Procesamiento del Lenguaje Natural, núm. 35 (2005), pp. 269-276 recibido 29-04-2005; aceptado 01-06-2005

Comparación de modelos de lenguaje en tareas de transcripción automática de noticiarios televisivos*

Javier Diéguez Tirado
ETSI Telecomunicación
Universidad de Vigo
jdieguez@gts.tsc.uvigo.es

Carmen García Mateo
ETSI Telecomunicación
Universidad de Vigo
carmen@gts.tsc.uvigo.es

Antonio Cardenal López
ETSI Telecomunicación
Universidad de Vigo
cardenal@gts.tsc.uvigo.es

Resumen: En el presente artículo se investigan diversas técnicas de modelado de lenguaje para una tarea de transcripción automática de noticiarios bilingües. Se compara una aproximación no adaptada con varios esquemas basados en interpolación de modelos. Mediante una estrategia de adaptación dinámica, utilizando reconocedores en paralelo, se ha conseguido reducir la tasa de errores de reconocimiento en un 20.7 % con respecto al modelo no adaptado. El artículo también analiza los problemas del habla espontánea que han limitado las mejoras.
Palabras clave: adaptación del modelo de lenguaje, transcripción de voz, programas de noticias

Abstract: In this paper several language models for a bilingual broadcast news transcription task are investigated. A non-adapted approach is compared to various schemes based on mixture models. Through the use of a dynamic adaptation strategy, employing several decoders in parallel, a 20.7 % reduction in the word error rate was achieved with respect to the non-adapted model. This paper also analyzes the problems of spontaneous speech, which have limited the improvements.
Keywords: language model adaptation, speech transcription, broadcast news

1. Introducción

La transcripción de noticiarios televisivos (Lamel et al., 2004) constituye un marco de trabajo idóneo para medir las prestaciones de un reconocedor de voz. La gran diversidad de locutores, estilos de habla y temas tratados a lo largo de un programa de noticias supone una exigente prueba para un reconocedor, obligándole a ser capaz de funcionar de manera robusta para un abanico de situaciones diferentes. Es por ello que esta tarea ha centrado una buena parte de la investigación en reconocimiento de voz en la última década.

El sistema Transcrigal de la Universidad de Vigo (Diéguez Tirado et al., 2004) fue diseñado para la transcripción de noticiarios en lengua gallega, que se caracterizan por la presencia frecuente de locutores que emplean el idioma castellano. El bilingüismo inherente a esta tarea constituye una nueva variable a tratar, que aumenta la complejidad y el interés del sistema.

* Este proyecto ha sido parcialmente apoyado por el MCyT de España, bajo el proyecto TIC2002-02208, y la Xunta de Galicia bajo el proyecto PGIDT03PXIC32201PN. También agradecemos la colaboración prestada por la Televisión de Galicia (TVG)

Para abordar esta variabilidad es conveniente recurrir a esquemas adaptados, tanto en lo relativo a los modelos acústicos como en el modelo de lenguaje. En este artículo se estudian varias aproximaciones al modelado de lenguaje en Transcrigal. Se comparará un esquema adaptado con un esquema no adaptado:

- Aproximación de fuerza bruta, que consiste en concatenar todo el texto de entrenamiento disponible.
- Adaptación con modelos de mezclas (Clarkson, 1999), en la cual el texto se divide en fuentes, y el modelo resultante se obtiene como combinación lineal de modelos componentes. La elección de los pesos permite la adaptación.

La adaptación por modelos de mezclas es bien conocida, si bien este artículo utiliza dos variaciones frente a las aproximaciones tradicionales: (i) aplicar el modelo adaptado en la fase de Viterbi del reconocedor, en lugar de esperar a la fase N-best; (ii) aumentar la especificidad de los modelos de lenguaje mediante la selección de subconjuntos dentro de dominio, en lugar de la aproximación habitual de entrenar "modelos de temas" fuera

de dominio, eg. (Gotoh y Renals, 1999). Se compararán varias estrategias realizando un aumento progresivo de la especificidad. El uso de estas técnicas ha permitido obtener una disminución del 20.7 % en la tasa de errores de reconocimiento con respecto a la aproximación por fuerza bruta.

