

Selección de unidades léxicas para reconocimiento automático del habla continua en euskera

Karmele López de Ipiña Peña

University of the Basque Country
Sistemen Ingeniaritza eta Automatika
EUITI Nieves Cano 12 Vitoria-Gasteiz
isplopek@vc.ehu.es

Manuel Graña Romay

University of the Basque Country
Konputazio-Zientzia eta Adimen Artifiziala
Informatika Fakultatea, Ibaeta Donostia
ccpgrrom@si.ehu.es

Ekaitz Zulueta Guerrero

University of the Basque Country
Sistemen Ingeniaritza eta Automatika
EUITI Nieves Cano 12 Vitoria-Gasteiz
isplopek@vc.ehu.es

Aitzol Ezeiza Ramos

University of the Basque Country
Konputagailuen Arkitektura eta Teknologia
Informatika Fakultatea, Ibaeta Donostia
aitzol@si.ehu.es

Resumen: El euskera es una lengua aglutinante, lo que implica que el vocabulario de un corpus no puede definirse mediante palabras porque crece combinatorialmente y se hace intratable para tareas de medio y gran vocabulario. Los pseudo-morfemas, generados mediante una herramienta de segmentación automatizada pueden ser una buena alternativa para la construcción del lexicón y de un modelo del lenguaje, puesto que reducen notablemente el tamaño del vocabulario. En euskera el número de morfemas cortos y acústicamente muy parecidos es muy alto. Este es un fenómeno que debe de tenerse en cuenta ya que el proceso de decodificación acústico fonética puede influir en el CSR, al aumentar la posibilidad de confusión e inserción de ciertas unidades léxicas (unidades muy cortas y con alta tasa de confusión acústica). Una posible forma de abordar el problema es no segmentar estas unidades. El siguiente paso en la mejora del sistema de CSR en euskera es la utilización de un modelo de lenguaje para guiar el proceso de reconocimiento.

Palabras clave: reconocimiento del habla continua, unidades léxicas, modelo de lenguaje, euskera

Abstract: Basque is an agglutinative language, which implies that corpusvocabulary can not be defined with words, because they growcombinationally making medium and large vocabulary tasks intractable. Pseudo-morphemes, generated with an automatic segmentation tool, could be an alternative choice for building the lexicon and the language model, for they notably reduce the vocabulary size. In Basque, there are many short and acoustically very similar morphemes. This phenomenon has to be taken into account, because the acoustic-phonetic decodification process can influence the CSR task, increasing the possibility of confussion and insertion of certain lexical units (very short units with high rates of acoustic confussion). A feasible way to deal with this problem is to avoid the segmentation of those units. The next step to improve the CSR system in Basque is the use of a language model in order to guide the recognition process.

Keywords: Continuous Speech Recognition, Lexical Units, Language Model, Basque.

1 Introducción

La definición de unidades léxicas apropiadas (Lexical Unit, LU) es crucial para el desarrollo de un sistema de CSR (Continuous Speech Recognition, CSR) ya que, mediante ellas, se define el vocabulario de la tarea. Además, los modelos de las unidades subléxicas (Sublexical Unit, SLU) se integran en el sistema de CSR combinándose en un nivel superior a la decodificación acústico-fonética formando modelos de LUs. Por lo tanto, la definición de un conjunto adecuado de LUs es imprescindible no sólo para integrar las SLUs sino para construir modelos de lenguaje (Language Model, LM) apropiados para el sistema.

Generalmente, la palabra ha sido la unidad utilizada en los sistemas de CSR como LUs. Sin embargo, en los últimos años han empezado a aparecer diferentes propuestas de LUs de tipo no-palabras (Otsuki *et al.*, 1999) (Peñagarikano *et al.*, 1999) (Larson *et al.*, 2000) (Larson 2001) (Carký, Geutner y Schultz, 2000) para lenguas donde la palabra no es una propuesta de LU viable. El euskera es una lengua aglutinante que posee una cierta estructura en el interior de las palabras (Alegria *et al.* 1996) por lo que este último tipo de unidades puede ser una buena alternativa.

