Procesamiento del Lenguaje Natural, núm. 43 (2009), pp. 85-92 recibido 1-05-2009; aceptado 5-06-2009

Algoritmo de inserción de pausas para una lengua declinada

Break insertion algorithm for an inflected language

Eva Navas
UPV/EHU
Alda. Urquijo s/n
eva.navas@ehu.es

Iñaki Sainz
UPV/EHU
Alda. Urquijo s/n
inaki@aholab.ehu.es

Jon Sanchez
UPV/EHU
Alda. Urquijo s/n
jon.sanchez@ehu.es

Ibon Saratxaga
UPV/EHU
Alda. Urquijo s/n
ibon.saratxaga@ehu.es

Inmaculada Hernáez
UPV/EHU
Alda. Urquijo s/n
inma.hernaez@ehu.es

Resumen: Los sistemas de conversión de texto en habla necesitan un módulo para insertar las pausas que no vienen indicadas en el texto. En este trabajo se desarrolla un algoritmo de inserción de pausas para euskara estándar que utiliza información morfológica, sintáctica y del caso gramatical. El algoritmo se ha evaluado tanto objetiva como subjetivamente y los resultados confirman que este tipo de información es adecuada para realizar la predicción de las pausas en euskara estándar. Los tres tipos de información son relevantes y contribuyen a la mejora de los resultados del algoritmo.
Palabras clave: Inserción de pausas, Etiquetado morfosintáctco, Conversión de texto en habla

Abstract: Text to speech synthesis systems must have a module to insert breaks that are not indicated in the text. In this paper an algorithm to insert breaks for standard Basque is developed. This algorithm uses morphology, syntax and information about the grammatical case. The algorithm has been evaluated both objectively and subjectively and the results confirm that this kind of information is suitable to make the prediction of the location of the breaks in standard Basque. All three types of information contribute to the improvements of the results of the algorithm.
Keywords: Break insertion, Morpho-syntactic labelling, Text to speech conversion

1 Introducción

Los sistemas de conversión de texto en habla (CTH) deben producir una señal de voz sintética que sea lo más natural posible. La correcta ubicación de las pausas en la señal sintética contribuye a lograr naturalidad y es crítica ya que las pausas tienen influencia en el proceso de asignación de duración a los sonidos, en la generación de la curva de entonación y en la transcripción fonética. Proporcionan una estructura rítmica a la señal de voz y en algunos casos llegan hasta a modificar el sentido de una frase (Hirschberg y Prieto, 1996).

Sin embargo, en la mayoría de los casos, sólo parte de las pausas están indicadas en el texto mediante signos de puntuación. Aunque es común hacer una pausa cuando en la lectura de un texto se encuentran algunos signos de puntuación como puntos, exclamaciones, interrogaciones, dos puntos, punto y coma y puntos suspensivos, otros signos de puntuación no llevan siempre aparejada una pausa. Este es el caso de las comas, paréntesis, comillas y guiones. Aunque hay algunas indicaciones sobre dónde es conveniente hacer una pausa, a menudo el lector las coloca donde lo cree más conveniente. Es por ello que todo sistema CTH necesita un módulo de inserción de pausas.

La tarea de insertar pausas automáticamente en un texto es difícil. Un campo en el que el uso de un algoritmo de inserción de pausas es crítico es el de la traducción voz-voz. El objetivo de un sistema de este tipo es permitir la comunicación oral entre personas que no

ISSN: 1135-5948

comparten una lengua en común. Para ello se hace uso de la tecnología de reconocimiento de voz para transcribir la señal original, se aplican técnicas de traducción automática para pasarlas a la lengua objetivo y finalmente, mediante un sistema CTH se produce la señal oral en la lengua final.

Los primeros modelos de inserción automática de pausas se basaban en reglas que localizaban el mejor punto para insertar una pausa en un texto basándose en el contexto de palabras que la rodeaban (Bachenko y Fitzpatrick, 1990). Liberman y Church (1991) insertan las pausas teniendo en cuenta la puntuación y la existencia de una palabra función seguida por una palabra contenido. Estos sencillos algoritmos de inserción de pausas se han usado en sistemas CTH como el MITalk (Allen et al., 1987) y el sistema multilingüe de telefónica (Castejón et al., 1994). Son modelos que han dado buenos resultados a pesar de su sencillez, pero no son adecuados para lenguas declinadas en las que se utilizan muy pocas palabras función.

