

INCORPORACIÓN DE INFORMACIÓN SINTÁCTICO-SEMÁNTICA EN LA TRADUCCIÓN DE VOZ A LENGUA DE SIGNOS

B. Gallo, R. San-Segundo, J.M. Lucas, R. Barra, F. Fernández, L.F. D'Haro
 Grupo de Tecnología del Habla. Universidad Politécnica de Madrid.
 E.T.S.I. Telecomunicación. Ciudad Universitaria SN 28040 Madrid
 lapiz@die.upm.es

RESUMEN

Este artículo presenta un conjunto de experimentos para evaluar la mejora obtenida cuando se incorpora información sintáctico-semántica en la traducción estadística de voz a lengua de signos. La traducción se realiza utilizando dos alternativas tecnológicas: la primera basada en modelos de subsecuencias de palabras y la segunda basada en traductores de estados finitos ("FST"). En cuanto a la evaluación de dichos resultados, se utilizan varias métricas, como WER (tasa de error de palabras), BLEU y NIST. Las pruebas realizadas incluyen experimentos con las frases de referencia en castellano y Lengua de Signos y con frases obtenidas del reconocedor de voz. Para evaluar la mejora obtenida se muestran los resultados con y sin información sintáctico-semántica. Los mejores resultados se obtuvieron con la solución de traductores de estados finitos con unas tasas de error de 26,06% para las frases de referencia y de 33,01% para las salidas del reconocedor cuando se incorpora información sintáctico-semántica.

1. INTRODUCCIÓN

Con la realización de este trabajo se pretende evaluar la incorporación de información sintáctico-semántica en una plataforma de traducción capaz de transformar, en base a un conjunto de modelos probabilísticos, frases de castellano a Lengua de Signos Española (LSE). Con estos experimentos se pretende mejorar una herramienta de traducción muy útil para las personas sordas puesto que el coste de un intérprete signante (que conoce la Lengua de Signos) es muy elevado. A la vez supone una aportación importante para las personas sordas prelocutivas (aquellas que se quedaron sordas antes de poder hablar), ya que su capacidad de comprensión del castellano escrito es muy inferior al resto.

Este artículo se centra en la mejora del módulo de traducción basado en métodos estadísticos. Estos métodos son un paradigma de traducción automática donde se generan traducciones en base a modelos estadísticos y de teoría de la información cuyos parámetros se obtienen del análisis de corpus de textos bilingües (documentos que constituyen la base de datos

con pares de frases castellano-LSE). Con estos experimentos se pretende mejorar los resultados anteriormente presentados en [1].

Para la realización de los experimentos se dispuso de unos textos en castellano y en LSE con frases típicas que un funcionario de la administración pronuncia al atender a una persona en el servicio de solicitud o renovación del DNI.

2. ARQUITECTURA DEL SISTEMA

El sistema completo está formado por tres módulos: el del reconocedor de voz, el módulo de traducción estadística y finalmente, la representación de los signos mediante un avatar. Se muestra a continuación el diagrama del sistema:

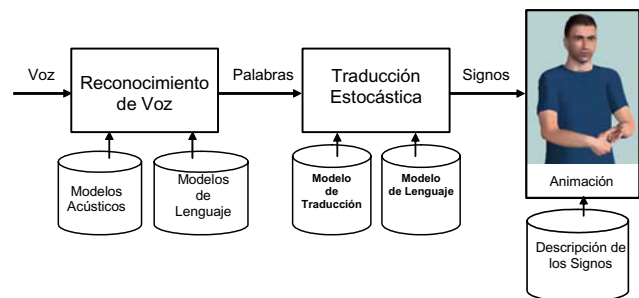


Figura 1. Arquitectura completa del sistema

El reconocedor del habla realiza la conversión del lenguaje natural (habla continua) a una secuencia de palabras basándose en un modelo del lenguaje y varios modelos acústicos.

En segundo lugar, el módulo de traducción estadística consiste en un algoritmo de búsqueda dinámica que utiliza un modelo estadístico para obtener la mejor secuencia de signos resultado de la traducción de una secuencia de palabras obtenidas del reconocedor de voz. Este modelo integra información de dos tipos de probabilidades: la probabilidad de traducción, que recoge información sobre qué palabras se traducen por qué signos y la probabilidad de secuencia de signos, que aporta información sobre qué secuencias de signos son más probables en la LSE.