Por otra parte, la proliferación de sistemas multilingües en los últimos años (Schultz, Westphal y Waibel, 1997), donde se integran lenguas con inflexiones como el alemán, lenguas sin palabras definidas como el japonés o lenguas aglutinantes como el turco y el finlandés¹, ha provocado una investigación más profunda en la búsqueda de LUs alternativas a la palabra que permitan definir de un modo más preciso los LMs. En este trabajo² se abordará la selección de LUs apropiadas para el euskera siguiendo la línea de las investigaciones realizadas para otras lenguas de sus características similares. El artículo se estructura de la siguiente manera: el punto 2 analiza el modelado de LUs para el euskera, el punto 3 aborda la experimentación y en el punto 4 se resumen las conclusiones.

¹ En Europa varias de las 11 lenguas oficiales poseen este tipo de características. Por ejemplo: el alemán, finlandés o el euskera.

² Financiado por los proyectos: TIC2002321 UPV/EHU 2002 e HIZKING21 plan estratégico ETORTEK del Gobierno Vasco

2 Modelado de LUs para el euskera

La razón de la utilización de la palabra como LU en los sistemas de CSR, es que los textos escritos (en el sentido estricto) se componen de palabras y, en muchos idiomas, puede considerarse como la LU natural: inglés, francés, español, etc. Sin embargo en algunos casos, las palabras no están definidas en la frase como es el caso del japonés (Otsuki *et al.*, 1999), o poseen cierta estructura interna dentro de la palabra como es el caso de finlandés, alemán, euskera, turco, etc. (Sivola, Kurimo y Lagus, 2001) (Ney, Essen y Kneser, 1994) (Lopez de Ipina *et al.*, 2000) (Schultz y Waibel, 2001) (Mengusoglu y Ris, 2001).

En estos casos, la utilización de la palabra como unidad base para definir el vocabulario de la tarea, provoca el crecimiento combinatorio del tamaño de este haciendo inviable su definición. Esto provoca la imposibilidad de entrenar apropiadamente un LM y, además, la tasa de palabras fuera del vocabulario (Out of Vocabulary, OOV) se dispara (Larson, 2001) (Otsuki *et al.*, 1999).

En el caso de los lenguajes aglutinantes (como el euskera), el uso de la palabra como unidad de reconocimiento no tiene en cuenta este tipo de estructuras y limita y deshecha información sintáctica importante para el LM, puesto las palabras pueden contener información morfo-sintáctica. Además, el entrenamiento de un LM de palabras, en este caso, sería muy pobre ya que no hay regularidades gramaticales a nivel de palabras en el sentido estricto. Por otra parte, una selección no apropiada del conjunto de LUs puede desembocar en un lexicón poco eficiente y degradar el rendimiento del reconocimiento.

Muchas de las lenguas anteriormente mencionadas (japonés, turco, finlandés etc.) utilizan el morfema como primera aproximación de LU en los sistemas de CSR.

A continuación analizaremos varias propuestas de LUs, que podrían ser apropiadas para el euskera.

2.1 Morfemas

El morfema puede considerarse como la entidad gramatical más pequeña. Las palabras se componen de morfemas o de otras palabras. La utilización de morfemas como LUs posibilita la reducción del vocabulario de la tarea considerablemente. En el caso del japonés (Otsuki *et al.*, 1999), turco (Mengusoglu y Ris 2001) (Carký, Geutner y Schultz, 2000) (Schultz y Waibel, 2001) y el finlandés (Creutz y Lagus, 2002) (Sivola, Kurimo y Lagus, 2001) se han propuesto los morfemas como LUs.

En el caso del euskera, la estructura morfológica dentro de la palabra es muy rica. Por ejemplo, considérense las siguientes palabras:

Etzekoa	-> etxe+ko+a
Etxea	-> etxe+a
Etxeak	-> etxe+ak
Lana	-> lan+a
Lanak	-> lan+ak
Atekoa	-> ate+ko+a
Atea	-> ate+a
Ateak	-> ate+ak

1) **Palabras** {etzekoa, etxea, etxeak, lana, lanak, atekoa, atea, atek} 9 palabras

2) **Morfemas** Su descomposición en morfemas con significación gramatical es la siguiente:

En esta descomposición se han utilizado los siguientes morfemas {etxe, lan, ate, ko, a, ak}; en total, 6 morfemas. Se observa cómo, en este caso tan simple, la propuesta de morfemas reduce el tamaño del vocabulario por lo que puede ser una opción válida de LUs para el euskera.

