

TT2: Sistema de Traducción Asistida

TransType2 es un proyecto enmarcado en el ámbito de la traducción asistida. El objetivo del mismo es proporcionar traducciones de alta calidad combinando la experiencia de traductores humanos con sistemas de traducción automática.

Introducción

La tecnología disponible en la actualidad sólo permite abordar adecuadamente aplicaciones de traducción automática en dominios limitados. Cuando el dominio de aplicación se generaliza (o incluso se extiende a una lengua completa), no es posible obtener traducciones automáticas con la suficiente calidad (al menos en la actualidad). La traducción asistida puede ser de utilidad en estos casos.

Existen varias aproximaciones para afrontar esta tarea. La adoptada en este proyecto se denomina Traducción Predictiva Interactiva (TPI). En ella se engloban técnicas de traducción asistida en las que el sistema trata de asistir a un traductor humano en su labor.

El proyecto TransType2 (TT2), que se está llevando a cabo en el ITI, tiene como objetivo el desarrollo de un sistema de traducción asistida que permita hacer frente a la creciente demanda de traducciones de alta calidad. El sistema estará capacitado para realizar traducciones entre inglés, francés, español y alemán.

La solución propuesta por TT2 se basa en la incorporación de la Traducción Automática dentro de un entorno de TPI. De esta manera, el sistema combina las ventajas de dos paradigmas: por una parte la traducción asistida, en la cual el traductor humano asegura una salida de alta calidad. Por otra parte, la traducción automática, en la que la máquina proporciona la eficiencia necesaria para lograr un aumento de productividad.

Para abordar la tarea de la traducción automática, en TT2 se emplean transductores estocásticos de estados finitos, los cuales han demostrado su adecuación para la traducción automática en aplicaciones de dominio restringido. Éstos se caracterizan por su simplicidad y por la

El proyecto TT2 combina las ventajas de la traducción asistida y la automática

posibilidad de inferir modelos automáticamente a partir de corpora de entrenamiento bilingües. Además, la utilización de estos modelos está justificada por su eficiencia en la traducción. Por otra parte, se pueden emplear técnicas híbridas de estados finitos y traducción estadística para producir transductores más precisos. En este caso el aprendizaje se basa en el empleo de pares de entrenamiento alineados a nivel de palabra mediante técnicas estadísticas.

Proceso de traducción

Dentro del proceso de traducción encontramos dos fases claramente diferenciadas. La primera consiste en obtener el transductor a partir del cual se generarán las traducciones. En la segunda fase se obtiene la traducción de una frase determinada aplicando un proceso de búsqueda sobre el transductor. A continuación veremos con más detalle cada una de ellas.

1. Inferencia de los transductores

Para obtener el transductor, es necesario disponer de un conjunto de ejemplos compuestos de posibles frases y sus correspondientes traducciones, a partir de las que se tratará de generalizar el proceso de traducción. Este procedimiento se lleva a cabo mediante la técnica GIATI (*Grammatical Inference and Alignments for Transducer Inference*).

A continuación se muestra un ejemplo de inferencia en el que se dispone del siguiente conjunto de frases que corresponden a traducciones de inglés a español:

Es posible obtener traducciones de calidad tecleando solo un 15% de los caracteres

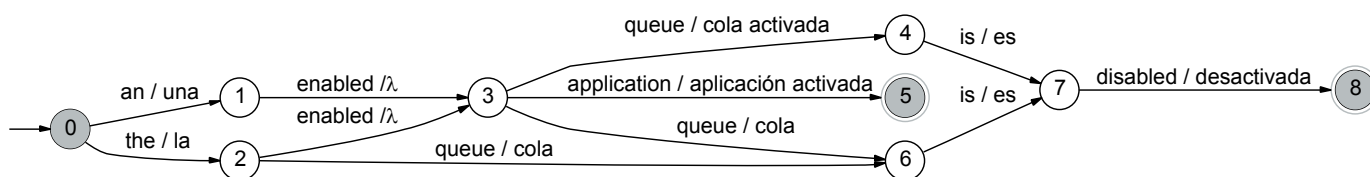


Figura 1: Ejemplo de transductor para la traducción inglés-español.

TT2: Sistema de Traducción Asistida

Ejemplo 1

an enabled queue is disabled → una cola activada es desactivada
 the enabled application → la aplicación activada
 the queue is disabled → la cola es desactivada

Aplicando GIATI sobre este conjunto de muestras se obtiene el transductor que se puede ver en la Figura 1.

Un transductor está formado por un conjunto de estados y transiciones que los unen entre sí. Cada una de estas transiciones, que está etiquetada con una palabra de la frase de entrada, lleva asociada una posible traducción y un valor que indica la probabilidad de dicha arista.