En este trabajo se presenta el desarrollo y la evaluación de un algoritmo de inserción de pausas que pueda ser utilizado en aplicaciones de traducción voz-voz para euskara estándar, una lengua declinada.

2 Descripción del corpus

Dado que al iniciar el trabajo no existía ningún corpus adecuado para el estudio de la inserción de pausas en euskara estándar, se construyó uno con este fin siguiendo la metodología empleada por Hirschberg y Prieto (1996). Los textos que forman el corpus pertenecen a una revista y un periódico electrónicos y a diferentes páginas de Internet escritas en euskara estándar.

En este trabajo se asume que se hace una pausa siempre que en el texto esté presente uno de los signos de puntuación que actúan como frontera de frase: punto, punto y coma, dos puntos, exclamación, interrogación, y puntos suspensivos. En el contexto de un sistema de traducción voz-voz estas pausas se pueden insertar en el texto de entrada al sistema CTH de acuerdo a las pausas encontradas por un sistema de detección de actividad vocal. Existen diferentes trabajos en los que se han detectado con éxito las pausas de fin de frase a partir de la voz y el texto reconocido automáticamente (Kim y Woodland, 2001). Sin embargo para un sistema de reconocimiento automático de voz es muy difícil situar los signos de puntuación internos de la frase, incluso teniendo en cuenta información acústica de la señal original (Chen, 1999). En el contexto de un sistema de traducción voz-voz estas pausas no vendrán indicadas en el texto, o si lo están no serán fiables. Por ello, en este trabajo se eliminaron todos los signos de puntuación internos de las frases y la predicción de la ubicación de las pausas se hace sin esta información. De este modo se emulan los textos de entrada al sistema CTH cuando funciona como parte de un sistema de traducción voz-voz.

En este trabajo se predicen las pausas internas de la frase, ya que pueden situarse en lugares donde no hay ningún signo de puntuación que las indique, y las pausas finales de frase, ya que no todos los signos de puntuación relacionados con el final de frase lo indican siempre: los puntos se pueden encontrar en abreviaturas, en direcciones WEB, se pueden encontrar interrogaciones y exclamaciones en el interior de las frases, etc.

2.1 Etiquetado de las pausas

Las pausas fueron manualmente etiquetadas por un lingüista que marcó los puntos en los que era probable que un lector hiciera una pausa en la lectura en voz alta del texto. Sólo se consideró un nivel de pausa, por lo tanto cada frontera entre dos palabras fue etiquetada bien como ‘pausa’ o como ‘no pausa’. Las principales características del corpus una vez etiquetado se recogen en la Tabla 1.

Nº palabras	15.867
Nº frases	1.470
Nº pausas	3.589
Nº pausas internas	2.119
% palabras función	17.5

Tabla 1: Características del corpus

Como puede verse en la Tabla 1, únicamente hay un 17.5% de palabras función en el corpus. De ellas sólo el 12% están seguidas de pausa y un 17% están precedidas por una pausa. Por esta razón no es posible utilizar un algoritmo basado en la distinción entre palabras función y contenido para predecir la situación de las pausas en este corpus.

2.2 Etiquetado morfológico del corpus

La información del etiquetado morfológico (POS) ha sido tradicionalmente utilizada en los

sistemas de predicción de pausas para diferentes idiomas como el inglés (Black y Taylor, 1997), francés (Boula de Mareüil y d'Alessandro, 1998), euskara estándar (Navas et al., 2002), tailandés (Tesprasit et al., 2003), castellano (Bonafonte y Agüero, 2004), coreano (Kim et al., 2006), etc.