El último módulo corresponde al avatar en 3D, que se encarga de la representación de los signos

provenientes de la traducción estadística. El avatar utilizado es “VGuido” del proyecto eSIGN [2].

3. TRADUCCIÓN BASADA EN SUBSECUENCIAS DE PALABRAS

La traducción estadística basada en modelos de subsecuencias (o subfrases) requiere la obtención de un modelo de traducción a partir del alineamiento entre las palabras de las lenguas origen y destino utilizando un corpus paralelo. Después del alineamiento de palabras se extraen y puntúan las subsecuencias de palabras que formarán el modelo de traducción. Además, es necesario generar un modelo de lenguaje de la lengua destino. En la fase de traducción, dada una frase de entrada se obtiene la secuencia de signos, que luego se evalúa para calcular los aciertos y fallos en la traducción. La arquitectura completa se muestra en la figura 2.

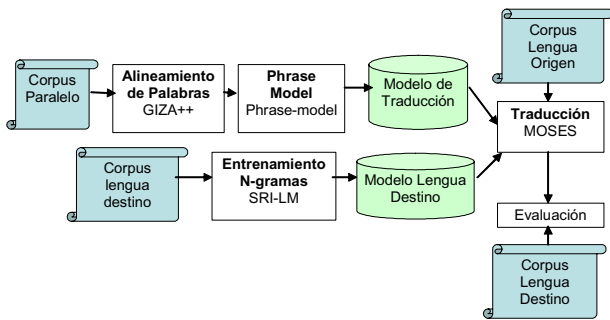


Figura 2. Traducción basada en subsecuencias

3.1 Generación de modelos

En primer lugar debe crearse el Modelo de Lenguaje de la lengua destino y el Modelo de Traducción (a partir de un corpus paralelo tanto en lengua origen (LO) como destino (LD)). El problema de la traducción se centra en conocer la probabilidad $p(d|o)$ de que una cadena o de LO genere una cadena d en LD. Estas probabilidades se calculan utilizando técnicas de estimación de parámetros a partir del corpus paralelo. Aplicando el Teorema de Bayes a $p(d|o)$ esta probabilidad se representa como el producto $p(o|d) \cdot p(d)$, donde el Modelo de Traducción $p(o|d)$ es la probabilidad de que la cadena origen se traduzca por la cadena destino, y el Modelo de Lenguaje $p(d)$ es la probabilidad de ver aquella cadena origen.

Para la creación del Modelo de Lenguaje, se utiliza la herramienta SRILM [3], que realiza la estimación de los modelos de lenguaje tipo N-grama (en la que la probabilidad de una palabra depende de las N anteriores), a partir del corpus de entrenamiento. La generación de los Modelos de Traducción se hace mediante una traducción basada en subfrases. Para esto la herramienta utilizada es el GIZA++ [4], que permite obtener los alineamientos entre palabras de la lengua origen y palabras de la lengua destino, y un módulo de

modelos de subfrases a partir de estos alineamientos. Para esto se necesita un corpus paralelo. Los pasos para la generación de los modelos son:

1. Obtención del alineamiento entre palabras: a partir de los dos textos en castellano y LSE se identifican qué palabras de uno se alinean con los signos de LSE. El alineamiento se calcula en ambos sentidos: palabras-signos y signos-palabras.

2. Cálculo de una tabla de traducción léxica: se calcula a partir del alineamiento, obteniendo los valores de $w(d|o)$ y su inversa $w(o|d)$ para todos los pares de palabras.

3. Extracción de subsecuencias de palabras: se recopilan todos los pares de subsecuencias que sean consistentes con el alineamiento.