2. Búsqueda de la traducción

Una vez generado el transductor, el proceso de traducción consiste en encontrar el camino más probable (mediante

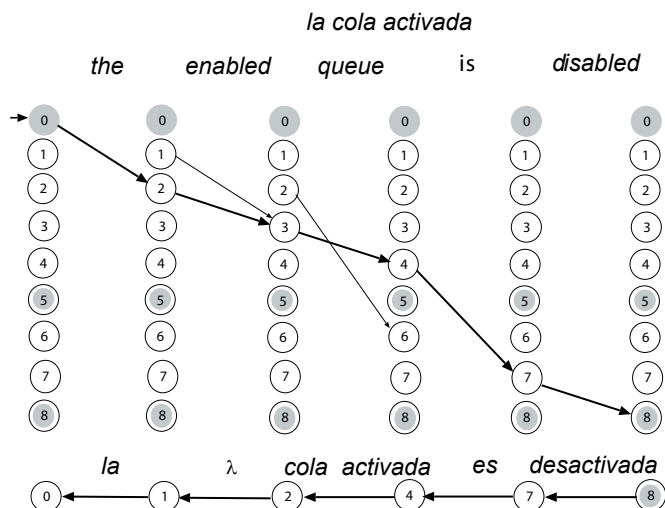
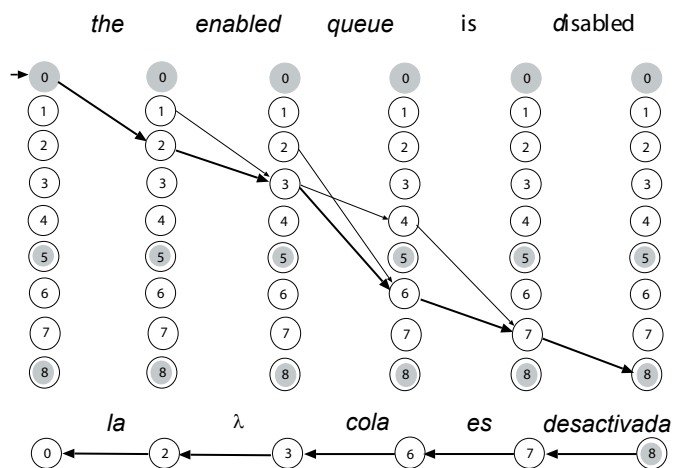


Figura 2: Búsqueda de la mejor traducción.

el algoritmo de Viterbi) compatible con la frase a traducir dentro del transductor.

Es importante resaltar que las traducciones obtenidas mediante este sistema no garantizan un nivel óptimo de calidad. En este punto, el traductor humano debe intervenir para evaluar y corregir, si procede, dichas traducciones. Concretamente, el sistema proporcionará las n traducciones más probables (siendo n un número fijado de antemano por el usuario) de entre las cuales el usuario seleccionará aquélla que considere más adecuada. A continuación, el usuario procederá a examinar secuencialmente la traducción para determinar cuál es la primera palabra incorrecta y la corregirá, formando así un prefijo que el sistema de traducción deberá completar. Este proceso se repetirá hasta obtener la traducción deseada.

Ejemplo 2

Un ejemplo de este procedimiento, basado en el transductor representado en la figura anterior se presenta en la Figura 2.

Dada la frase fuente en inglés 'the enabled queue is disabled', se aplica el algoritmo de búsqueda obteniéndose como traducción más probable "la cola es desactivada".

La traducción proporcionada por el sistema tiene en común con la traducción correcta el prefijo 'la cola'. Por tanto, el usuario debería introducir la palabra 'activada' tras el prefijo. Una vez insertado el prefijo adecuado se repite el proceso obteniendo así la traducción correcta 'la cola activada es desactivada'.

El prototipo de TT2

Este procedimiento de traducción se plasma en un prototipo (figura 3) compuesto por el motor de traducción asistida empotrado en un interfaz java.

En las siguientes líneas se describirá el funcionamiento del mismo. En primer lugar, se seleccionará el par de idiomas entre los que se desea realizar la traducción (así se cargará el transductor correspondiente, mediante el cual se llevará a cabo el proceso de búsqueda de las traducciones) y un fichero que contenga el conjunto de frases a traducir. Una vez hecha esta selección, el sistema estará listo para ser utilizado. De todas las frases disponibles, que aparecen en el panel izquierdo de la figura, habrá que elegir aquella a traducir. Seguidamente el sistema proporcionará las n traducciones más probables (en este caso se ha ajustado el parámetro n a 2). Estas traducciones aparecen en el panel derecho de la figura en un desplegable.