El resto del artículo está organizado de la siguiente forma. A continuación, en el apartado 2 se proporciona una visión general del sistema completo de transcripción de noticias. En el apartado 3 se introducen algunos conceptos básicos sobre adaptación del modelo de lenguaje. El apartado 4 describe las soluciones propuestas para el modelo de lenguaje de Transcrigal. Seguidamente, el apartado 5 recoge los resultados experimentales. Finalmente, se realiza una discusión de los resultados obtenidos (Apdo. 6) y se proporcionan las conclusiones y las líneas futuras del trabajo (Apdo. 7).

2. *El sistema Transcrigal de transcripción de noticias*

En el presente apartado se resumen los componentes del sistema Transcrigal. Una descripción más detallada puede encontrarse en (Diéguez Tirado et al., 2004).

2.1. Bases de datos

Para la construcción de Transcrigal se utilizaron varias bases de datos:

- Transcrigal-DB. Es la base de datos propia del sistema. Está formada por grabaciones de noticiarios de la Televisión de Galicia, tanto en audio como en vídeo, así como su transcripción de texto. Esta base de datos ha sido ampliada recientemente de 14 a 31 programas (Tabla 1). Cada programa dura aproximadamente 1 h., distinguiéndose tres secciones: "noticias" (N), "deportes" (D) y "el tiempo" (T). Las transcripciones constan de un total de 315K palabras (2MB de texto). Los 31 programas se dividen en entrenamiento, validación y test (26, 2 y 3, respectivamente).
- Bases de datos de audio. Se utilizaron 25 horas en castellano y 15 horas de gallego tomadas de SpeechDAT, como material de entrenamiento para los modelos acústicos.
- Bases de datos de texto. Fueron utilizadas para para entrenar los modelos de lenguaje, y corresponden a la edición Internet de varios diarios en lengua gallega y castellana. También se pudo contar con las transcripciones utilizadas por los presentadores de los informativos, conocidas habitualmente como escaletas, proporcionadas por la TVG (Tabla 2).

	Remesa1	Remesa2
Núm. programas	14	17
Fecha captura	2002	2003–2004
Edición	mediodía	tarde
Codif. audio	PCM 16Khz	PCM 48Khz
Codif. vídeo	AVI (Indeo)	MPEG2

Tabla 1: Base de datos Transcrigal-DB

Nombre	id.	fechas	tamaño
El Correo Gallego	ES	12/00–01/05	366 MB
El Correo Gallego	GA	12/00–01/05	122 MB
Galicia Hoxe	GA	05/03–01/05	117 MB
Vieiros	GA	03/01–02/04	11 MB
Escaletas	GA	06/01–12/04	154 MB

Tabla 2: Material de texto

2.2. Estructura del sistema

El sistema Transcrigal consta de tres bloques fundamentales (Fig. 1): (i) segmentador acústico, divide cada programa de noticias en una serie de turnos de locutor; (ii) reconocedor de voz, transcribe cada turno de locutor utilizando un modelo de lenguaje y una serie de modelos acústicos determinados. El modelo de lenguaje se integra con el reconocedor en la fase de alineamiento de patrones (Fig. 2); (iii) visualizador: permite acceder a los contenidos multimedia en base a búsquedas sobre las transcripciones (Fig. 3).

3. *El mecanismo de adaptación del modelo de lenguaje*

A lo largo de este apartado se exponen algunos conceptos básicos sobre adaptación de modelos de lenguaje. En primer lugar, se presenta la adaptación de modelos de lenguaje como una herramienta capaz de solucio-

