Procesamiento del lenguaje Natural, nº 41 (2008), pp. 251-258 recibido 7-05-2008; aceptado 16-06-2008

Aplicación de métodos estadísticos para la traducción de voz a Lengua de Signos

Using statistical methods for translating speech into Sign Language

B. Gallo, R. San-Segundo, J.M. Lucas, R. Barra, L.F. D'Haro, F. Fernández
Grupo de Tecnología del Habla. Universidad Politécnica de Madrid.
ETSIT. Ciudad Universitaria SN 28040. Madrid. Spain.
lapiz@die.upm.es

Resumen: Este artículo presenta un conjunto de experimentos para la realización de un sistema de traducción estadística de voz a lengua de signos para personas sordas. El sistema contiene un primer módulo de reconocimiento de voz, un segundo módulo de traducción estadística de palabras en castellano a signos en Lengua de Signos Española, y un tercer módulo que realiza el signado de los signos mediante un agente animado. La traducción se hace utilizando dos alternativas tecnológicas: la primera basada en modelos de subsecuencias de palabras y la segunda basada en transductores de estados finitos. De todos los experimentos, se obtienen los mejores resultados con el modelo que realiza la traducción mediante transductores de estados finitos con unas tasas de error de 26,06% para las frases de referencia, de 33,42% para la salida del reconocedor.
Palabras clave: Traducción Automática Estadística, Lengua de Signos, subfrase, Transductor de Estados Finitos, Modelo de Lenguaje, Modelo de Traducción, alineamiento, tasa de errores de palabras.

Abstract: This paper presents a set of experiments used to develop a statistical system from translating speech to sign language for deaf people. This system is composed of an Automatic Speech Recognition (ASR) system, followed by a statistical translation module and an animated agent that represents the different signs. Two different approaches have been used to perform the translations: a phrase-based system and a finite state transducer. The best results were obtained with the finite state transducer, with a word error rate of 26.06% for the reference text, and 33.42% using the ASR output.
Keywords: Statistical Machine Translation, Sign Language, phrase, Finite State Transducer, Language Model, Translation Model, alignment, word error rate.

1 Introducción

Con la realización de este trabajo se pretende el desarrollo y evaluación de una Plataforma de Traducción capaz de transformar, en base a un conjunto de modelos probabilísticos, frases de una lengua a otra, concretamente, de castellano a Lengua de Signos Española (LSE). La importancia de esta plataforma radica en la necesidad cada vez mayor de una herramienta que permita una traducción rápida y relativamente precisa entre lenguas. El coste de un intérprete signante (que conoce la Lengua de Signos) es muy elevado. Se debe tener en cuenta que la capacidad de comprensión del español de las personas sordas prelocutivas (aquellas que se quedaron sordas antes de poder hablar) es muy inferior a la de los oyentes. Así, presentan una capacidad lectora y de escritura en español muy inferior a la LSE, ya que no son capaces de extraer la información semántica de todas las palabras o construcciones, o no pueden formar una imagen mental de aquello que se les está comunicando. Se intenta, por lo tanto, desarrollar un software que permita traducir conjuntos de frases de castellano a una secuencia de signos de la LSE, que un avatar (agente visual) se encargará de representar.

2 Estado del arte

Numerosos proyectos de investigación se han enfocado a la traducción de habla natural, como por ejemplo en los casos de C-Star, ATR, Vermobil, Eutrans, LC-Star, PF-Star y TC-Star. A excepción de TC-Star, estos proyectos se dedican a la traducción de vocabularios medios (de unas 10000 palabras) en dominios restringidos de aplicación. Los sistemas de

ISSN: 1135-5948

traducción que dan mejores resultados son los basados en soluciones estadísticas (Och y Ney, 2002) (como el estudiado en este artículo), incluyendo técnicas basadas en ejemplos (Sumita et al, 2003), transductores de estados finitos (Casacuberta y Vidal, 2002) ("FST" en inglés) y técnicas basadas en subfrases (Koehn et al, 2003). Los avances importantes que se han conseguido en traducción de habla natural se debe a la aparición de medidas de error (Papineni et al, 2002), la mejora de eficiencia de algoritmos de entrenamiento (Och y Ney, 2003), el desarrollo de modelos dependientes del contexto (Sumita et al, 2003) y algoritmos de generación eficientes (Koehn et al, 2003).