Se plantean varios problemas con la utilización de los morfemas como LUs en el sistema de CSR:

1) ¿Cómo automatizar adecuadamente el proceso de segmentación morfológica de los textos? Sin duda, es necesaria la utilización de una herramienta de segmentación morfológica automática de los textos para poder establecer las correspondencias entre los resultados del reconocimiento y la clasificación auténtica de la señal de voz. Esta segmentación, tiene que estar adaptada a las necesidades que son propias de las tareas de CSR para que la segmentación producida sea coherente con el mensaje

contenido en la señal de voz. En nuestro trabajo sobre el euskera utilizamos la herramienta *AHOSEG* (Lopez de Ipina *et al.*, 2002).

2) ¿Cómo componer de nuevo las palabras tras el proceso de reconocimiento basado en morfemas, para poder realizar la comparación de los resultados?. En este caso, utilizamos una herramienta simple de alineamiento a palabras.

3) Por último, ciertos problemas de confusión acústica pueden surgir debido a la estructura morfémica del euskera, ya que la frecuencia de acumulación del conjunto de morfemas muy cortos y/o acústicamente muy parecidos es alta (Lopez de Ipina *et al.*, 2002).

Por ejemplo, la (k) es un morfema con una alta RFO y significado sintáctico (denota el plural) pero problemático en cuanto a su reconocimiento acústico. Estos morfemas muy cortos y problemáticos no deberían ser segmentados.

2.2 Selección de LUs alternativas mediante criterios orientados a CSR

Nuestra propuesta, (López de Ipina *et al* 2000) consiste en realizar la selección basándonos en medidas orientadas a CSR. La metodología que se utilizará para generar estas nuevas LUs será la siguiente:

1) Utilizando la herramienta de segmentación *AHOSEG* se obtiene el vocabulario de pseudo-morfemas.

2) Se realizan experimentos de reconocimiento con el conjunto de LUs (pseudo-morfemas) obtenido.

3) En un refinamiento posterior, los pseudo-morfemas muy cortos y con un alto grado de confusión acústica no se segmentan. Además, las LUs tienen que tener una mínima frecuencia para entrenar adecuadamente el LM, por lo que aquellos pseudo-morfemas que no cumplan este requisito tampoco serán segmentados.

4) Para analizar los resultados se utilizará un alineamiento automático a palabras.

A modo de ejemplo analizaremos las palabras: *etxea*, *etxekoa*, *etxekoarena* y *etxekoarenak*). El lema de todas ellas es *etxea*.

La segmentación en morfemas sin restricciones sería:

Etxea	-> etxe+a
Etxekoa	-> etxe+ko+a
Etxearena	-> etxe+a+ren+a
Etxekoarenak	-> etxeko+a+ren+a+k

Segmentado de acuerdo a los anteriores criterios, obtendríamos:

Etxea	-> etxea
Etxekoa	-> etxekoa
Etxearena	-> etxea+rena
Etxekoarenak	-> etxekoa+renak

De manera que se aumenta el número de unidades con respecto al conjunto inicial de pseudo-morfemas, pero se reduce la posibilidad de confusión acústica entre las LUs. Denominaremos a esta nueva propuesta NOPALABRAS.

3 Experimentación

En esta sección se presenta la experimentación realizada con los diferentes conjuntos de LUs: palabras, pseudo-morfemas y LUs alternativas a la palabra. Primeramente, presentamos el marco experimental. A continuación, los criterios de evaluación y por último, la realización experimental. Para generar los modelos de LUs se utilizará un conjunto de unidades subléxicas incontextuales

de tipo fonético de, 28 unidades + un modelo de silencio modelado mediante HMMs discretos de 4 *codebooks*.

Para realizar los experimentos de reconocimiento y evaluar las diferentes propuestas de LUs se utilizarán las tareas MLA y VBE, (López de Ipina *et al* 2000):

1) La tarea MLA está compuesta por 400 frases grabadas por 20 locutores; 10 hombres y 10 mujeres. El vocabulario de la tarea es de: 47 palabras y 34 pseudo-morfemas.

