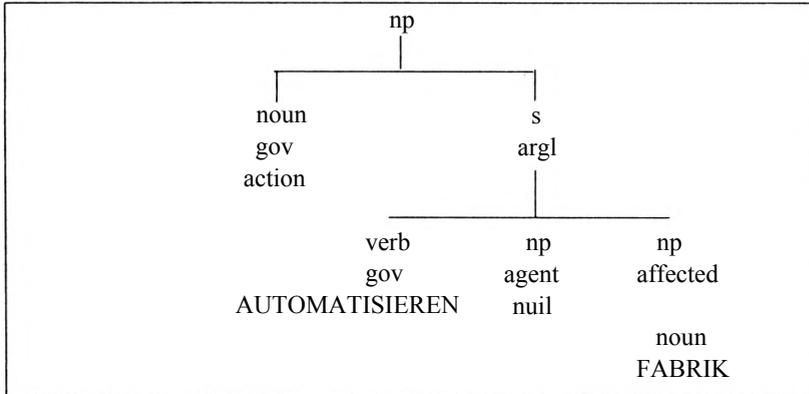
Hay otras reglas que reconocen derivaciones como infinitivos nominalizados (das Lesen - “la lectura”), derivaciones de estado (Schoenheit - “belleza”), sustantivos en -ik y -ie (Linguistik - “lingüística”, Biologie - “biología”), en -ismus (Automatismus - “automatismo”), o los que terminan en -er, pero que se derivan de otros sustantivos (Mathematiker - “matemático”, Frankfurter - “frankfurtense”).

Para demostrar visualmente este análisis, he aquí la estructura arborescente de un sustantivo derivado (y además compuesto) en EMS:

Fabrikautomation (automatización de fábricas):



Esta descripción morfo-sintáctica de las palabras forma parte del análisis sintáctico de la oración completa en los niveles de sintaxis superficial (ECS) y relacional (ERS). En el nivel IS, tanto la estructura de la oración como la de las palabras se describe semánticamente, es decir, con roles semánticos tal como “agent”, “affected”, “effected”, “processor”, etc. y rasgos semánticos como [+ -abstr], [+ -hum], [+ -count]. He aquí la estructura arborescente de la palabra “Fabrikautomation”:



Para realizar tal estructura, existen entradas en el diccionario para las raíces de las palabras (AUTOMATISIEREN, FABRIK) y unas categorías abstractas que indican el tipo de derivación, es decir, derivación de acción, de agente, o de cambio de estado:

{cat = verb, lu = automatisieren, argl = agent, arg2 = affected,...}.

{cat = noun, deriv = nominal, type = action,...}.

{cat = noun, deriv = nominal, type = agent,...}.

{cat = noun, deriv = nominal, type = State,...}.

Además de las entradas en el diccionario, hay reglas estructurales en el nivel IS. Estas reglas describen tanto las estructuras semánticas de las oraciones como de las palabras. Por ejemplo, las nominalizaciones de verbales tienen la misma estructura como una oración con el verbo correspondiente, es decir, la misma relación de dependencia:

{cat = s,...}

[{cat = verb, argl = ARG1, arg2 = ARG2,...},

{role = ARG1},

{role = ARG2}].

5. Problemas por resolver

5.1 Problemas lingüísticos

Dos de los problemas más difíciles para lograr una descripción adecuada de derivaciones son:

- a) los rasgos semánticos que adquiere la palabra derivada y
- b) qué derivaciones pueden o no realizarse

a) Acerca de los rasgos semánticos, veamos el siguiente ejemplo: el sustantivo *Leitung* (dirección), derivado por una nominalización de acción, del verbo *leiten* (dirigir), tendría los atributos [+ hum] y [H- abstr]. El sustantivo *Leiter* (director), derivado por una nominalización agentiva, los atributos [+ hum], [-abstr]. Por otra parte, el sustantivo *Zeiger* (manecilla), igualmente derivado por nominalización de agente, del verbo *zeigen* (enseñar), tendría los rasgos [-hum] y [-abstr], ya que este sustantivo no representa una “persona que hace algo”, sino un instrumento. Sin embargo, la regla para el morfema “-er” que aparece tanto en el sustantivo *Leiter* como en *Zeiger* no permite diferenciar entre ambos:

{cat = morph, lu = '#er',...,abstr = no, count = yes,...}.

La regla no tiene rasgos de valor semántico de “hum” (humano). De este tipo de derivación pueden crearse palabras correspondientes a los “seres humanos” (*Leiter*) o a las “cosas” (*Zeiger*). No hay forma, por el momento, de indicar cuándo una derivación agentiva resulta en una persona y cuándo en un objeto (instrumento). Esto depende del verbo del que se deriva el sustantivo. Para poder hacer la diferenciación habría que explicar qué tipo de verbos conduce a qué derivación. Como por el momento no hay una teoría semántica para definir en términos generales qué características presentan los verbos que conducen a sustantivos de persona/instrumento, es necesario indicar en el diccionario con cada verbo el valor semántico [+ hum] o [-hum] que va a tener su nominalización agentiva.

La entrada de un verbo con esta posibilidad de descripción se escribiría así:

{cat = verb, lu = leiten,...,type = agent-hum = yes,...}.