En los últimos años, varios grupos de investigación han mostrado su interés en los sistemas de traducción de voz a Lengua de Signos desarrollando varios prototipos: basados en ejemplos (Morrissey, 2005), reglas (San-Segundo, 2006; Lynette, 2003), frases completas (Cox, 2002) o métodos estadísticos (Bungeroth, 2002) como el sistema de IBM SiSi (Say It Sign It). Este artículo presenta la evaluación de métodos estadísticos para la traducción a LSE de las explicaciones que un policía da a una persona que quiere renovar el DNI (Documento Nacional de Identidad).

3 Arquitectura del Sistema

El sistema completo está formado por tres módulos: el reconocedor de voz, el traductor estadístico y finalmente, la representación por un agente animado de los signos obtenidos:

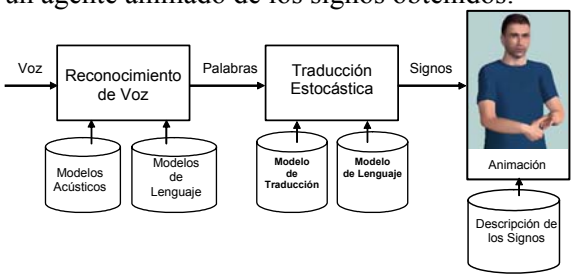

Figura 1: Arquitectura completa del sistema

3.1 Reconocimiento de Voz

Este módulo realiza la conversión del habla en lenguaje natural (habla continua) a una secuencia de palabras independiente del locutor. De esta manera, a partir de unos Modelos de Lenguaje y Acústicos de los que se dispone previamente, puede hacerse el análisis de la señal de habla ofreciendo a su salida una secuencia de palabras resultado.

3.2 Traducción Estadística

La 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 principalmente información de dos tipos de probabilidades:

- Probabilidad de traducción: recoge información sobre qué palabras se traducen por qué signos.
- Probabilidad de la secuencia de signos: aporta información sobre qué secuencias de signos son más probables en la LSE.

En este paso se realiza una traducción de las palabras provenientes del reconocedor a signos correspondientes, en este caso, a la Lengua de Signos Española. Para esto se utilizan métodos estadísticos cuyos Modelos se aprenden a partir de un corpus paralelo, compuesto por documentos de texto en castellano y sus equivalentes en Lengua de Signos. El documento de texto contendrá *palabras* en castellano, mientras que el de LSE contendrá GLOSAS. Las glosas son palabras (en mayúsculas) que representan los *signos*. Por ejemplo la glosa FOTO representa el signo cuyo significado es el de "fotografía".

3.3 Representación de Signos

El último módulo corresponde al agente animado en 3D, que se encarga de la representación de los signos provenientes de la Traducción Estadística. El agente utilizado es "VGuido" del proyecto eSIGN (http://www.sign-lang.unihamburg.de/eSIGN). Este módulo está incorporado en el sistema como un control ActiveX. Cada glosa (representación de un signo) está asociada a un fichero de texto XML con la descripción detallada de los movimientos que tiene que realizar el avatar para representar dicho signo. Para representar varias glosas seguidas basta con ir accediendo a los ficheros XML correspondientes e ir dándole las instrucciones necesarias al avatar para que realice los movimientos oportunos.

4 Traducción Estadística basada en Modelos de Subsecuencias de Palabras

La traducción estadística basada en modelos de subsecuencias de palabras (o subfrases) consiste en la obtención de un Modelo de Traducción a partir del alineamiento y extracción de

subsecuencias utilizando un corpus paralelo, y la generación de un modelo de lenguaje de la lengua destino. Estos modelos se utilizan por el módulo de traducción (Moses) para obtener la secuencia de signos/glosas dada una frase de entrada. La arquitectura completa de este sistema de traducción es la siguiente:

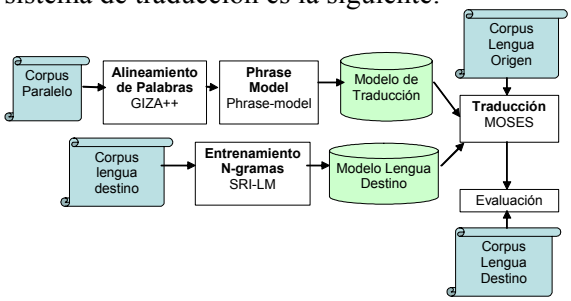

Figura 2: Arquitectura de la Traducción basada en Subsecuencias de Palabras

4.1 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 como destino). Las ideas que hay detrás de la traducción automática estadística vienen de la teoría de la información. Esencialmente, el problema de la traducción se centra en conocer la probabilidad p(d|o) de que una cadena *o* de la lengua origen genere una cadena *d* en la lengua destino. Estas probabilidades se calculan utilizando técnicas de estimación de parámetros a partir del corpus paralelo.

Aplicando el Teorema de Bayes a p(d|o) se puede representar esta probabilidad como el producto p(o|d)·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. Matemáticamente, encontrar la mejor traducción $\tilde{o}$ se consigue escogiendo aquella secuencia de signos que permita obtener la probabilidad máxima:

$$\tilde{o} = \arg\max_{o \in o^*} p(d/o) = \arg\max_{o \in o^*} p(o/d) \cdot p(d) \qquad (1)$$

Para la creación del Modelo de Lenguaje, se utiliza la herramienta SRILM (Stolcke, 2002), una herramienta que realiza la estimación de los modelos de lenguaje tipo N-grama, a partir del corpus de entrenamiento, y su evaluación calculando la probabilidad de un corpus de test. Estos Modelos se utilizan ampliamente en muchos ámbitos: reconocimiento de habla, OCR (Reconocimiento Óptico de Caracteres), etc. En cuanto a los Modelos de Traducción, su generación se hace mediante una traducción basada en subfrases. Para esto la herramienta utilizada es el GIZA++ (que es una implementación de los modelos IBM de traducción (Och y Ney, 2000)), un sistema de traducción estadística automática capaz de entrenar estos modelos para cualquier par de lenguas (http://www.statmt.org/moses). Para esto se necesita una colección de textos traducidos, que será el corpus paralelo. Los pasos para la generación de los modelos son:

1. Obtención del alineamiento entre palabras: consiste en que a partir de los dos textos en castellano y LSE se identifican qué palabras de uno corresponden con las del otro. Para esto se utiliza el programa GIZA++. El alineamiento se hace tanto en el sentido palabras-glosas como en la dirección glosas-palabras. Un ejemplo de un alineamiento es el siguiente:

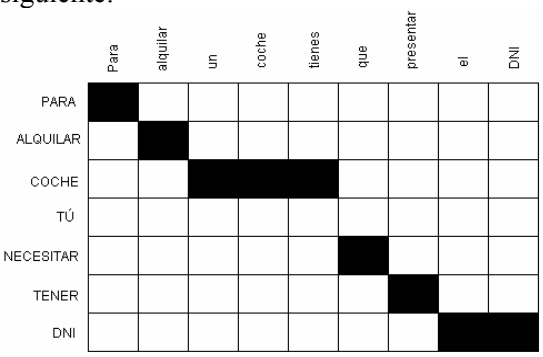

Figura 3: Ejemplo de un alineamiento entre palabras en castellano y signos en LSE representados mediante glosas.

2. Cálculo de una tabla de traducción léxica: a partir del alineamiento, se realiza una estimación de la tabla de traducción léxica más probable, obteniendo los valores de *w(d|o)* y su inversa *w(o|d)* para todas las palabras, es decir, las probabilidades de traducción para todos los pares de palabras. Un ejemplo para la palabra "por" con el texto utilizado es:

```
por PRIMER 0.5000000
por POR 0.3333333 …
```

3. Extracción de subsecuencias de palabras: se recopilan todos los pares de subsecuencias que sean consistentes con el alineamiento. El archivo generado en este paso tiene la forma siguiente, donde la subfrase "a los siguientes

países" se traduce por la subsecuencia de glosas "ESTOS PLURAL PAÍS":

```
a los siguientes países ||| ESTOS PLURAL
PAÍS ||| 0-0 2-0 1-1 3-2
a los siguientes ||| ESTOS PLURAL ||| 0-0
2-0 1-1
```