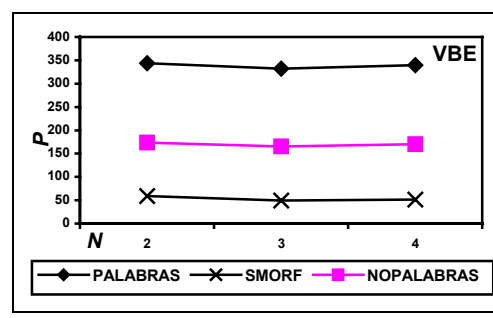
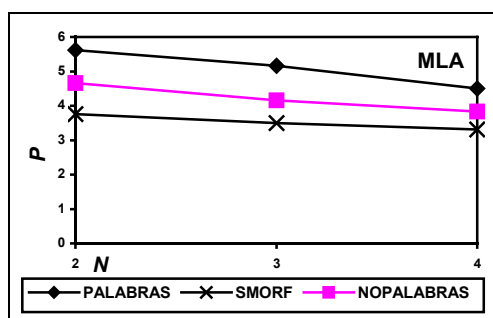
2) La tarea VBE está compuesta por 800 frases grabadas por 20 locutores; 10 hombres y 10 mujeres. El vocabulario de la tarea es de: 3.500 palabras y 1900 pseudo-morfemas.

Para los experimentos de reconocimiento se ha diseñado un subconjunto de la tarea VBE (MVBE). Este subconjunto tiene un tamaño de vocabulario de 550 en el caso de palabras y 490 unidades alternativas y 400 pseudo-morfemas. Por último, los textos de las tareas se han dividido de la siguiente manera para entrenar el LM:

1) MLA: 14.500 frases para entrenamiento y 500 frases para test.

2) VBE: 4.000 frases para entrenamiento y 500 frases para test.

La calidad de las propuestas de conjuntos de LUs será medida utilizando los criterios de evaluación que a continuación se describen y el dominio de experimentación anteriormente expuesto.



(a)

(b)

Figura 1 Medidas de perplejidad (P) para las tareas MLA y VBE con los tres conjuntos de unidades PALABRAS, SMORF y NOPALABRAS

Para realizar la experimentación se seguirá el siguiente proceso:

1) Evaluación sobre texto: Análisis de la Perplejidad (P) y la Perplejidad Fonética (PP) de un LM tipo N-grama sobre las tareas MLA y VBE.

2) Evaluación sobre voz: Extracción de la tasa %PRA (exactitud), RTR y %PTU sobre las tareas MLA y MVBE. Definiéndose el retraso respecto del tiempo real (RTR), como el ratio del tiempo de cálculo con respecto al tiempo real, el cual se estima debe ser inferior al tiempo de un frame (10 ms). El porcentaje de la tasa de reconocimiento con respecto al tiempo de cómputo del experimento %PTU, se calcula de acuerdo a la expresión:

$$\%PTU = \frac{\%PRA}{TiempoComputo} * 100$$

donde %PRA es tasa anteriormente calculada y $TiempoComputo$ es el tiempo de cómputo en ms/frame. Cuando las unidades de reconocimiento son las palabras, la tasa de %PRA directamente evalúa la bondad del conjunto de LUs. Cuando las LUs son inferiores a la palabra, se realizará una doble evaluación: con y sin alineamiento a palabras para poder comparar los resultados de reconocimiento de los diferentes conjuntos de unidades (Peñagarikano *et al.*, 1999). En la experimentación se analizarán los conjuntos de LUs planteados en la sección anterior:

1) **PALABRAS**: se utilizan como LUs todas las posibles palabras en el sentido estricto. Esta propuesta, como se ha mencionado anteriormente, puede provocar un vocabulario intratable a partir de tareas de complejidad media-alta.

2) **SMORF**: pseudo-morfemas. Estas unidades se obtienen mediante el segmentador automático *AHOSEG*.

3) **NOPALABRAS**: la nueva propuesta de LUs planteada en la sección anterior. En esta propuesta se han utilizado concretamente los siguientes criterios de segmentación morfémica:

a) Para reducir la confusión acústica, aquellos pseudo-morfemas de longitud menor que 3 caracteres o cuya tasa de confusión sea alta no serán segmentados. Para ello, se han definido reglas automáticas (basadas en las tablas de confusión del experimento de reconocimiento con el conjunto *SMORF*) dentro del segmentador *AHOSEG*.

b) Por otra parte, se asegurará que las nuevas unidades tienen frecuencias relativas mínimas que permitan entrenar adecuadamente LMs basados en las nuevas unidades. Las tareas se han adecuado a las nuevas LUs.