En este trabajo se ha utilizado la información del etiquetado POS obtenido automáticamente con el analizador morfológico para euskara MORFEUS (Ezeiza et al., 1998) desarrollado por el grupo IXA de la UPV/EHU[1]. Este analizador proporciona el lema de cada palabra y su categoría POS; en la mayor parte de los casos además proporciona la subcategoría y el caso. En este trabajo se ha utilizado la información de POS que ha demostrado ser influyente en la colocación de las pausas en numerosas lenguas, entre ellas el euskara y la información de caso, dado que al ser el euskara una lengua declinada se prevé que pueda tener influencia en la colocación de las pausas.

Etiqueta	Descripción	Seguida de pausa
ADB	Adverbio	27.8%
ADI	Verbo principal	17.0%
ADJ	Adjetivo	18.5%
ADL	Verbo auxiliar	49.4%
ADT	Verbo sintético	44.5%
BST	Otros	0.0%
DET	Determinante	16.5%
ERL	Sufijo de relativo	54.8%
IOR	Pronombre	3.7%
ITJ	Interjección	66.7%
IZE	Nombre	18.6%
LAB	Abreviatura	100.0%
LOT	Conjunción	18.0%
PRT	Partícula	11.4%
PUNT	Signo de puntuación	99.2%
SIG	Acrónimo	8.3%

Tabla 2: Etiquetas de categoría POS utilizadas

El conjunto de etiquetas proporcionadas por este analizador a nivel de categoría POS es demasiado detallado para los objetivos del trabajo y dado que utilizar características no significativas a la hora de entrenar un clasificador introduce ruido que puede degradar su funcionamiento (Read y Cox, 2004), las etiquetas se agruparon manualmente. El conjunto de etiquetas utilizado a nivel de categoría y caso se muestran en las Tablas 2 y 3 respectivamente. También se muestra en estas tablas el porcentaje de etiquetas de cada tipo que van seguidas de pausa en el corpus etiquetado.

Como puede verse en la Tabla 2, la categoría PUNT va seguida de pausa en el 99.2% de los casos. En realidad y teniendo en cuenta que se han eliminado los signos internos de puntuación, todos los signos restantes deberían ser seguidos por una pausa. Sin embargo, el porcentaje en el corpus no es del 100% debido a que varios puntos correspondientes a abreviaturas se han etiquetado erróneamente como signos de puntuación y no van seguidos de pausa. En esta misma tabla se observa que los pronombres (IOR) casi nunca aparecen seguidos por una pausa.

Etiqueta	Descripción	Seguida de pausa
ABL	Ablativo	23.2%
ABS	Absolutivo	26.8%
ABZ	Orientativo	16.7%
ABU	Terminativo	50.0%
ALA	Adlativo	26.2%
DAT	Dativo	24.7%
DESK	Descriptivo	0.0%
DES	Benefactivo	71.4%
ERG	Ergativo	34.2%
GEL	Genitivo para lugar	2.4%
GEN	Genitivo para persona	0.6%
INE	Inesivo	40.1%
INS	Instrumental	38.1%
MOT	Causal	52.0%
PAR	Partitivo	19.0%
PRO	Prolativo	20.0%
SOZ	Sociativo	27.0%

Tabla 3: Etiquetas de caso utilizadas

En la Tabla 3 se observa que el caso descriptivo y el genitivo prácticamente no se encuentran en el corpus seguidos de pausa. Además de las etiquetas listadas en esta tabla, en el caso se ha añadido la etiqueta NONE para las categorías que no lo deben tener y la etiqueta MISSING para las categorías que deberían haber llevado un caso asociado pero no fue asignado por MORFEUS.

2.3 Etiquetado sintáctico del corpus

En algunos trabajos sobre predicción de la ubicación de las pausas se ha utilizado también información sintáctica (Ingulfsen et al., 2005)(Koehn et al., 2000)(Ostendorf y Veilleux, 1994). Aunque no hay una relación directa entre sintaxis y estructura prosódica, sí

[1] http://ixa.si.ehu.es/Ixa

es cierto que ambas están muy relacionadas (Frazier et al., 2004). El corpus se etiquetó automáticamente para obtener información relacionada con la sintaxis utilizando el analizador sintáctico desarrollado por el grupo IXA (Aduriz et al., 2004). Esta herramienta etiqueta sintácticamente cada palabra y además las agrupa en sintagmas. Las pausas se producen generalmente entre sintagmas, por ello esta información se consideró muy importante para la predicción de su ubicación.