4. Cálculo de las probabilidades de traducción de cada subsecuencia ("Phrase Scoring"): En este paso, se calculan las probabilidades de traducción para todos los pares de subfrases en los dos sentidos: subfrase en castellano- signo en LSE y signo en LSE – subfrase en castellano. Un ejemplo del archivo obtenido es:

```
a los siguientes países ||| ESTOS PLURAL
PAÍS ||| (0) (1) (0) (2) ||| (0,2) (1) (3)
||| 1 0.0283293
a los siguientes ||| ESTOS PLURAL ||| (0)
(1) (0) ||| (0,2) (1) ||| 1 0.0661018
```

4.2 Ajuste

Para realizar el proceso de traducción se deben combinar los modelos generados en la fase anterior de entrenamiento. Esta composición se hace mediante una combinación lineal de probabilidades cuyos pesos se deben ajustar. El proceso de ajuste de los pesos consiste en probar el traductor Moses con un conjunto de frases (conjunto de validación) y, conociendo la traducción correcta, evaluar las salidas del traductor automático en función de los valores diferentes asignados a los pesos. Estos valores se eligen aleatoriamente y después de una búsqueda también aleatoria se eligen los valores que hayan ofrecido los mejores resultados.

4.3 Traducción

Utilizando un nuevo conjunto de frases (conjunto de test) se evalúa el sistema. Tanto para la fase de ajuste como para la de evaluación se utiliza el traductor Moses que emplea los modelos obtenidos anteriormente (modelo de traducción y modelo de lenguaje de la lengua destino), combinados según los pesos ajustados. Moses(http://www.statmt.org/moses) es un sistema de traducción automática estadística basado en subsecuencias de palabras, que implementa un algoritmo de búsqueda para obtener, a partir de una frase de entrada, la secuencia de signos que con mayor probabilidad corresponde a su traducción. Permite trabajar con redes de confusión de palabras como las que se obtienen en gran cantidad de sistemas de reconocimiento de voz. Por otro lado, también permite la integración de varios modelos de traducción entrenados con los diferentes factores con los que se puede etiquetar las palabras de las frases.

5 Traducción Estadística basada en Transductores de Estados Finitos

Los transductores de Estados Finitos ("FSTs: Finite State Transducers") se están usando en diferentes áreas de reconocimientos de patrones y lingüística computacional. Los FSTs parten de un corpus de entrenamiento que consta de pares de frases origen-destino, 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, por último, en un traductor de Estados Finitos. Una de las principales razones del interés de esta técnica es que las máquinas de estados finitos pueden aprenderse automáticamente a partir de ejemplos (Vidal et al, 2000).

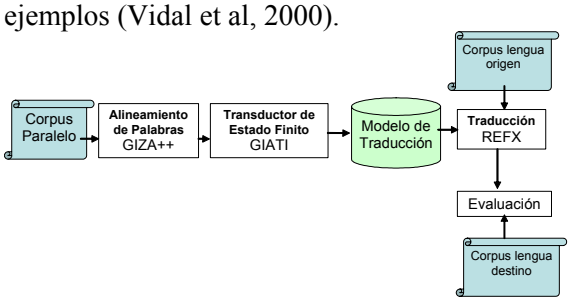

Figura 4: Arquitectura de la Traducción basada en Transductores de Estados Finitos

Un FST se caracteriza por la topología y por las distribuciones de probabilidad, dos características distintivas que se pueden aprender de un corpus bilingüe mediante algoritmos eficientes, como el GIATI ("Grammar Inference and Alignments for Transducers Inference"). En la Figura 4 se muestra la arquitectura de esta solución. Los pasos de esta estrategia de traducción son los que se explican a continuación.

5.1 Alineamiento con GIZA++

En esta fase se pretende el alineamiento de las palabras de las frases de entrada (en castellano) con los signos/glosas de sus traducciones correspondientes (en LSE). Este alineamiento se realiza en los dos sentidos: tanto en el sentido palabras-glosas como en la dirección glosas-palabras. Para realizar este alineamiento se utiliza el programa GIZA++ como se comentó en el apartado 3.1.1.