4. Cálculo de las probabilidades de traducción de cada subsecuencia.

3.2 Traducción

Para realizar el proceso de traducción se combinan los modelos generados en la fase anterior de entrenamiento mediante una combinación lineal de probabilidades cuyos pesos se deben ajustar. Este proceso de ajuste consiste en la ejecución iterativa del traductor Moses [5] sobre un conjunto de validación. En cada iteración se van modificando los pesos con el objetivo de maximizar los resultados de BLEU sobre ese conjunto de validación. Finalmente, y utilizando un nuevo conjunto de test se evalúa finalmente el sistema. No se realiza ningún tipo de reordenamiento previo a la fase de traducción.

4. TRADUCCIÓN BASADA EN TRADUCTORES DE ESTADOS FINITOS

Los traductores de estados finitos (“FST”) parten también de un corpus paralelo de entrenamiento y, usando métodos de alineamiento basados en GIZA++, generan un conjunto de cadenas a partir de las cuales se puede inferir una gramática racional. Esta gramática se convierte en un FST caracterizado por su topología y distribuciones de probabilidad, características aprendidas con el programa GIATI [6]. En la figura 3 se muestra la arquitectura de esta solución.

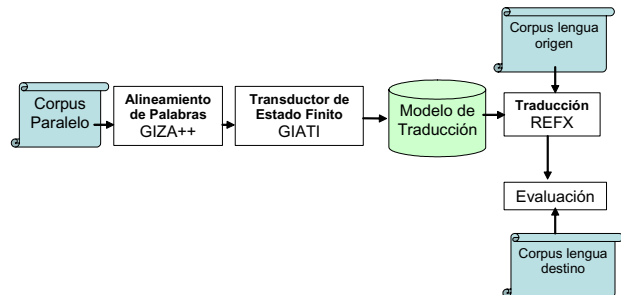


Figura 3. Traducción basada en Traductores de Estados Finitos

Las fases para la generación de los modelos son:

1. Alineamiento con GIZA++ a nivel de palabras. Igual que la primera fase del sistema de traducción basado en subsecuencias.

2. Transformación de pares de entrenamiento a frases. Se construye ahora un corpus extendido a partir de cada uno de los pares de subsecuencias de entrenamiento y sus correspondientes alineamientos obtenidos con GIZA++: se asignarán por tanto palabras de LO a su correspondiente palabra en LD gracias a su alineamiento. Se muestra a continuación un ejemplo de pares castellano / LSE y su alineamiento:

- el denei es obligatorio desde los catorce años #
DNI(2) SE-LLAMA(3) OBLIGATORIO(4)
DESDE(5) CATORCE(7) PLURAL(6) AÑO(8)
EDAD(8)
- el denei es el documento oficial #
DNI(2) SE-LLAMA(3) DOCUMENTO(5)
OFICIAL(6)

A continuación se forman las palabras extendidas (“extended words”, unión de palabras y signos alineados), que representan la traducción propuesta. En este ejemplo:

- (el, λ) (denei, DNI) (es, SE-LLAMA) (obligatorio, OBLIGATORIO) (desde, DESDE) (los, PLURAL), (catorce, CATORCE) (años, AÑO EDAD)
- (el, λ) (denei, DNI) (es, SE-LLAMA) (el, λ) (documento, DOCUMENTO) (oficial, OFICIAL)

3. Inferencia de un Gramática Estocástica y posteriormente de un traductor de estados finitos. Se obtiene un FST a partir de las frases con las palabras extendidas. Las probabilidades de saltos entre nodos de un FST se computan por las cuentas correspondientes en el conjunto de entrenamiento de palabras extendidas. Se ilustra este proceso en la siguiente figura, donde los nodos grises indican que la frase de salida puede terminar en ese punto

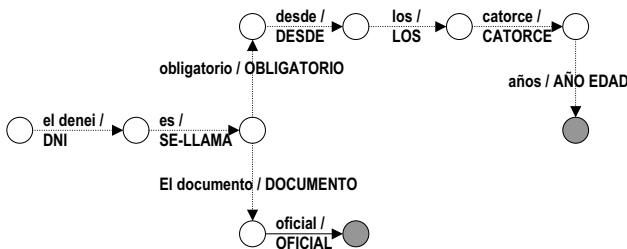


Figura 4. FST para el ejemplo anterior

5. INCORPORACIÓN DE INFORMACIÓN SINTÁCTICO-SEMÁNTICA A LA TRADUCCIÓN

5.1. Traducción basada en subfrases

La incorporación de información adicional a las subfrases utilizadas para la traducción constituye una

valiosa información en fases de pre o post-procesado de textos [7], ya que la traducción basada en subfrases se limita al “mapeo” de estos pequeños pedazos de texto sin incluir una información lingüística específica (morfológica, sintáctica o semántica). Con la hipótesis de que en la traducción automática se hace un uso pobre de la información morfológica que proporciona una palabra en sí misma, se plantea añadir unas categorías a las palabras para mejorar la tasa de traducción.

