

Tecnologías de la Traducción: Actividades opcionales *

Mikel L. Forcada, Felipe Sánchez Martínez

Departament de Llenguatges i Sistemes Informàtics
Universitat d'Alacant

Curso 2011–2012, segundo cuatrimestre

última actualización: 5 de febrero de 2012

Índice

1. Textos y formatos	2
2. Estrategias y limitaciones de la traducción automática	2
3. Evaluación de la traducción automática	4

Notas

- Este documento contiene actividades opcionales que podéis hacer y entregar a través del Campus Virtual (opción Evaluación→Controles) antes de la fecha del examen final.
- Para escribir la respuesta a estas actividades tendréis que usar la plantilla `tt-plantilla-act-aula-es.odt` que os podéis descargar de la sección de materiales del Campus Virtual.

* © 2012 Universitat d'Alacant. Este material puede ser distribuido, copiado y exhibido si los nombres de los autores se muestran en los créditos. Las obras derivadas tienen que distribuirse bajo los mismos términos de licencia que el trabajo original. Más detalles: <http://creativecommons.org/licenses/by-sa/3.0/deed.es>. Podéis pedir los fuentes LaTeX a los autores.

1. Textos y formatos

Act. 1.1 Cuando alguien quiere utilizar el ordenador para hacer una traducción, en algún punto tiene que manipular o generar un *texto informatizado*. Indicad cómo se hace para generar este texto y qué herramientas informáticas —software, hardware— se usan (procurad dar nombres y definiciones tan precisas como podáis). No olvidéis que los textos se pueden generar también sin tener que teclearlos.

2. Estrategias y limitaciones de la traducción automática

Ambigüedad

Act. 2.1 Indicad qué hace que los textos ambiguos siguientes sean especialmente difíciles de tratar por un sistema de traducción automática, aunque la mayoría de las personas puedan elegir la interpretación correcta:

1. (en) “The soldiers shot at the women and I saw them fall” (“Los soldados dispararon a las mujeres y los vi caer” — los soldados— o “...y las vi caer” — las mujeres—).
2. “Cogió las llaves de la silla” (¿hay llaves que abran sillas? ¿tienen cerradura las sillas?)
3. “Las asesorías fiscales ayudan mucho cuando se tienen que preparar las declaraciones de la renta pero aun así las odio”.
4. “¿Tuviste relaciones con tu marido antes de casarte con él?” “Yo no, ¿y tú?” “Yo sí, pero no sabía que acabaría casándose contigo”.

Subact. 2.1.1 ¿Qué tipo de ambigüedad representa cada caso?

Subact. 2.1.2 ¿Cómo podría un sistema de traducción automática intentar resolver la ambigüedad en cada caso? ¿Qué información usaría? ¿Cómo la obtendría?

Traducción automática basada en reglas

Act. 2.2 Un caso de ambigüedad interesante sucede cuando se traduce una frase que es ambigua en la lengua origen y resulta que las traducciones correspondientes a cada una de las posibles interpretaciones son idénticas. Esto se denomina *free ride* (pase gratuito). ¿Podrías poner algún ejemplo?

Act. 2.3 Los sistemas de traducción automática entre dos lenguas con sintaxis muy diferente necesitan hacer reordenamientos. Imaginad un sistema de traducción automática que traduce sintagmas nominales del catalán al vasco y usa reglas para reordenar secuencias de categorías gramaticales que se aplican de izquierda a derecha, reordenando la secuencia más larga detectada y sin operar dos veces sobre la misma palabra.

Fijaos en los ejemplos siguientes, donde se indica la traducción producida por el programa y, donde es incorrecta (*), la traducción correcta:¹

1. *La casa*: etxea
2. *La casa vermella*: etxe gorria
3. *La casa de la dona*: emakumearen etxea
4. *La casa vermella de la dona*: emakumearen etxe gorria
5. *La casa de la dona jove*: *emakumearen etxea gazte (corr.: emakume gaztearen etxea)
6. *La casa vermella de la dona jove*: *emakumearen etxe gorria gazte (corr.: emakume gaztearen etxe gorria)

¿Cuáles son las reglas activas en cada frase? ¿Por qué se producen las traducciones incorrectas? Justificad vuestras respuestas.