5.2 Transformación de pares de entrenamiento a frases de palabras extendidas

Partiendo de un alineamiento como el explicado en la sección 3.1, se realiza un proceso de etiquetado, en el cual se construyen un *corpus extendido* a partir de cada uno de los pares de subsecuencias de entrenamiento y sus correspondientes alineamientos: se asignarán por tanto palabras de lengua origen a su correspondiente palabra en lengua destino gracias a su alineamiento. Se muestra a continuación un ejemplo de pares castellano / LSE (en glosas) 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 obligatorio # DNI(2) SE-LLAMA(3) OBLIGATORIO(4)

el denei es el documento oficial # DNI(2) SE-LLAMA(3) DOCUMENTO(5) OFICIAL(6)

el denei es oficial # DNI(2) SE-LLAMA(3) OFICIAL(4)

Para que no se produzca una violación en el orden secuencial de las palabras en la lengua destino, se sigue el siguiente criterio de etiquetado: cada palabra de lengua destino se une con su correspondiente palabra en lengua origen a partir del alineamiento si el orden de las palabras objetivo no se altera. Si fuera así, la palabra en lengua destino se une con la primera palabra en lengua origen que no viole el orden de las palabras objetivo. Por lo tanto, el ejemplo anterior quedaría de la siguiente manera, con la formación de *palabras extendidas* ("extended words", unión de palabras y signos alineados):

(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) (obligatorio, OBLIGATORIO)

(el, λ) (denei, DNI) (es, SE-LLAMA) (el, λ) (documento, DOCUMENTO) (oficial, OFICIAL)

(el, λ) (denei, DNI) (es, SE-LLAMA) (oficial, OFICIAL)

Si se hace un refinamiento de este etiquetado, se puede implementar que las palabras origen que hayan quedado aisladas se unan a la primera palabra extendida que tenga palabra(s) destino asignadas. Por lo tanto, el ejemplo anterior se convierte en:

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

(el denei, DNI) (es, SE-LLAMA) (obligatorio, OBLIGATORIO)

(el denei, DNI) (es, SE-LLAMA) (el documento, DOCUMENTO) (oficial, OFICIAL)

(el denei, DNI) (es, SE-LLAMA) (oficial, OFICIAL)

5.3 Inferencia de un Gramática Estocástica y posteriormente de un Traductor de Estados Finitos

Consiste en la obtención de un Transductor de Estados Finitos 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. La probabilidad de una palabra extendida z_j a partir de una palabra origen s_i y una palabra destino t_i: $\mathbf{z}_j = (s_i, \mathbf{t}_i)$, dada una secuencia de palabras extendidas $\mathbf{z}_{i-n+1}$, $\mathbf{z}_{i-1} = (s_{i-n+1},\mathbf{t}_{i-n+1})$ $(s_{i-1},\mathbf{t}_{i-1})$ es:

$$p_n(z_i \mid z_{i-n+1}...z_{i-1}) = \frac{c(z_{i-n+1},...,z_{i-1},z_i)}{c(z_{i-n+1},...,z_{i-1})} \quad (2)$$

Donde c(·) es el número de veces que ocurre un evento en el conjunto de entrenamiento. A partir del resultado del apartado anterior se infiere un modelo tipo bigrama. Se ilustra este proceso en la siguiente figura, donde los nodos grises indican que la subfrase puede terminar en ese punto:

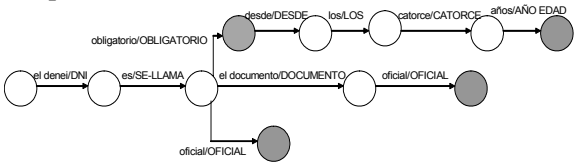

Figura 5: Transductor de estado finito inferido a partir del bigrama del ejemplo anterior

6 Evaluación

6.1 Medidas de evaluación

Con el objetivo de evaluar la calidad de la traducción, es necesario comparar la salida del sistema automático con una referencia y calcular algunas medidas de evaluación. WER ("Word Error Rate", porción de palabras con error) es una medida comúnmente utilizada en la evaluación de sistemas de reconocimiento del habla o de traducción automática. Calcula el número de inserciones, borrados y sustituciones de palabras cuando se comparan frases. Esta medida se basa en la distancia de edición o de Levensthein. En tareas tanto de traducción automática como de reconocimiento del habla,

se calcula el WER entre la frase generada por el sistema de traducción y una frase que es de referencia *correcta* (en este caso, frases de signos o glosas).