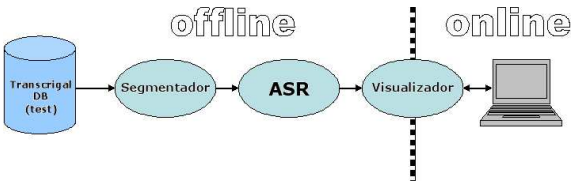

Figura 1: Diagrama de bloques de Transcrigal

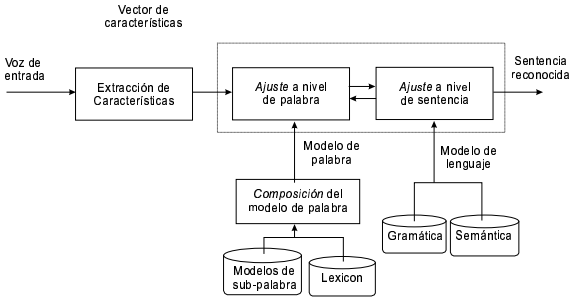

Figura 2: Integración del modelo de lenguaje en el reconocedor

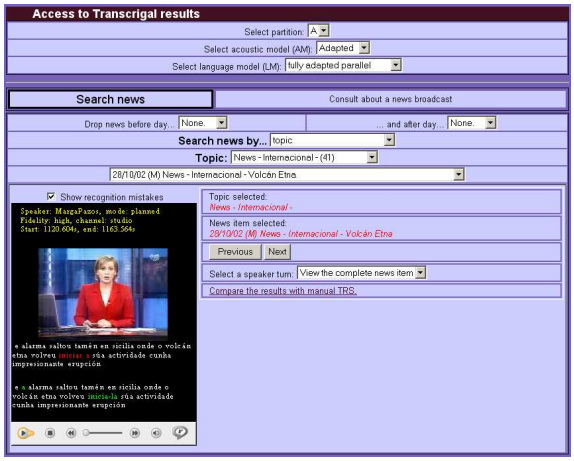

Figura 3: Visualizador de Transcrigal

nar algunas limitaciones de los modelos de n-gramas. Seguidamente se proporcionan los fundamentos de la adaptación por modelos de mezclas empleada en Transcrigal.

3.1. Adaptación de modelos de lenguaje

El reconocedor de voz de la Universidad de Vigo, utilizado en Transcrigal, está preparado para utilizar modelos de lenguaje trigramas, basados en descuento Good-Turing con backoffs (Manning y Schütze, 1999). El reconocedor aplica estos modelos de trigramas de manera eficiente en la fase de Viterbi del reconocedor por medio de un mecanismo de predicción avanzado (Cardenal-Lopez, Dieguez-Tirado, y Garcia-Mateo, 2002).

Para la construcción de un modelo de lenguaje para Transcrigal, podría optarse por la simple concatenación de todo el texto disponible, y entrenar a partir del mismo un modelo de trigramas (esquema de fuerza bruta). Sin embargo, esto proporcionaría unas prestaciones muy pobres, debido fundamentalmente a dos razones:

- Para que un modelo de n-gramas funcione correctamente, el texto de entrenamiento debe ser abundante, y estar ajustado a la tarea en tema y estilo. En cambio, en Transcrigal, se distingue una pequeña fracción de texto muy bien ajustado (Transcrigal-DB) y una serie de bases de datos complementarias más cuantiosas pero menos adecuadas (Tabla 2).
- Los n-gramas se entrenan sobre un texto estático, pero se aplican a dominios cambiantes. Es preferible un modelo capaz de ajustar sus parámetros a las características de cada realización, en lugar de un modelo universal.

Para superar estas limitaciones, es habitual recurrir a técnicas de adaptación de modelos de lenguaje (Bellegarda, 2004), en la cual un modelo de lenguaje de referencia, construido a partir de una gran cantidad de texto, se modifica para disminuir su perplejidad sobre un pequeño corpus de adaptación.

De entre las diversas técnicas de adaptación existentes, hemos seleccionado la interpolación de modelos de lenguaje, debido a las siguientes razones:

- El corpus de texto de Transcrigal (Tabla 2) está dividido de manera natural en fuentes diversas. La técnica de mezclas permite combinarlas de manera efectiva.
- El modelo resultante puede convertirse en un modelo de n-gramas autónomo, lo cual permite su aplicación eficiente en la fase de Viterbi del reconocedor. La aplicación temprana del modelo adaptado ayudará a mejorar las prestaciones.