La información de nivel de sintagma proporcionada por el etiquetador automático se postproceso para producir las etiquetas mostradas en la Tabla 4.

Etiqueta	Descripción	Seguida de pausa
BEG	Inicial	1.2%
CEN	Central	3.9%
END	Final	24.5%
UNI	Una palabra	18.5%

Tabla 4: Etiquetas de sintagma utilizadas

En la Tabla 4 se ve que las palabras en posición inicial o central en el sintagma tienen muy poca probabilidad de ser seguidas por pausa. De hecho, los casos en que esto ocurre se deben a errores de etiquetado.

Etiqueta	Descripción	Seguida de pausa
AL	Complemento verbal	23.8%
JAL	Verbo auxiliar	25.5%
JAL_M_AL	Verb. aux. como compl. verbal	50.0%
JAL_M_O	Verb. aux. como objeto subordinado	66.7%
JAN	Verb. principal	9.3%
JAN_MP	Verb. subordinado	29.1%
JAN_M_AL	Verb. subord. como compl. verbal	8.7%
JAN_M_KM	Verb. subord. como nombre	0.0%
JAN_M_IL	Verb. subord. como compl. adj.	2.5%
JAN_M_O	Verb. subord. como objeto	21.2%
JAN_IL	Verb. ppal como compl. adj.	10.9%
LOK	Coordinada	59.2%
MP	Subordinada	100.0%
IL	Complemento adj.	4.8%
OBJ	Objeto	14.4%
PJ	Conjunción coordinante	1.2%
SUBJ	Sujeto	25.9%
ZOBJ	Objeto indirecto	10.8%

Tabla 5: Etiquetas de sintaxis utilizadas

De la misma manera que con el etiquetado POS, el conjunto de etiquetas sintácticas se redujo dando como resultado el conjunto mostrado en la Tabla 5.

3 Medidas de error

Para evaluar el funcionamiento de los algoritmos de inserción de pausas es necesario definir los parámetros con los que se van a medir sus resultados. En este trabajo se han utilizado medidas objetivas y se ha definido también un procedimiento subjetivo para realizar la evaluación.

3.1 Medidas objetivas

A la hora de evaluar objetivamente un algoritmo de inserción de pausas, cada frontera entre palabras mal clasificada se considera un error. Hay dos tipos posibles de error: inserciones (I) cuando se clasifica como 'pausa' una frontera que no estaba etiquetada como tal y omisiones (O) cuando se clasifica como 'no pausa' una frontera que estaba etiquetada como 'pausa' en el corpus. La puntuación total se define habitualmente como la proporción de fronteras bien clasificadas con respecto al número total de fronteras (N) y se calcula de acuerdo con la expresión (1).

$$S(\%) = \frac{N - O - I}{N}.100 \qquad (1)$$

Este dato deber ser interpretado con cuidado, ya que depende de la proporción de fronteras de cada clase presentes en el corpus. En nuestro caso el 77.4% de las fronteras pertenecen a la clase 'no pausa', por lo que un algoritmo que no insertara ninguna pausa obtendría una puntuación total del 77.4%. Para evitar este problema y poder hacer comparaciones con los resultados obtenidos en otros trabajos se ha propuesto el uso del estadístico kappa (κ) sugerido inicialmente por Carletta (1996) para tareas de clasificación lingüística y usado en varios trabajos de inserción de pausas (Navas et al., 2002). Este estadístico se calcula de acuerdo con la expresión (2), en la que se compara la puntuación total obtenida por el algoritmo (S), definida según la expresión (1), con la proporción inicial de fronteras etiquetadas como 'no pausa' (NPT/N), eliminando de esta manera la dependencia del dato de dicha proporción.

$$\kappa = \frac{S - \frac{NPT}{N}}{1 - \frac{NPT}{N}} \quad (2)$$

En la expresión (2) S es la puntuación total y NPT el número total de fronteras etiquetadas como 'no pausa' en el corpus. Si el algoritmo no inserta ninguna pausa, el valor de κ es 0. Si el método predice todas las fronteras correctamente, κ es 1. Valores negativos de κ indican que el método predice pausas, pero no en los lugares correctos.