BLEU (Bilingual Evaluation Understudy) (Papineni et al, 2002) es un método de evaluación de la calidad de traducciones realizadas por sistemas de traducción automática. Una traducción tiene mayor calidad cuanto más similar es con respecto a otra de referencia, que se supone correcta. BLEU puede calcularse utilizando más de una traducción de referencia. Esto permite una mayor robustez a la medida frente a traducciones *libres* realizadas por humanos. BLEU se calcula normalmente a nivel de frases y halla la precisión en n-gramas entre la traducción del sistema y la de referencia. Estas medidas surgen con el objetivo de encontrar medidas automáticas que correlen con la evaluación que un experto haría de la traducción.

Otra medida es NIST, que se basa en la BLEU con algunas modificaciones: en primer lugar, BLEU utiliza la media geométrica de la precisión de los N-gramas, mientras que NIST utiliza una media aritmética para reducir el impacto de bajas concurrencias para órdenes altos de n-gramas. También, BLEU calcula la precisión de n-gramas utilizando pesos iguales para cada n-grama, mientras que NIST considera la calidad de la información que proporciona un n-grama particular en sí mismo (por ejemplo, cuanto menos frecuente sea un n-grama más peso se le asignará).

6.2 Base de Datos

La base de datos utilizada para los experimentos consiste en un corpus paralelo que contiene 414 frases típicas de un contexto restringido: aquellas que diría un funcionario cuando asiste a gente que quiere renovar el pasaporte y/o el Documento Nacional de Identidad, o información relacionada. En este contexto concreto, un sistema de traducción de voz a LSE es muy útil puesto que la mayoría de estos empleados no conocen este lenguaje y tienen dificultades a la hora de interactuar con personas sordas.

El conjunto de frases se dividió aleatoriamente en tres grupos: entrenamiento (conteniendo aproximadamente el 70% de las frases), evaluación (con el 15% de las frases) y test (15% de frases). Esta concentración se hace de forma arbitraria. Se muestra a continuación un resumen de la base de datos:

		Castellano	**LSE**
Total	**Pares de Frases**	**414**	
	Nº de palabras/glosas	**4847**	**4564**
Entrena-miento	Pares de Frases	314	
	Nº de palabras/glosas	3071	2876
Validación	Pares de Frases	50	
	Nº de palabras/glosas	582	574
Test	Pares de Frases	50	
	Nº de palabras/glosas	570	505

Tabla 1: Estadísticas de la base de datos

6.3 Resultados de los experimentos realizados

Las 414 frases fueron pronunciadas por 14 personas para evaluar el reconocedor de voz. En este caso se ha realizado tres experimentos diferenciados que se describen a continuación:

- En la primera situación, se evalúa el sistema de reconocimiento de voz utilizando tanto el modelo de lenguaje como el vocabulario generados a partir del conjunto de entrenamiento. Esta situación es la más realista.
- En el segundo caso, el modelo de lenguaje se genera a partir del conjunto de entrenamiento, mientras que el vocabulario incluye todas las palabras. De esta forma se evita el efecto de las palabras fuera de vocabulario.
- En el último experimento se utilizan todas las frases tanto para el entrenamiento como para el vocabulario. En este caso, se intenta reproducir la situación en la que se disponga de tantas frases de entrenamiento que las frases de test estén contenidas en ellas.