3.2. Fundamentos de la adaptación por modelos de mezclas

Los fundamentos de la adaptación de modelos de lenguaje por mezclas son los siguientes (Clarkson, 1999): el texto disponible se organiza manual o automáticamente en una serie de *fuentes*. Con cada fuente de texto se entrena un modelo de n-gramas P_j, y los distintos modelos son combinados por medio de interpolación lineal, para dar lugar al siguiente modelo adaptado:

$$P(w_i|h_i) = \sum_{j=1}^{N} \lambda_j P_j(w_i|h_i)$$

$$h_i = w_{i-n+1}, \cdots, w_{i-1}$$

Nombre	Descripción
trs-f	Transcrigal-DB entrenam. (filtrada)
trs	Transcrigal-DB entrenam. (total)
esc	Escaletas (filtrada)
p-GA	ECG-GA, Galicia Hoxe, Vieiros
p-ES	ECG-ES

Tabla 3: Organización del texto en fuentes

donde la probabilidad P se calcula para una palabra w_i con una historia dada h_i. Los pesos de interpolación $\{\lambda_j\}$ se obtienen por medio del algoritmo EM, minimizando la perplejidad de un cierto corpus de adaptación.

En el caso de Transcrigal, el texto disponible se organizó en cinco fuentes distintas, según la tabla 3. El corpus de adaptación fue extraído del conjunto de validación de Transcrigal-DB. El esquema para obtener los modelos adaptados fue el siguiente: en primer lugar, los LMs componentes fueron entrenados como trigramas por medio del paquete SRILM (Stolcke, 2002) con suavizado de Katz. Tras el cálculo de los pesos, según las estrategias explicadas en el Apartado 4, cada modelo resultante se convirtió a un modelo de trigramas autónomo. Finalmente, se aplicó poda basada en entropía con umbral $2,5 \cdot 10^{-8}$, y el vocabulario se limitó a 20K palabras, para permitir su uso con el reconocedor.

4. *Esquemas adaptados para Transcrigal*

En este apartado, se proponen tres esquemas diferentes para el modelado de lenguaje de Transcrigal, basados en adaptación con modelos de mezclas. En primer lugar, se describe un esquema de adaptación al dominio. Seguidamente, se describe una aproximación que aprovecha la estructura temporal de la tarea. Finalmente, se detalla un esquema dinámico, que consigue una adaptación a tema, estilo e idioma.

La principal novedad de estas estrategias consiste en el aumento de la especificidad de los modelos escogiendo subconjuntos homogéneos en el texto dentro de dominio, en lugar del procedimiento habitual de identificar estos subconjuntos en el corpus de referencia.

4.1. Adaptación al dominio

Para realizar una adaptación al dominio, únicamente es necesario entrenar un conjunto de pesos, de manera que se obtenga un modelo de lenguaje universal. Este modelo será aplicado a todos los turnos de locutor de manera ciega, al igual que si fuese un modelo de lenguaje obtenido por fuerza bruta, pero con la ventaja de haber combinado el material de entrenamiento de manera adecuada con respecto a la tarea.

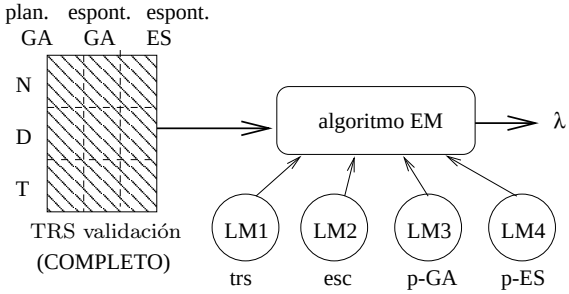

Figura 4: Creación del LM universal

LM	trs	esc	p-GA	p-ES
univ.	0.173	**0.471**	0.268	0.088

Tabla 4: Pesos LM universal

Para entrenar el modelo de lenguaje universal, se combinaron cuatro de las fuentes de la Tabla 3, y se utilizó como corpus de adaptación la totalidad del texto de validación de Transcrigal-DB. Este proceso se ilustra en la Figura 4. Los pesos obtenidos se muestran en la Tabla 4. Se observa que la fuente de mayor peso corresponde a las escaletas, a pesar de carecer de transcripciones de habla espontánea, por ofrecer el mejor compromiso entre ajuste a la tarea y cantidad de texto. Si bien la fuente "trs" corresponde a material más ajustado, no se le asigna un peso importante debido a su escasez.