De acuerdo con esto, se han añadido una serie de datos (o factores) sobre las palabras. Esta información ha sido:

- Información sintáctico-semántica: una categoría que ofrezca cierta información sintáctico-semántica de la palabra. Para esta información se han aprovechado las categorías utilizadas en el sistema basado en reglas presentado en [1].
- Partes de la Oración (“Part of Speech”): una información sintáctica sobre la palabra. Se pueden clasificar las palabras, por ejemplo, en nombres, artículos, adjetivos, pronombres, verbos, adverbios, preposiciones, conjunciones, intersecciones, posesivos, demostrativos y conjunciones.
- Información adicional sobre el género o número de las palabras, el tiempo de los verbos y las características de los adverbios, por ejemplo.

Se ha realizado una labor manual para la categorización de las frases que componen la base de datos. Con este proceso de categorización, las palabras no son vistas únicamente como las unidades fundamentales del texto, sino como un vector de factores que representa varios niveles de anotación, como se observa en la siguiente figura:

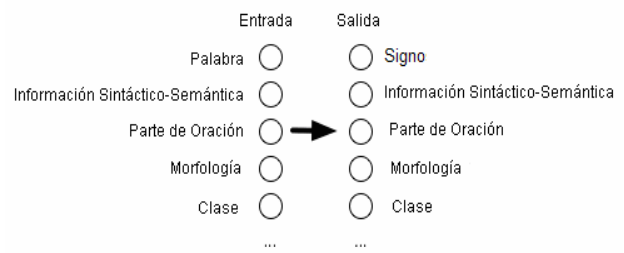


Figura 5. Diferentes tipos de factores

La plataforma de traducción basada en subsecuencias de palabras permite el entrenamiento de varios modelos considerando diferentes factores. Esos modelos se combinan para realizar la traducción. Esta combinación se hace mediante linealmente con unos pesos que se entrenan en la fase de validación.

5.2. Traducción basada en FSTs

En el caso de la traducción basada en FST se ha realizado una categorización previa del texto de entrada para luego aprender los modelos de traducción con los textos categorizados. En este caso hemos utilizado una

única categoría por palabra pero sólo para la lengua origen (castellano). Dicha categoría ha sido el primero de los factores comentados en el apartado anterior. El esquema de la arquitectura se muestra en la figura 6.

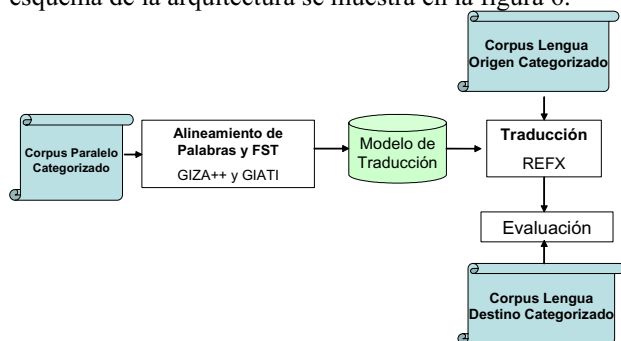


Figura 6. Utilizando textos categorizados

6. RESULTADOS Y DISCUSIÓN

La base de datos utilizada para los experimentos consiste en un corpus paralelo que contiene 414 frases típicas que diría un funcionario en una comisaría. El conjunto de frases se dividió aleatoriamente en tres grupos: entrenamiento (conteniendo el 70% de las frases), evaluación (15%) y test (15%). En relación con los experimentos de reconocimiento de voz se consideró únicamente el experimento 2 comentado en [1]. En el experimento 2 es aquel en el que el modelo de lenguaje se genera a partir del conjunto de entrenamiento, mientras que el vocabulario (540 palabras) incluye todas las palabras (entrenamiento y test). De esta forma se evita el problema de las OOVs (Out of Vocabulary words). Los resultados de reconocimiento obtenidos en este caso fueron WER (Word Error Rate) = 15,84, I (ins.) = 1,19%, B (borr.) = 5,93%, S (sus.) = 8,72%.