A continuación se expresan en forma de tabla los resultados obtenidos para los tres experimentos. Como parámetros de medida se incluyen: WER (Word Error Rate), I (inserciones), D (borrados) y S (sustituciones):

	WER	**I(%)**	**D(%)**	**S(%)**
Experimento 1	24,08	2,61	6,71	14,76
Experimento 2	15,84	1,19	5,93	8,72
Experimento 3	4,74	0,86	1,94	1,94

Tabla 2: Resultados del Reconocedor de Voz para los tres experimentos realizados

A continuación se muestran los resultados de traducción obtenidos aplicando las técnicas de traducción estadística descritas en los apartados 3 y 4. En la Tabla 3 se observan los resultados de los experimentos de traducción realizados, tanto con las frases de referencia del corpus paralelo castellano-LSE (Referencia), como utilizando la salida del reconocedor de voz para los tres experimentos de reconocimiento comentados anteriormente (Experimento 1-3). Por otro lado se diferencian dos situaciones principales: en la primera parte de la tabla se muestran los resultados habiendo entrenado el modelo de traducción con las frases de referencia, en segundo lugar, la segunda parte de la tabla muestra los mismos resultados pero en este caso considerando la salida de reconocedor (de las frases de entrenamiento) para entrenar el modelo de traducción. Para todos los casos se muestran los resultados de WER (tasa de error de signos a la salida de la traducción), tasas de signos insertados, borrados o sustituidos en la traducción y las medidas de BLEU y NIST descritas anteriormente.

Modelo de traducción generado con las frases de referencia del conjunto de entrenamiento				
		WER	BLEU	NIST
Traducción basada en subfrases	Exp 1	39,17	0,4853	6,2806
	Exp 2	37,99	0,4900	6,4006
	Exp 3	33,72	0,5301	6,7238
	REF	31,75	0,5469	6,8652
Traducción basada en FST	Exp 1	35,85	0,5090	6,6473
	Exp 2	33,89	0,5238	6,8610
	Exp 3	**29,32**	**0,5804**	**7,3100**
	REF	**28,21**	**0,5905**	**7,3501**
Modelo de traducción generado con la salida del reconocedor para las frases de entrenamiento				
		WER	BLEU	NIST
Traducción basada en subfrases	Exp 1	40,04	0,4775	6,2076
	Exp 2	37,46	0,4939	6,4738
	Exp 3	32,44	0,5449	6,8606
	REF	31,75	0,5469	6,8652
Traducción basada en FST	Exp 1	36,33	0,5188	6,5273
	Exp 2	33,42	0,5235	6,8344
	Exp 3	**29,27**	**0,5698**	**7,1953**
	REF	**28,21**	**0,5905**	**7,3501**

Tabla 3: Resultado de los experimentos de traducción

En esta tabla se puede observar que los resultados de la Referencia siempre serán los mejores resultados (menor WER y mayor BLEU y NIST) en comparación con los obtenidos en la traducción de la salida del reconocedor de voz puesto que la referencia no contiene errores de reconocimiento que dificultan la traducción posterior. Además, se puede ver que cuanto peor es la tasa de reconocimiento, peor es la tasa de traducción que se consigue traduciendo la salida del reconocedor. En general, 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 entrenar el modelo de traducción con las salidas de reconocimiento se permite entrenar dicho modelo con los posibles errores del reconocedor, de forma que el modelo de traducción puede aprender de estos errores y corregirlos durante el proceso de traducción. Si bien es cierto que los resultados mejoran, las diferencias son muy pequeñas. Finalmente se puede concluir que el mejor sistema es el de la traducción basada en FST entrenando con las salidas del reconocedor.

7 Conclusiones

En este artículo se ha presentado un sistema de traducción estadística de voz a lengua de signos para personas sordas. En concreto se ha estudiado la traducción de castellano a Lengua de Signos Española, utilizando como dominio de aplicación el de frases que un policía pronuncia cuando informa sobre cómo renovar o solicitar el DNI. Estas situaciones requieren de un intérprete signante que sea capaz de traducir cualquier frase a personas sordas que quieran realizar estas acciones por lo que un sistema automático puede ser de gran ayuda. En cuanto al sistema desarrollado, contiene un primer módulo de reconocimiento de voz, un módulo de traducción en el que se ha centrado este artículo, y un último módulo de representación de los signos. Se han estudiado dos soluciones tecnológicas de traducción estadística, la primera utiliza un modelo de traducción basado en subsecuencias de palabras y la segunda utiliza un transductor de estados finitos (FSTs). En ambos casos se utilizan programas de libre distribución que se comentan a lo largo del texto (GIZA++, Moses y GIATI). Estos programas permiten hacer la traducción tanto de textos en castellano (que contienen las frases originales) como de textos

que contienen las frases obtenidas a la salida del reconocedor de voz (que contendrán los fallos propios del reconocimiento) a LSE.