4.2. Adaptación por bloques

El siguiente esquema propuesto aprovecha que cada programa de noticias está separado en tres bloques bien diferenciados: noticias, deportes y tiempo. Precediendo a cada bloque se encuentra una sintonía característica que permite su fácil identificación. Por tanto, puede asumirse que para cada turno de locutor que se desee reconocer, el bloque al que pertenece será conocido, siendo factible la aplicación de un modelo adaptado a ese bloque.

Conforme a este planteamiento, se entrenó un modelo diferente para cada tema, utilizando como corpus de adaptación el sub-

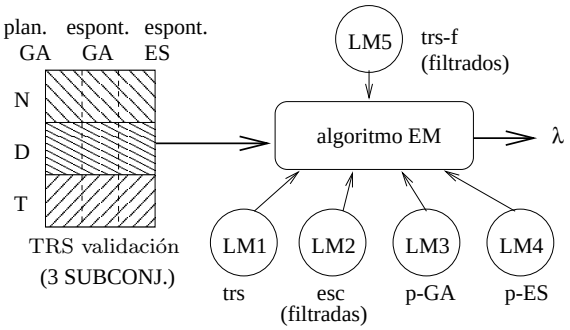

Figura 5: Creación de los LMs por bloques

LM	trs-f	trs	esc	p-GA	p-ES
N	0.07	0.02	**0.49**	0.34	0.08
D	0.15	0.04	**0.52**	0.14	0.15
T	**0.83**	0.02	0.11	0.04	0.00

Tabla 5: Pesos LMs por bloques

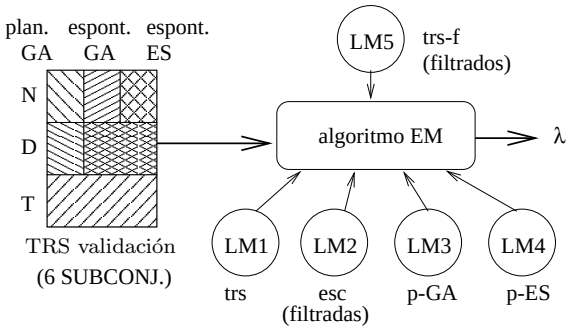

Figura 6: Creación de los LMs para adapt. dinámica

LM	trs-f	trs	esc	p-GA	p-ES
N-pl-GA	0.05	0.01	**0.60**	0.33	0.01
N-sp-GA	0.19	0.07	0.13	**0.53**	0.09
N-sp-ES	0.23	0.01	0.01	0.03	**0.71**
D-pl-GA	0.06	0.03	**0.74**	0.16	0.01
D-sp	0.33	0.01	0.02	0.05	**0.59**
T	**0.83**	0.02	0.11	0.04	0.00

Tabla 6: Pesos LMs para adapt. dinámica

conjunto correspondiente de la parte de validación. Se combinaron cinco fuentes y se hizo uso de filtrado en algunas de ellas, para mejorar la correspondencia entre corpus de adaptación y texto de entrenamiento, de esta manera:

- Las escaletas fueron divididas en noticias y deportes en base a sus etiquetas. Para los bloques de noticias y deportes, se escogió el subconjunto correspondiente. Para el bloque de el tiempo, se escogió la totalidad de las escaletas.
- La fuente "trs-f" está formada por el subconjunto de entrenamiento de Transcrigal-DB correspondiente al bloque. La fuente "trs" corresponde a la totalidad del texto de entrenamiento, sin filtrado.