En esta nueva propuesta, la tarea MLA tiene un vocabulario de 40 LUs; la tarea VBE, 2.700 unidades y la tarea reducida MVBE, 500. Por lo tanto, con respecto a PALABRAS, se reduce el vocabulario alrededor de un %25 al aumentar el número de unidades con respecto a la propuesta SMORF.

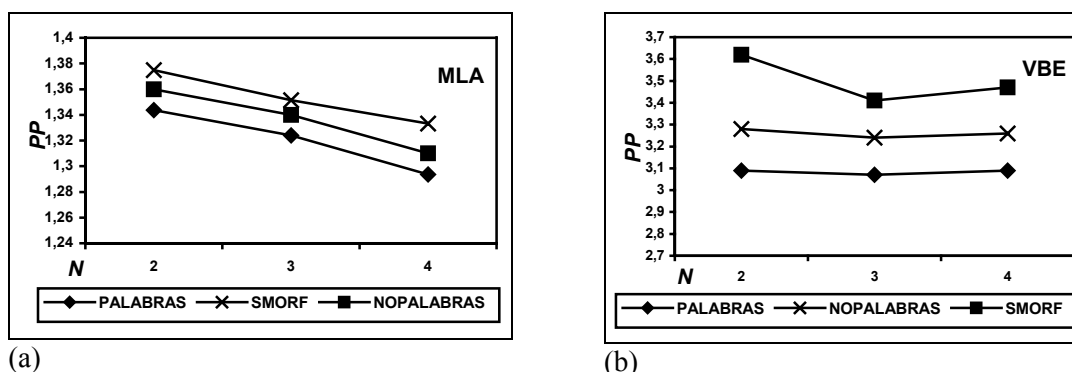


Figura 2 Perplejidad Fonética (PP) para las tareas LMA y VBE con los conjuntos PALABRAS, SMORF y NOPALABRAS

3.1 Medidas de perplejidad para un LM de tipo N-grama

Se han calculado medidas de perplejidad clásica (P) (figura 1) y perplejidad fonética (PP) (figura 2). Examinando los resultados en las dos figuras extraemos las mismas conclusiones:

El conjunto NOPALABRAS tiene unos valores de perplejidades situados entre los valores de los conjuntos PALABRAS y SMORF. Representa, por tanto, la solución intermedia que estamos buscando, al menos desde el punto de vista de la perplejidad.

Esta opción puede ser válida para diseñar LM apropiados para tareas de un tamaño de vocabulario medio o grande.

3.2 Experimentos de CSR

En la tabla 1 aparecen los resultados obtenidos en los experimentos, previamente al alineamiento a palabras. Estudiando los resultados extraemos las siguientes conclusiones:

1) Tanto para la tarea MLA como para la tarea MVBE, los mejores resultados se obtienen con el conjunto PALABRAS. De todas formas, la propuesta de unidades alternativas NOPALABRAS funciona adecuadamente en la medida en que mejora a SMORF en las dos tareas.

2) La segunda de las tareas MVBE presenta con NOPALABRAS unos resultados con un %PRA menor que MLA debido a la dificultad de la tarea, a la alta inserción y a la confusión entre las LUs, sobre todo las más cortas.

Tras el alineamiento a palabras (ver tabla 2), se pueden observar las siguientes mejoras:

1) El comportamiento de NOPALABRAS mejora en las dos tareas y sigue superando la tasa de reconocimiento de SMORF debido a un mejor reconocimiento acústico.

2) El estudio detallado de la salida producida por el sistema de reconocimiento CSR muestra que la inserción de unidades se reduce.

La tabla 3 muestra los resultados tras aplicar un LM de tipo bigrama:

1) La tasa de reconocimiento %PRA para NOPALABRAS mejora la obtenida por SMORF en las dos tareas y se acerca a la tasa de PALABRAS.

2) La tarea MVBE presenta una tasa %PRA aceptable también para este conjunto de unidades que, sin duda, mejoraría claramente mediante un buen LM.