Los resultados que se muestran corresponden a pruebas con el texto original (de referencia) y el texto obtenido a la salida del reconocedor. Para este estudio, la base de datos de frases se ha dividido en tres subconjuntos: entrenamiento, validación y test (comprobación). En general, 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 subsecuencias de palabras. El hecho de entrenar el modelo de traducción con las salidas de reconocimiento permite aprender de los errores y corregirlos durante el proceso de traducción. Si bien es cierto que los resultados mejoran, las diferencias son muy pequeñas. Finalmente se puede concluir que el mejor sistema es el de la traducción basada en FST entrenando con las salidas del reconocedor con una WER de 29,27% y un BLEU de 0,5698.

8 *Prototipo desarrollado*

Figura 6: Interfaz del prototipo.

Con este trabajo se ha desarrollado un prototipo (Figura 6) de traducción de voz a LSE que ha sido evaluado con frases pronunciadas por estudiantes. El siguiente paso es evaluar el sistema en condiciones reales considerando interacciones reales entre los policías y personas sordas.

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)

Bibliografía

Bungeroth J. and Ney H. Statistical Sign Language Translation. Workshop on Representation and Processing of Sign Languages, LREC'04.

Casacuberta F., E. Vidal. "Machine Translation with Inferred Stochastic Finite-State Transducers". Comp. Linguistics, V30, n2, 205-225. 2002.

Cox S. J., Lincoln M., J Tryggvason, M Nakisa, M. Wells, Mand Tutt, and S Abbott. TESSA, a system to aid communication with deaf people. In ASSETS 2002, pages 205-212, Edinburgh.

Koehn P., Och J., Marcu D. "Statistical Phrase-Based Translation". Human Language Technology Conference 2003 (HLTNAACL 2003), Edmonton, Canada, pp. 127-133.

Lynette van Zijl Stellenbosch. "South African Sign Language Machine Translation System" Proc. of the 2nd international conference on Computer graphics, virtual Reality, visualisation and interaction in Africa table of contents Pages: 49 – 52. 2003

Morrissey S. and Way A. 2005. An example-based approach to translating sign language. In Workshop Example-Based Machine Translation (MT X–05), pages109–116, Phuket, Thailand, September.

Papineni K., S. Roukos, T.Wardm W.J.Zhu. "BLEU: a method for automatic evaluation of machine translation". 40th Annual Meeting of the ACL, Philadelphia, PA, pp. 311-318. 2002

Och J., Ney H. "Improved Statistical Alignment Models". Proc. of the 38th Annual Meeting of the Association for Computational Linguistics, pp. 440-447, Hongkong, China, Octubre 2000.

Och J., H.Ney. "Discriminative Training anda Maximum Entropy Models for Statistical Machine Translation".Annual Meeting of the Ass.ACL, Philadelphia, PA, pp.295-302. 2002

Och J., H. Ney. "A systematic comparison of various alignment models". Computational Linguistics, Vol.29, No.1, pp.19-51. 2003.

San-Segundo R., Barra R., L.F. D'Haro, J.M. Montero, R. Córdoba, J. Ferreiros. "A Spanish Speech to Sign Language Translation System". Interspeech 2006.

Stolcke A. "SRILM – An Extensible Language Modelling Toolkit" ICSLP. 2002.

Sumita E., Y.Akiba, T.Doi et al. "A Corpus-Centered Approach to Spoken Language Translation". Conf. Of Ass. For Computational Linguistics (ACL) Hungary, pp.171-174.2003.

Vidal E., Casacuberta F, García P. "Gramatical Inferrence and Automatic Speech Recognition". *New Advances and Trends in Speech Recognition and Coding* (volume 147 of NATO-ASI Series F: Casacuberta and Vidal. 2000.