Act. 2.4 Fijaos en el ejemplo de la actividad anterior. Las limitaciones observadas son las de un sistema que no hace análisis sintáctico y que, por lo tanto, es incapaz de identificar sintagmas y manipularlos como tales. La solución consiste en hacer el análisis sintáctico y aplicar

¹Vocabulario: *emakume* n, *dona*; *etxe* n, *casa*; *gazte* adj, *jove*; *gorri* adj, *vermell*, -a; -a art, *el*, *la*; -ren gen, *marca de genitivo* (de).

a las frases analizadas reglas que transforman la sintaxis. Por ejemplo, las nuevas reglas para los sintagmas nominales se podrían escribir:²

$$R_1: \text{tr}([_{SN} \text{ } SN_1 \text{ } [_{SP} \text{ de } SN_2 \text{ }] \text{ }]) = [_{SN} [_{SP} \text{ tr}(SN_2) \text{ gen }] \text{ tr}(SN_1) \text{ }]$$

$$R_2: \text{tr}([_{SN} \text{ art } n \text{ }]) = [_{SN} \text{ tr}(n) \text{ tr}(\text{art}) \text{ }]$$

$$R_3: \text{tr}([_{SN} \text{ art } n \text{ adj } \text{ }]) = [_{SN} \text{ tr}(n) \text{ tr}(\text{adj}) \text{ tr}(\text{art}) \text{ }]$$

donde $\text{tr}(\dots)$ representa “la traducción de ...” y los corchetes $[_X \dots]$ indican la estructura sintáctica producida por el analizador o la enviada al generador.

Subact. 2.4.1 En el nuevo modelo, ¿cuáles serían las traducciones automáticas al vasco de las frases incorrectas del ejemplo anterior?

Subact. 2.4.2 ¿Son correctas?

Subact. 2.4.3 Si en vasco *fill n* (hijo) se dice *seme*, ¿cuáles serían las traducciones de “la casa vermella del fill jove de la dona” i “la casa del fill de la dona jove”?

Subact. 2.4.4 ¿Qué reglas se han aplicado? (fijaos en que es posible que en estas frases tengáis que aplicar más de una vez la misma regla).

Act. 2.5 Indicad, si podéis, algunos problemas que todavía no resuelve el modelo de transferencia sintáctica propuesto en la actividad anterior y que se tendrían que resolver en un modelo más avanzado de sistema de traducción automática. Podéis usar las oraciones siguientes como inspiración:

1. (en) I like roasted peanuts → (ca) *Jo agrade cacaus torrats* (pero: I buy roasted peanuts → *Jo compre cacaus torrats*).
2. (ca) Porta les claus de la porta verda → (en) *?Bring the keys from the green door*
3. (en) Cleaning fluids can be toxic → (ca) *?Netejar líquids pot ser tòxic*
4. (en) The ship sank → (ca) **El vaixell afonà* (pero: We sank the ship → *[Nosaltres] Vam afonar el vaixell*).

²Usando la representación clásica con corchetes $[\dots]$ en lugar de árboles.

Traducción automática estadística

Act. 2.6 Los sistemas de traducción automática estadística aprenden a traducir a partir de bitextos (textos paralelos que son traducción mutua). ¿Cómo creéis que se puede aprender un diccionario bilingüe a partir del siguiente bitexto?

the house		la casa
the white house		la casa blanca
the flower		la flor

Act. 2.7 A la hora de traducir, hacer uso siempre de la traducción más probable dada por un diccionario bilingüe probabilístico como el obtenido en la actividad anterior es equivalente a la traducción palabra por palabra (“modelo 0”), que como ya hemos visto no proporciona buenas traducciones:

1. (es) *El oso apareció tarde* → (en) *The I dare appeared afternoon* (correcta: *The bear appeared later*)
2. (en) *The computer expert's large desk is full* → (es) *El ordenador experto grande escritorio es lleno* (correcta: *El escritorio grande del experto en ordenadores está lleno*)
3. (es) *Menos mal que sólo sesenta y cinco personas murieron* → (en) *Less wrong that only sixty and five persons died* (correcta: *Fortunately, less than sixty five people died*)

Probad a introducir algunas de estas traducciones (o mejor, segmentos de texto más pequeños) entre comillas en un buscador de Internet y veréis cuántos resultados encuentra. Probad después con las traducciones corregidas (también con segmentos más pequeños de estas).

¿Se os ocurre alguna manera de hacer uso de esta información (el número de resultados encontrados) y de un diccionario bilingüe probabilístico para que no siempre se use la traducción más probable y mejoren las traducciones? Dad tanto detalle como podáis.

Act. 2.8 Los sistemas que habéis esbozado en la actividad anterior trabajan traduciendo palabras

sueltas. ¿Cuál creéis que es el principal problema de esta aproximación a la traducción? Como lo solucionaríais? Pensad en la traducción de segmentos de más de una palabra como los *phrasal verbs* o las colocaciones.

3. Evaluación de la traducción automática

Act. 3.1 Considerad el concepto de *evaluación predictiva* de la calidad de las traducciones, que podemos definir como sigue: un proceso de evaluación organizado de tal manera que, con los datos recogidos, podamos predecir razonablemente el comportamiento de un sistema de traducción automática en situaciones nuevas respecto de las que se han evaluado. Explicad como habría que organizar un proceso de evaluación predictiva y qué conocimientos son necesarios.