El proceso de creación de los tres modelos se representa en la Figura 5. Los pesos obtenidos para cada uno de los tres modelos se recogen en la Tabla 5. Se observa cómo la importancia de cada fuente varía claramente en función del bloque considerado. En todos los casos, se le asigna un mayor peso a la fuente filtrada "trs-f" que a la fuente "trs", debido a que está más ajustada, a pesar de contar con un tamaño menor.

4.3. Adaptación dinámica

Si bien el esquema anterior consigue una adaptación al tema, se realizaron ciertas modificaciones para incorporar adaptación al estilo (habla planeada o espontánea) y al idioma (castellano o gallego). Procediendo de manera análoga al caso anterior, se crearon 6 modelos adaptados, utilizando como corpus de adaptación subconjuntos concretos del corpus de validación:

- Para el bloque de noticias, se crearon 3 modelos: habla planeada en gallego, habla espontánea en gallego y habla espontánea en castellano. No se creó un modelo para habla planeada en castellano debido a que el texto en Transcrigal-DB para esta condición es insuficiente.
- En el bloque de deportes, debido a que casi toda el habla espontánea es en idioma castellano, se creó un único modelo para habla espontánea cubriendo ambos idiomas.
- Para el bloque de el tiempo, únicamente se entrenó un modelo de lenguaje, debido a que todos los turnos corresponden al locutor principal.

Los pesos de los seis modelos entrenados se recogen en la Tabla 6, y los subconjuntos de validación utilizados se resumen en la Figura 6.

Para aplicar el modelo adecuado a cada turno de locutor, podría haberse procedido con un esquema multipase. Sin embargo, debido a que únicamente existe incertidumbre acerca de un máximo de 3 modelos, ha resul-

LM	PPL (int.)	%OOV (20K)	%WER
fuerza bruta	155.8	5.72	37.13
mezcla estática	109.9	4.76	32.75
mezcla bloques	95.9	4.39	31.67
mezcla dinámica	**84.7**	**3.94**	**29.55**

Tabla 7: Resultados globales

tado factible una implementación basada en reconocedores en paralelo, decidiendo el modelo adecuado en base a la puntuación final de reconocimiento. Este esquema proporciona dos ventajas sobre el anterior:

- Permite realizar adaptación dinámica sin utilizar transcripciones erróneas, lo cual evita una fuente de problemas.
- Facilita la implementación en tiempo real, lo cual puede ser útil para aplicaciones futuras de subtitulado en directo.

5. Resultados experimentales

El el presente apartado se analizan las prestaciones de cada uno de los mecanismos de modelado de lenguaje propuestos, frente a la parte de test de Transcrigal-DB (3 programas de noticias, con un total de 31577 palabras etiquetadas). Se realizaron experimentos de texto, en base a perplejidad y tasa de palabras fuera de vocabulario (OOV), y experimentos de reconocimiento. La tabla 7 presenta los resultados.

5.1. Experimentos de texto

La columna PPL de la tabla 7 representa la perplejidad de cada LM frente al texto de test de Transcrigal-DB. La perplejidad se obtuvo para los modelos de lenguaje ya podados por entropía, restringiéndolos a un vocabulario común de manera que se permitiera la comparación. Se utilizó el vocabulario intersección, formado por 6571 palabras presentes en todos los modelos de lenguaje obtenidos, con una tasa de palabras fuera de vocabulario de 21.07 %. En el caso de los modelos por bloques y dinámicos, se aplicaron únicamente a los subconjuntos del test correspondientes. La columna "OOV" muestra la tasa de palabras fuera de vocabulario, de cada modelo de lenguaje, una vez podado a 20K palabras para poder ser utilizado por el reconocedor.

Ambas variables analizan distintos aspectos acerca la calidad del LM resultante. Mientras la tasa de OOV mide la capacidad del LM para seleccionar un lexicón adaptado, la perplejidad mide el poder predictivo del modelo de lenguaje frente al test. Ambos aspectos influirán de manera complementaria en la tasa de reconocimiento final.