Otras medidas apropiadas para evaluar los resultados de los algoritmos de inserción de pausas son la precisión (P) y la *recall* (R), ampliamente utilizadas en el campo de recuperación de información para evaluar la eficiencia de las búsquedas. Estos valores y su media armónica conocida como *F-score* (F) se han aplicado también en la evaluación de algoritmos de inserción de pausas (Sun y Applebaum, 2001). En este contexto, la precisión mide la proporción de pausas bien identificadas entre todas las fronteras a identificar y la *recall* la proporción de pausas bien identificadas entre todas las fronteras marcadas como 'pausa' en el corpus. La *F-score* se calcula de acuerdo a la expresión (3).

$$F = \frac{2P.R}{(P+R)} \quad (3)$$

Si el algoritmo no inserta ninguna pausa, R es 0 y tanto P como F están indefinidas. Si el algoritmo inserta pausas, pero en lugares incorrectos P y R son 0 y F está indefinida.

En este trabajo se han utilizado κ y *F-score* como medidas objetivas del funcionamiento de los algoritmos, ya que ambas independizan los resultados de la proporción inicial de clases en el corpus y facilitan la comparación de los resultados.

3.2 Medidas subjetivas

Medir el adecuado funcionamiento de un algoritmo de inserción de pausas comparando sus decisiones con el etiquetado de referencia tiene el inconveniente de que la importancia subjetiva del error no se tiene en cuenta. Para evitar este efecto se diseñó una evaluación subjetiva de los resultados del algoritmo en la que se seleccionaron aleatoriamente 30 frases de la parte del corpus reservada para pruebas y se pidió a diferentes evaluadores que insertaran las pausas donde las creyeran necesarias. Su etiquetado se compara después con los resultados del algoritmo a evaluar y se clasifican las pausas en tres tipos:

- Pausas correctas: pausas insertadas por el algoritmo que también han sido indicadas por algún evaluador.
- Pausas omitidas: pausas indicadas por cualquiera de los evaluadores, que el algoritmo no ha insertado.
- Pausas incorrectas: pausas insertadas por el algoritmo que ninguno de los evaluadores ha indicado.

Para medir los resultados del algoritmo de inserción se calculan las siguientes medidas:

- Número absoluto de pausas correctas, omitidas e incorrectas (Nº). Este número debe ser alto para las pausas correctas y bajo para el resto.
- Número de pausas seleccionadas por al menos el 80% de los evaluadores (<80%). Esta medida se calcula para las pausas correctas y las omitidas y da una idea de qué pausas son importantes para la mayoría de los evaluadores.
- Puntuación media de las pausas, entendida como el número medio de evaluadores que indican esa pausa (Media). Esta medida se calcula para las pausas correctas y omitidas. Para las pausas correctas debe ser alta ya que esto indica que gran parte de los evaluadores están de acuerdo con la pausa. En el caso de las pausas omitidas debe tener un valor pequeño, dado que esto indica que pocos evaluadores seleccionaron la pausa.

4 Algoritmo de predicción

La inserción de pausas en un texto es un problema de clasificación en el que se debe decidir si cada frontera entre dos palabras incluye o no una pausa. En este trabajo sólo se ha considerado un nivel de pausa, por lo cual sólo hay dos clases posibles: 'pausa' y 'no pausa'.

Como técnica de clasificación se han utilizado árboles de clasificación (CART), construidos aplicando validación cruzada de 10 bloques y teniendo en cuenta las probabilidades iniciales de las clases en el corpus. En todos los casos el 75% de los datos disponibles se han utilizado para el entrenamiento del algoritmo y el 25% restante para las pruebas y a los dos tipos de error de clasificación posibles se les ha asignado el mismo peso.

Además, se ha definido un sistema ficticio de referencia que inserta una pausa en todos los finales de frase. De este modo se pueden comparar los resultados de los algoritmos entrenados con los de este sistema que tendría unos resultados mínimos.