{cat = verb, lu = zeigen,...,type = agent-hum = no,...}.

b) Otro problema que queda por resolver es: cuáles derivaciones son posibles y cuáles no. Por ejemplo, a partir del verbo *faltren* (manejar), **Fahr-ung* (approx. *manej-encia) no es posible, pero ER-FAHR-UNG (experiencia) se puede derivar. Aquí intervienen razones históricas, de “lexicalización”, y del uso de la lengua, que aún no pueden formalizarse, por falta de reglas sistemáticas.

5.2 Problemas de traducción

El lingüista de la traducción automática se enfrenta a la tarea de elaborar un sistema semántico coherente que le permitirá definir las reglas del programa de computación. Debido a la característica lógica de las computadoras, mismas que no entienden supuestos socio-culturales, este sistema tiene que ser explícito, coherente y libre de contradicciones. Entonces, se plantea la siguiente cuestión:

¿Hay razones lingüísticas semánticas formalizables, que indiquen la forma en que la derivación se realiza en cada caso?

En primer lugar, habrá que construir un esquema de los matices semánticos que pueden tomar los distintos tipos de derivación, para después encontrar los rasgos distintivos en cada caso. En segundo lugar, se necesita un análisis contrastivo español- alemán, para formular reglas de transferencia, porque las diferencias en el tipo de derivación en las dos lenguas y, sobre todo, en el diferente valor semántico que pueden tomar las derivaciones, presentan problemas serios de traducción.

En el campo de la descripción semántica hay trabajos muy importantes, como el de la derivación en español de Elisabeth Beniers, quien demuestra por ejemplo los diferentes valores semánticos que puede tomar la agentivización en -encia, -dor y -ero. La autora concluye que las derivaciones de ninguna manera son unívocas (Beniers 1990, p.105).

Uno de los ejemplos es la palabra “dirección”, derivada de “dirigir”, y que se contrasta aquí con su equivalente en alemán:

Español

Alemán

1. Nominalización de acción

dirección (de un proyecto)
del verbo dirigir

Leitung
del verbo “leiten”
type = action

2. Nominalización agentiva

dirección (la gente que dirige)

Leitung
(p.ej.en Schulleitung)
type = agent

Hay además la palabra derivada del latín:
“Direktion”

En alemán el verbo subyacente “dirigieren”
se refiere sólo a dirigir a una orquesta, con
la correspondiente nominaliza-
ción “Dirigent”.

3. En el caso de “dirección” como lugar, la situación también es distinta para el español y el alemán:

dirección (p.ej., las oñcinas del director)	Direktion *Leitung En alemán la derivación del verbo “leiten” no se usa en este contexto, hay que usar la derivación del latín.
---------------------------------------------------	---------------------------------------------------------------------------------------------------------------------------------------------

Concluyendo, se puede decir que no es posible adoptar la fórmula “el sustantivo agentivo en español de un verbo determinado se transfiere en el sustantivo agentivo de este mismo verbo en alemán”.

Con el proyecto de investigación que realizamos actualmente en el Centro de Enseñanza de Lenguas Extranjeras y en el Instituto de Investigaciones Filológicas de la UNAM * queremos hacer una contribución a sistemas semánticos para la descripción de derivaciones postverbales como el del sistema de traducción automática EUROTRA.

BIBLIOGRAFIA

- BENIERS, E. (1990) “La derivación en la gramática y en el diccionario”. ELA, Año 8, Núm. 11. México: CELE-UNAM, pp. 102-116.
- CARULLA, M./MAAS, D./SANCHEZ, F. (1990) **Report on the Experimental Implementation of Morphology**. Luxembourg (manuscrito).
- FLEISCHER, W. (1982) **Wortbildung der deutschen Gegenwartssprache**. Tübingen: Max Niemeyer.
- HUTCHINS, W.J. (1986) **Machine Translation - Past, Present, Future. Computers and their Applications**. Chichester: Ellis Horwood Publ.
- PERLMUTTER, D.M. (ed.) (1983/84) **Studies in Relational Grammar**. Chicago: Univ. of Chicago Press, 2 vols. (Vol. 2 coedited by Carol G. Rosen).
- TALLOWITZ, U. (1990) “El proyecto de traducción automática EUROTRA.” ELA, Año 8, Núm. 12. México: CELE-UNAM, pp. 113-125.
- TOUSSAINT, G.T. (ed.) (1988) **Computational morphology: a computational geometric approach to the analysis of form**. Amsterdam: North-Holland.
- ZELINSKY-WIBBELT, C. (1989) **Machine Translation Based on Cognitive Linguistics: What Lexical Semantics Contributes to the Semantic Unity of a Sentence**. Saarbrücken: EUROTRA-D Working Papers, No. 16.
- ZWICKY, A. / WALLACE, R. E. (eds.) (1984) **Papers on morphology**. Columbus, Ohio: Dept. of Linguistics.

* Tallowitz, U./E. Beniers: “La morfología en la traducción automática”, CELE-IIF, UNAM, México. Proyecto en curso, en colaboración con el grupo alemán de EUROTRA, en Saarbrücken,