Los resultados obtenidos indican que la aproximación por fuerza bruta, al haber sido realizada sin tener en cuenta la naturaleza de la tarea, se ve claramente superada por las técnicas basadas en mezclas. A medida que la estrategia tiene en cuenta las variaciones puntuales de tema y estilo, se obtienen mejoras tanto en tasa de OOV como en perplejidad.

5.2. Experimentos de reconocimiento

Los experimentos de reconocimiento se realizaron utilizando modelos acústicos adaptados a locutores masculinos, femeninos, y locutores principales, según el procedimiento explicado en (Diéguez Tirado et al., 2004). Los parámetros de poda del reconocedor fueron ajustados para una ejecución en 3 veces tiempo real. Se partió de una segmentación manual del material de test, para evitar errores derivados de una segmentación automática imperfecta.

La columna WER de la Tabla 7 muestra la tasa de errores de reconocimiento para la parte de test de Transcrigal-DB, utilizando cada uno de los esquemas de modelado de lenguaje propuestos. Se observan resultados bastante correlados con los valores de perplejidad obtenidos.

En la Tabla 8 se desglosa la WER para la aproximación de fuerza bruta y la adaptación dinámica. También se incluye la proporción del test que corresponde a cada grupo desglosado. Si bien se observan mejoras para todos los grupos de locutores, el porcentaje de errores en la parte de habla espontánea (en negrilla) sigue siendo muy alto, si bien corresponde únicamente a un 19 % del test.

6. Discusión. Los problemas del habla espontánea

En este artículo se han presentado algunos mecanismos de adaptación al modelo de lenguaje para una tarea de transcripción de noticias. El mejor de los mecanismos propuestos, ha proporcionado una mejora relativa de un 20.7 % en tasa de reconocimiento, con respecto a una aproximación basada en fuerza bruta (Tabla 7). No obstante, un desglose de los resultados (Tabla 8) indica una gran di-

Bloque	Locutores	%test	% WER	
			f.bruta	ad.din.
N	Loc. ppal	21.61	18.41	14.24
	Reporteros	34.85	34.33	26.85
	Entrev-GA	7.46	59.79	**53.42**
	Entrev-ES	6.44	61.18	**57.64**
D	Loc. ppal	5.46	27.39	17.93
	Reporteros	11.97	42.57	29.60
	Entrev-GA	1.52	80.38	**75.57**
	Entrev-ES	3.71	72.76	**66.10**
T	Loc. ppal	6.98	32.56	18.50
Total		100.0	37.13	29.55

Tabla 8: Desglose de la WER

ferencia entre los resultados para habla planeada (WER entre 14 y 29 %) y aquellos obtenidos para habla espontánea (WER entre 53 y el 75 %). Es necesario por tanto profundizar en las razones de este comportamiento, para intentar solucionar el problema de cara a próximos trabajos.

Una de las razones fundamentales para el pobre funcionamiento con habla espontánea consiste en la ausencia de material específico de entrenamiento. Tanto los corpora periodísticos utilizados, como las escaletas, corresponden fundamentalmente a habla de tipo planeado. Únicamente se ha sacado partido de la parte de habla espontánea de Transcrigal-DB, si bien su escasez no ha permitido grandes mejoras. El hecho de trabajar en idioma gallego implica una mayor dificultad para la adquisición de corpora de texto. Actualmente estamos investigando el uso de guiones de series de televisión y películas, para paliar este problema. También estamos utilizando mecanismos de recuperación de información para intentar aislar la pequeña fracción de habla espontánea que puede estar presente en nuestros corpora actuales.

Al margen de la inadecuación de los corpora de texto, existen otros problemas asociados al habla espontánea. En primer lugar, se dan todo un conjunto de disfluencias (repeticiones, muletillas, palabras inacabadas, etc.) no presentes en habla planeada, que dificultan el modelado de lenguaje utilizando n-gramas. Un modelo de lenguaje basado en conteos no modelará nunca de manera correcta estos fenómenos, que surgirán siempre de manera aleatoria independientemente de lo observado en el corpus de entrenamiento. Asimismo, el reconocedor de voz siempre intenta ajustar la secuencia acústica a palabras presentes en el vocabulario. Para tener en cuenta los fenómenos mencionados, habría que liberarlo de esta restricción.