De entre todas las traducciones, el usuario deberá quedarse con aquella que considere más adecuada pudiendo refinarla, situando el cursor al final del fragmento de traducción considerado más adecuado. Tras la inserción de

TT2: Sistema de Traducción Asistida

un nuevo carácter o palabra, el sistema procederá automáticamente a mostrar una nueva sugerencia.

En principio, la interacción con el prototipo se realiza con métodos tradicionales, es decir, utilizando periféricos como el teclado y el ratón. No obstante, en este proyecto se incorpora también la voz para hacer más cómoda la interacción con el sistema.

Para finalizar este apartado, se presenta un ejemplo completo de funcionamiento del proceso de búsqueda interactivo llevado a cabo en el prototipo.

Ejemplo 3

En la siguiente secuencia se analiza cómo se obtiene la traducción al español para una frase dada en inglés.

Frase fuente: *It also contains a section to help users of previous software versions adapt more quickly to the new software.*

Hipótesis 0: Se se para ayudar a los usuarios de versiones anteriores del software a que se adapten más rápidamente a este nuevo software.

Prefijo 0: T

Hipótesis 1: También se ofrece una sección para ayudar a los usuarios de versiones anteriores del software a que se adapten más rápidamente a este nuevo software.

Prefijo 1: También c

Hipótesis 2: También contiene una sección para ayudar a los usuarios de versiones anteriores del software a que se adapten más rápidamente a este nuevo software.

Hipótesis final: También contiene una sección para ayudar a los usuarios de versiones anteriores del software a que se adapten más rápidamente a este nuevo software.

A partir de la frase de entrada, el sistema proporciona una primera traducción (hipótesis 0). Esta hipótesis es claramente incorrecta y, por tanto, el usuario cambia el prefijo "Se" por la letra "T" (prefijo 0) que es la primera letra de la traducción que él cree idónea. El sistema producirá a continuación una nueva predicción. La nueva hipótesis (hipótesis 1) todavía no puede tomarse como correcta y el usuario decide cambiar el prefijo por uno más completo (prefijo 1). Finalmente, el sistema es capaz de devolver la traducción adecuada (hipótesis final).

Como se ha podido observar en este ejemplo, efectuando sólo dos cambios en la hipótesis inicial ofrecida por el sistema, se consigue una traducción correcta para la frase de entrada. Con tan sólo tres pulsaciones de tecla (dos modificaciones y una validación final) se ha llegado a la

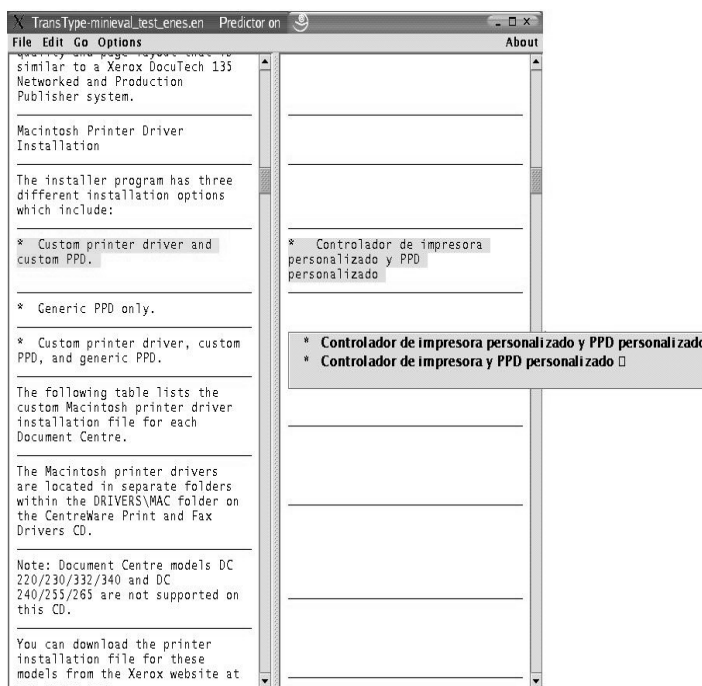


Figura 3: Prototipo TT2.

traducción deseada. De no haber empleado este sistema, habrían sido necesarias 148. En este caso, se muestran con claridad las ventajas que supone utilizar nuestro sistema.

El usuario refina la traducción más adecuada, de entre las proporcionadas por el sistema.

Se han llevado a cabo múltiples experimentos de los cuales puede concluirse que es posible obtener traducciones de calidad tecleando sólo un 15% de los caracteres que forman la traducción.

Autores: Elsa Cubel, Antonio Lagarda
Más información: grftl@iti.upv.es