Las tablas 4 y 5 presentan resultados de medidas de *TiempoComputo*, *RTR* y *%PTU* para las dos tareas:

1) Para ambas tareas se aprecia un mejor funcionamiento del subconjunto NOPALABRAS respecto a PALABRAS debido a una reducción en el número de LUs.

2) A pesar de que las medidas de cómputo sean ligeramente favorables a SMORF, sopesando tasa %PRA y medidas de Tiempo de Cómputo (*TiempoComputo*, *RTR* y *%PTU*) la alternativa del conjunto NOPALABRAS presenta los mejores resultados.

Tabla 1 %PRA para las dos tareas (MLA y MVBE) con los conjuntos de LUs PALABRAS, SMORF y NOPALABRAS antes del alineamiento

	MLA	MVBE
PALABRAS	80,61	43,71
NOPALABRAS	74,84	30,09
SMORF	60,29	28,98

Tabla 2 %PRA para las tareas MLA y MVBE con los conjuntos de LUs PALABRAS, SMORF y NOPALABRAS tras el alineamiento a palabras.

	MLA	MVBE
PALABRAS	80,61	43,71
PAL. desde NOPAL	76,30	32,60
PAL de SMORF	63,80	29,09

Tabla 3 %PRA para las tareas MLA y MVBE con los conjuntos de LUs, PALABRAS, SMORF y NOPALABRAS con un LM de bigramas

	MLA	MVBE
PALABRAS	91,34	48,44
NOPALABRAS	88,82	42,07
SMORF	82,38	39,85

Tabla 4 Tiempo de cómputo en ms por frame (tr=tiempo real) para las tareas MLA y MVBE con los conjuntos de LUs, PALABRAS, SMORF y NOPALABRAS

	MLA	MVBE
PALABRAS	6 (tr)	33 (3.3*tr)
NOPALABRAS	5 (tr)	28 (2.8 *tr)
SMORF	3 (tr)	25 (2.5*tr)

Tabla 5 %PTU para las tareas MLA y MVBE con los conjuntos de LUs, PALABRAS, SMORF y NOPALABRAS

	MLA	MVBE
PALABRAS	13,4	1,46
NOPALABRAS	14,96	1,50
SMORF	20,09	1,59

4 Conclusiones

Puesto que el euskera es una lengua aglutinante, la utilización de la palabra como unidad base para definir el vocabulario de la tarea, puede disparar considerablemente el tamaño de esta haciéndolo inviable. Entrenar apropiadamente un LM puede ser una tarea muy compleja, ya que las posibles combinaciones de palabras son enormes y la tasa de OOV muy alta.

La utilización de pseudo-morfemas obtenidos con el segmentador *AHOSEG* puede ser una buena opción para definir el vocabulario de la tarea porque reduce notablemente el número de LUs. El euskera posee una gran cantidad de morfemas cuyo reconocimiento acústico es problemático debido a su corta longitud y a que a menudo presentan sonidos difíciles de reconocer, como son las consonantes oclusivas.

Esto provoca una alta confusión acústica y una alta inserción.

La opción de los pseudo-morfemas funciona mejor cuando no se tratan de reconocer las unidades problemáticas anteriormente descritas. El segmentador morfológico debe adoptar reglas de segmentación morfológica que reduzcan estos casos. La nueva aproximación definida NOPALABRAS, es una buena propuesta puesto que combina una aceptable tasa de reconocimiento con un tiempo de cómputo más adecuado que la opción PALABRAS reduciendo en cierta forma la alta confusión acústica e inserción de unidades del conjunto SMORF. A pesar de que esta última opción (NOPALABRAS) mejora los resultados, todavía se deben explorar otro tipo de soluciones para tareas más complejas o para mejorar los resultados de PALABRAS. El alineamiento a palabras funciona apropiadamente pero se deben explorar metodologías más robustas de composición de LUs en palabras.

La utilización de bigramas mejora notablemente el funcionamiento del sistema incluso para el conjunto SMORF, donde se produce la mejora más evidente. Es necesario, por tanto, un buen LM de LUs de tipo morfema para subsanar los problemas anteriormente descritos aprovechando la información morfosintáctica que se encuentra en el interior de las palabras.

Por otra parte, el uso de modelos acústicos más potentes y menos sensibles al ruido puede mejorar los resultados obtenidos en este trabajo. Por ejemplo, el uso de HMM semicontinuos y continuos.