Finalmente, otro aspecto que puede estar influyendo son los parámetros utilizados para la detección de actividad dentro del reconocedor, los cuales están ajustados para habla planeada. El habla espontánea incluye normalmente pausas durante las frases, y el detector de actividad provoca que los segmentos entre dos pausas se consideren frases separadas, no utilizando la historia de palabras anterior a la pausa. Sería por tanto conveniente aplicar una detección de actividad adaptativa al tipo de hablante.

7. *Conclusiones y líneas futuras*

En el presente artículo, se han propuesto diversos mecanismos de modelado de lenguaje para una tarea de transcripción de noticiarios bilingües. Se ha comparado un esquema no adaptado, basado en concatenación del texto de entrenamiento, con diversos esquemas adaptados basados en modelos de mezclas. El uso de modelo de mezclas ha permitido aprovechar el texto de entrenamiento disponible de manera efectiva, y poder aplicar el modelo adaptado resultante desde el principio del proceso de reconocimiento. Mediante la identificación de condiciones típicas de tema, estilo e idioma dentro del corpus la tarea, se ha desarrollado un esquema de adaptación dinámica realista basada en reconocimientos en paralelo. Este esquema de adaptación dinámica proporciona dos ventajas con respecto a aproximaciones multipase (i) no es necesario depender de transcripciones incorrectas para la adaptación (ii) facilita el reconocimiento en tiempo real. El modelo adaptado dinámico ha permitido una mejora de un 20.7 % en tasa de error con respecto al esquema no adaptado. Se ha observado un comportamiento pobre para el habla espontánea, y se han identificado los problemas que dan origen a este comportamiento: falta de adecuación del corpus de entrenamiento, mal modelado de las disfluencias del lenguaje, y detección de actividad inadecuada.

El cuanto a próximas líneas de actuación, se está trabajando en la recopilación de corpora de habla espontánea, así como en técnicas para aislar habla espontánea de los corpora existentes. También se están investigando mecanismos de clustering jerárquico del material de entrenamiento, que permitan combi-

nar el material con una mayor granularidad, combatiendo los efectos de fragmentación.

Bibliografía

Bellegarda, J. 2004. Statistical language model adaptation: review and perspectives. *Speech Communication*, 42(1):93–108, January.

Cardenal-Lopez, A., F. J. Dieguez-Tirado, y C. Garcia-Mateo. 2002. Fast LM look-ahead for large vocabulary continuous speech recognition using perfect hashing. En *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, volumen 1, páginas 705–708, Orlando, FL, May.

Clarkson, Philip R. 1999. *Adaptation of Statistical Language Models for Automatic Speech Recognition*. Ph.D. tesis, University of Cambridge.

Diéguez Tirado, Javier, Carmen García Mateo, Laura Docío Fernández, y Antonio Cardenal López. 2004. Transcrigal: Sistema de transcripción de noticias de la universidad de vigo. En Emilio Sanchis Arnal, editor, *Terceras Jornadas en Tecnología del Habla*, páginas 243–248, Valencia, Spain, November.

Gotoh, Y. y S. Renals. 1999. Topic-based mixture language modelling. *J. Natural Language Engineering*, 5:355–375.

Lamel, L., J-L. Gauvain, G. Adda, M. Adda-Decker, L. Canseco, L. Chen, O. Galibert, A. Messaoudi, y H. Schwenk. 2004. Speech transcription in multiple languages. En *Proc. IEEE Int. Conf. Acoustics, Speech and Signal Processing*, volumen 3, páginas 757–760, Montreal, Canada, May.

Manning, Christopher D. y Hinrich Schütze. 1999. *Foundations of Statistical Natural Language Processing*. The MIT Press, Cambridge, Massachusetts.

Stolcke, A. 2002. SRILM – an extensible language modeling toolkit. En *Proc. Int. Conf. Spoken Language Processing*, volumen 2, páginas 901–904, Denver, CO, September.