Se han construido diferentes árboles utilizando en cada uno de ellos únicamente uno de los tipos de información descritos en las secciones 2.2 y 2.3 para evaluar el poder de predicción de cada tipo de etiqueta. También se ha construido un árbol utilizando los mejores subconjuntos de 2 y 3 etiquetas, que resultaron ser la sintaxis y el POS y la sintaxis, el POS y la información de sintagma respectivamente. Finalmente se ha entrenado un árbol que utiliza los cuatro tipos de información. En todos ellos, además de la correspondiente información morfológica y sintáctica, se ha considerado el número de sílabas restantes hasta el siguiente signo de puntuación de fin de frase como factor de predicción. Las etiquetas morfológicas y sintácticas se han considerado en una ventana de longitud 5 alrededor de la palabra que precede a la frontera a predecir, es decir, se tienen en cuenta las etiquetas de la palabra actual y de las dos anteriores y posteriores.

5 Resultados

5.1 Medidas objetivas

En la Tabla 6 se muestran los resultados de los árboles construidos utilizando únicamente un tipo de información de etiquetado. El que obtiene mejores resultados es el que considera la información sintáctica descrita en el apartado 2.3, es decir, éste es el tipo de información más influyente en la colocación de las pausas de las estudiadas en este trabajo. Teniendo en cuenta estos resultados, el POS es la siguiente característica a considerar, seguida de la información de sintagma y finalmente del caso.

Tipo info.	κ	F
POS	0.429	0.614
Caso	0.253	0.478
Sintaxis	0.498	0.713
Sintagma	0.346	0.571

Tabla 6: Resultados de los árboles entrenados con un solo tipo de información

Los valores de κ y *F-score* de los árboles construidos usando 1, 2 y 3 características, así como del árbol que considera los cuatro tipos de información se muestran en las Figuras 1 y 2 respectivamente, comparándolos con los resultados que obtiene el algoritmo de referencia que únicamente inserta pausas en los finales de frase. Todos ellos consiguen mejores resultados que el sistema de referencia, y cada tipo de información que se añade mejora los resultados del árbol anterior. El algoritmo para predecir de pausas que mejores resultados objetivos obtiene es el que emplea los cuatro tipos diferentes de información considerados, por ello es éste el que se sometió al proceso de evaluación subjetiva.

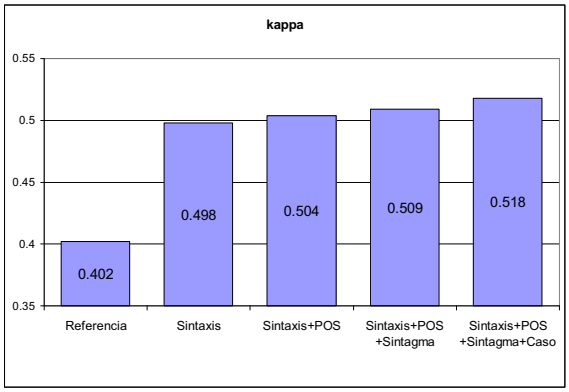

Figura 1: Valores de κ para los árboles construidos considerando diferente número de tipos de información

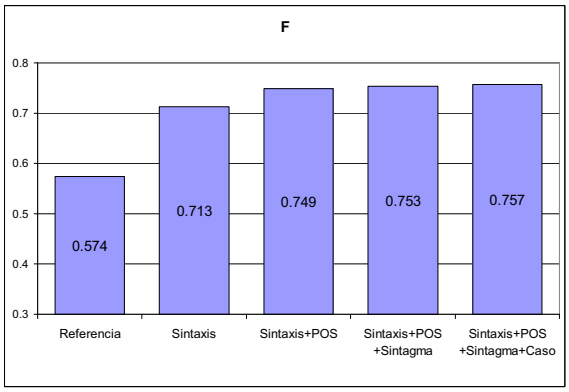

Figura 2: Valores de F para los árboles construidos considerando diferente número de tipos de información

5.2 Medidas subjetivas

En la prueba subjetiva del algoritmo tomaron parte 15 personas, todas ellas hablantes de euskara estándar con nivel fluido en dicha lengua.