Sin información sintáctico-semántica				
		WER	BLEU	NIST
Traducción basada en subsecuencias	RAH	37,46	0,4939	6,474
	Ref	31,75	0,5469	6,865
Traducción basada en FST	RAH	33,42	0,5235	6,834
	Ref	28,21	0,5905	7,350
Con información sintáctico-semántica				
		WER	BLEU	NIST
Traducción basada en subsecuencias	RAH	37,13	0,5124	6,606
	Ref	31,54	0,5581	7,006
Traducción basada en FST	RAH	33,01	0,5311	6,943
	Ref	26,06	0,6071	7,664

Tabla 1. Resultados de traducción con y sin información sintáctico semántica

En la tabla 1 se incluyen los experimentos de traducción considerando las dos situaciones: incluyendo o no la información sintáctico-semántica descrita anteriormente. Se muestran los resultados de traducción tanto para las frases de referencia (Ref) como la salida del reconocedor (RAH). Para evaluar la calidad de la traducción se utiliza la WER (Tasa de Error a la salida de la traducción), BLEU [8] y NIST [9]. Mientras WER es una medida negativa (cuando mejor es el sistema menor es esta medida), BLEU y NIST son medidas positivas (aumentan con los mejores sistemas de traducción).

Como se puede observar, con esta base de datos (del dominio de frases del DNI/pasaporte), la traducción estadística basada en FST ofrece mejores resultados que la solución tecnológica basada en subfrases. Se observa también que al incorporar la información sintáctico-semántica es posible mejorar los resultados (menor WER y mayor BLEU o NIST). Si bien es cierto que los resultados mejoran, las diferencias son muy pequeñas. El único caso en el que la mejora es importante es para el caso de traducción basada en FST de las frases de referencia. Finalmente se puede concluir que el mejor sistema es el de la traducción basada en FST incorporando información sintáctico semántica.

8. AGRADECIMIENTOS

Este trabajo ha sido posible gracias a la financiación de los siguientes proyectos: EDECAN (MEC Ref: TIN2005-08660-C04), ROBONAUTA (MEC Ref: DPI2007-66846-C02-02) y ANETO (UPM-DGUI-CAM. Ref: CCG07-UPM/TIC-1823) y al trabajo en colaboración con la Fundación CNSE.

9. BIBLIOGRAFÍA

- [1] R. San-Segundo, A. Pérez, D. Ortiz, L. F. D'Haro, M. I. Torres, F. Casacuberta. "Evaluation of Alternatives on Speech to Sign Language Translation". Interspeech 2007. Antwerp, Bélgica. (ISSN:1990-9772)
- [2] <http://www.sign-lang.uni-hamburg.de/eSIGN>.
- [3] Stolcke A. "SRILM – An Extensible Language Modelling Toolkit". Interspeech.
- [4] Koehn P., Och J., Marcu D. "Statistical Phrase-Based Translation". Human Language Technology Conference '03 (HLTNAACL 2003), Edmonton, Canada, pp. 127-133, May 2003.
- [5] <http://www.statmt.org/moses>
- [6] Casacuberta F., E. Vidal. "Machine Translation with Inferred Stochastic Finite-State Transducers". Comp. Linguistics, V30, n2, 2005-225.
- [7] Koehn P., Hoang H. "Factored Translation Models". Proc. of the 2007 Annual Meeting of the ACL, pp. 868-876, Prague, Junio 2007.
- [8] Papineni, K., Roukos, S., Ward, T., and Zhu, W. J. "BLEU: a method for automatic evaluation of machine translation" in ACL-2002: 40th Annual meeting of the Association for Computational Linguistics pp. 311-318.
- [9] <http://www.nist.gov/speech/tests/mt/>