Pueden utilizarse otro tipo de estructuras de LUs como el léxico en árbol con un tratamiento previo de clases de morfemas. Se pueden utilizar unidades alternativas automáticas para mejorar los resultados de reconocimiento y se deberían buscar alternativas donde no sea necesario la definición de LUs.

Bibliografía

- (Alegria *et al.*, 1996) Alegria I., X. Artola, K. Sarasola y M. Urkia."Automatic morphological analysis of Basque". *Literary & Linguistic Computing*, vol,11, nº 4, pp. 193-203, Oxford University Press. 1996.

- (Carký, Geutner y Schultz, 2000) Carký, K., P. Geutner y T. Schultz. "Turkish LVCSR: Towards better Speech Recognition for Agglutinative Languages". *Proceedings of IEEE ICASSP*, pp. 1563–1566, Istanbul, 2000
- (Creutz y Lagus, 2002) Creutz M. and K. Lagus. "Unsupervised discovery of morphemes". *Proceedings of the Workshop on Morphological and Phonological Learning of ACL-02*, pp. 21-30, Philadelphia, Pennsylvania, 2002.
- (Larson *et al.*, 2000) Larson M., D. Willett, J. Koehler y G. Rigoll, "Compound splitting and lexical unit recombination for improved performance of a speech recognition system for German parliamentary speeches" *Proceedings of ICSLP*, Pekin, 2000.
- (Larson, 2001) Larson M. "Sub-Word-Based Language Models for Speech Recognition: Implications for Spoken Document Retrieval". *Workshop on Language Modeling and Information Retrieval*, CMU. Pittsburgh, Pennsylvania, USA 2001.
- (Lopez de Ipina *et al.*, 2000) Lopez de Ipina K., I. Torres, L. Oñederra, A. Varona, N. Ezeiza, M. Peñagarikano, M. Hernández y L.J. Rodriguez. "First Selection of Lexical Units for Continuous Speech Recognition of Basque". *Proceedings of ICSLP*, vol II, pp. 531-535. Beijing 2000.
- (Lopez de Ipina *et al.*, 2002) Lopez de Ipina K., N. Ezeiza, G. Bordel. y M. Graña. "Automatic Morphological Segmentation for Speech Processing in Basque" *Proceeding of IEEE TTS Workshop*. Santa Monica USA, 2002.
- (Mengusoglu y Ris, 2001) Mengusoglu E. y C. Ris, "Use of Acoustic Prior Information for Confidence Measure in ASR Applications", *Proceedings of EUROSPEECH*, Aalborg Denmark, 2001.
- (Ney, Essen y Kneser, 1994) Ney H., U. Essen y R. Kneser. "On structuring probabilistic dependencies in stochastic language modeling". *Computer Speech and Language*. 8 pp 1-38, 1994.
- (Otsuki *et al.*, 1999) Otsuki K., T. Matsuoka, T. Mori, K. Yoshida, Y. Taguchi, S. Furoi and K. Shirai. "Japanese large-vocabulary continuous-speech recognition using a newspaper corpus and broadcast news", *Speech Communication*. Vol 28, pp 155-166, 1999.
- (Peñagarikano *et al.*, 1999) Peñagarikano M., G. Bordel, A. Varona y K. Lopez de Ipina. "Using non-word Lexical Units in Automatic Speech Understanding", *Proceedings of IEEE, ICASSP99*, Phoenix, Arizona, 1999.
- (Schultz, Westphal y Waibel, 1997) Schultz, T., M. Westphal y A. Waibel. The Global-Phone Project: Multilingual LVCSR with Janus-3. In: *Proc. SQEL, 2nd Workshop on Multi-lingual Information Retrieval Dialogs*, pp. 20-27, Plzeň 1997.
- (Schultz y Waibel, 2001) Schultz T y A. Waibel. "Language Independent and Language Adaptive Acoustic Modeling for Speech Recognition". *Speech Communication*, 2001.
- (Sivola, Kurimo y Lagus, 2001) Sivola V., M. Kurimo y K. Lagus. "Large Vocabulary Statistical Language Modeling for Continuous speech Recognition in Finnish", *Proceedings of EUROSPEECH*