El algoritmo desarrollado ha obtenido 17 pausas incorrectas, es decir, únicamente 17 de las pausas introducidas por el algoritmo no han sido seleccionadas por ninguno de los evaluadores. Para evaluar este dato hay que tener en cuenta que incluso el etiquetado de referencia obtiene pausas incorrectas, en este caso 3, ya que la inserción de pausas es un proceso que depende mucho de la persona y no hay acuerdo total sobre ellas.

En la Tabla 7 se muestran los resultados del algoritmo referidos a las pausas correctas, comparadas con los valores obtenidos por el etiquetado de referencia. El etiquetado de referencia tiene 79 pausas correctas, con una puntuación media de 0.68. De ellas 39 han sido seleccionadas por al menos el 80% de los evaluadores, lo que indica que hay un gran acuerdo sobre su conveniencia. El algoritmo desarrollado consigue insertar 62 pausas correctas de las cuales 34 son consideradas muy importantes por los evaluadores. Además la puntuación media de las pausas insertadas es 0.71, es decir, que las pausas correctas que inserta son las más importantes.

	Nº	>80%	Media
Etiquetado	79	39	0.68
Algoritmo	62	34	0.71

Tabla 7: Resultados subjetivos para pausas correctas

En la Tabla 8 se recogen los resultados referidos a pausas omitidas. Como era de esperar, el etiquetado de referencia omite menos pausas que el algoritmo desarrollado, y además la importancia media de esas pausas omitidas es menor, pero la diferencia entre ambos no es grande.

	Nº	>80%	Media
Etiquetado	52	6	0.30
Algoritmo	69	11	0.37

Tabla 8: Resultados subjetivos para pausas omitidas

6 *Conclusiones*

Para predecir la localización de las pausas en euskara estándar la información morfosintáctica es determinante, pero también es importante considerar el caso gramatical.

La comparación de resultados obtenidos en diferentes trabajos de predicción de la localización de las pausas es muy difícil, debido por un lado a las diferencias en las bases de datos utilizadas y por otro al distinto tratamiento de los signos de puntuación que se realiza en cada uno de ellos. Algunos trabajos utilizan la información de los signos de puntuación que facilita la tarea ya que está muy correlada con las pausas. Otros eliminan completamente los signos de puntuación. En este trabajo, debido a las especiales características de funcionamiento de los sistemas TTS dentro de un sistema de traducción voz-voz, se han mantenido los signos de puntuación de final de frase y se han eliminado todos los signos internos. Como indicación se dan los resultados obtenidos en trabajos recientes para lenguas declinadas como el de Yoon (2006), que utilizando CARTs con información sintáctica y de POS alcanza una F del 0.71 para el coreano, utilizando información sobre la puntuación. Zervas et al. (2005) también utilizan árboles de decisión e información de etiquetado POS para el griego y consiguen una $k > 0.75$.

Tras realizar la evaluación subjetiva se pone de manifiesto que el proceso de introducir pausas en un texto es muy dependiente de la persona. No hay un acuerdo total a la hora de decidir dónde colocarlas. Algunas de ellas sí son muy claras, y la mayor parte de los evaluadores las seleccionan (en los textos utilizados para la evaluación subjetiva el 21% de las pausas fueron marcadas por más del 80% de los evaluadores), pero otras son muy dependientes de la persona (el 19% de las pausas fue marcada por menos del 20% de los evaluadores).

7 *Agradecimientos*

Queremos agradecer a Nerea Ezeiza y al grupo IXA el etiquetado automático del corpus a nivel morfológico y sintáctico, a Iñaki Gaminde el etiquetado manual de las pausas y a todos los evaluadores su participación en la evaluación subjetiva.

Este trabajo ha sido parcialmente financiado por el Ministerio de Educación y Ciencia dentro del proyecto AVIVAVOZ (TEC2006-13694-C03-02, www.avivavoz.es) y por el Gobierno Vasco en su subvención a grupos de investigación del sistema universitario vasco (IT-444-07).

Bibliografía

Allen, J., D. Byron, M. Dzikovska, G. Aduriz I., Aranzabe M., Arriola J., Díaz de Ilarraza A., Gojenola K., Oronoz M., Uria L., 2004 A cascaded syntactic analyser for Basque *LNCS*, 2945: 124-135.

Allen, J., Hunnicut, S., Klatt, D. 1987. *From text to speech: the MITalk system*. Cambridge University Press, Cambridge.

Bachenko, J., Fitzpatrick, E., 1990 A Computational grammar of discourse-neutral prosodic parsing in English. *Computational Linguistics*, 16(3): 155-170.

Black, A.W.; Taylor, P., 1997. Assigning phrase breaks from part-of-speech sequences, en *Proceedings of Eurospeech'97*, páginas 995-998.

Boula de Mareüil, P., d'Alessandro, C., 1998. Text chunking for prosodic phrasing in French, en *Proceedings of 3rd ESCA/COCOSDA International Workshop on Speech Synthesis*, páginas 127-132.

Bonafonte, A., Agüero, P.D. 2004. Phrase break prediction using a finite state transducer, en *Proceedings de AST*.

Carletta, J. 1996. Assessing agreement on classification task: the kappa statistic. *Computational Linguistics*, 22(2):249-254.

Castejón, F., Escalada, J.G., Monzón, L., Rodríguez, M.A., Sanz, P., 1994. Un conversor texto-voz para español. *Comunicaciones de Telefónica I+D*, 10, 8.

Chen, C. J., 1999. Speech recognition with automatic punctuation, en *Proceedings of Eurospeech*, páginas 447-450.

Ezeiza N., Aduriz I., Alegria I., Arriola J.M., Urizar R., 1998. Combining stochastic and rule-based methods for disambiguation in agglutinative languages, en *Proceedings of COLING-ACL*, pp 379-384.

Frazier, L., Clifton, C., Carlson, K., 2004. Don't break, or do: prosodic boundary preferences. *Lingua*, 114(1): 3-27.

Hirschberg, J., Prieto, P., 1996. Training intonational phrasing rules automatically for English and Spanish text-to-speech. *Speech Communication*, 18: 281-290.

Ingulfsen, T., Burrows, T., Buchholz, S., 2005. Influence of syntax on prosodic boundary prediction, en *Proceedings of Interspeech*, páginas 1817-1820.

Kim, J., Woodland, P. C., 2001. The use of prosody in a combined system for punctuation generation and speech recognition, en *Proceedings of Eurospeech*, páginas 2757-2760.

Kim, S., Lee, J., Kim, B., Lee, G., 2006. Incorporating second-order information into two-step major phrase break prediction for Korean, en *Proceedings of Interspeech*, paper 1487.

Koehn, P., Abney, S., Hirschberg, J., Collins, M., 2000. Improving intonational phrasing with syntactic information, en *Proceedings of IEEE ICASSP*, páginas 1289-1290.

Liberman, M., Church, K., 1991. Text analysis and word pronunciation in text-to-speech synthesis, *Advances in Speech Signal Processing*, Dekker, New York, páginas 791-831.

Navas, E., Hernáez, I., Ezeiza, N., 2002. Assigning phrase breaks using CART's in Basque TTS, en *Proceedings of Speech Prosody*, páginas 527-531.

Ostendorf, M., Veilleux, N., 1994. A hierarchical stochastic model for automatic prediction of prosodic boundary location. *Computational Linguistics*, 20(1), páginas 27-54.

Read, I., Cox, S., 2004. Using part-of-speech for predicting phrase breaks, en *Proceedings of Interspeech*, páginas 741-744.

Sun, X., Applebaum, T.H., 2001. Intonational Phrase break prediction using decision tree and N-gram model, en *Proceedings of Eurospeech*, páginas 537-540.

Tesprasit, V., Charoenpornsawat, P., Sornlertlamvanich, V., 2003. Learning phrase break detection in Thai text-to-speech, en *Proceedings of Eurospeech*, páginas 325-328.

Yoon, K., 2006. A prosodic phrasing model for a Korean text-to-speech synthesis system. *Computer Speech & Language*, 20(1): 69-79.

Zervas, P., Xydas, G., Fakotakis, N., Kokkinakis, G., Kouroupetroglou, G., 2005. Experimental evaluation of tree-based algorithms for intonational breaks representation. *LNCS*, 3658: 334-341.