Procesamiento del Lenguaje Natural, Revista nº 49 septiembre de 2012, pp 217-220 recibido 07-05-12 revisado 01-06-12 aceptado 05-06-12

Sistema SAGAS: herramienta de soporte al subtitulado para personas sordas

Julio Villena[1], Lourdes Moreno[2], Paloma Martínez[2], José Carlos González[1]
[1]Daedalus – Data, Decisions and Language, S.A.
[2] Grupo LaBDA, Dpto. de Informática, Universidad Carlos III de Madrid
{jvillena, jmartinez, jgonzalez}@daedalus.es, {lmoreno, pmf}@inf.uc3m.es

Resumen: Siguiendo legislación en España, en televisión se deben alcanzar unas cuotas en el servicio de subtitulado para personas sordas, además, los subtítulos deben elaborarse siguiendo normativa. Este marco regulador conlleva una demanda de tecnología que facilite a los radiodifusores y productores de contenido la generación de subtitulado, como es la generación automática de subtitulado a partir de reconocimiento de audio. En este trabajo se presenta "SAGAS, Sistema Avanzado de Generación Automática de Subtítulos", que proporciona subtítulos adecuados a la norma española para contenido vídeo que vaya acompañado de un guión o transcripción.

Palabras clave: subtitulado, subtitulado para personas sordas, reconocimiento automático del habla, ASR, alineamiento

Abstract: Following legislation in Spain specific TV quotas must achieved in Subtitling Service for deaf people and additionally subtitles should be developed according to regulations. This regulatory framework implies a technology demand to facilitate broadcasters and content producers to generate subtitling, like automatic subtitling from Automatic Speech Recognition (ASR). This paper introduces "SAGAS prototype (Advanced System for automatic generation of subtitles)" which provides subtitles according to Spanish standard from an audio and a transcript.

Keywords: subtitling, subtitling for deaf people, Automatic Speech Recognition, alignment

1 Introducción

Las personas con sordera necesitan de subtitulado para poder acceder a los contenidos audiovisuales que se ofrecen en televisión (TV), entre otros medios. La TV es el medio de comunicación más influyente en la sociedad y es esencial asegurar el acceso a todos los ciudadanos. Dicho derecho está legislado en España y hay cuotas de servicio de subtitulado impuestas por ley a los radiodifusores y productores de contenidos de TV (BOE, 2010). Además, este subtitulado debe seguir buenas prácticas en su elaboración cumpliendo con la normativa en España (AENOR, 2003).

Para que haya un cumplimiento de este marco regulador, se debe disponer de tecnología de generación automática de subtitulado, ya que los procesos manuales son inviables por los excesivos costes. Por otro lado, las tecnologías involucradas, como las de reconocimiento del habla, tienen problemas tecnológicos aun sin resolver, sobre todo en el caso de reconocimiento del audio en tiempo real con independencia del locutor. La falta de tecnología que dé soporte a los subtituladores, junto con la gran demanda de contenido subtitulado en TV en España, ha motivado este trabajo.

El trabajo que aquí se presenta, sistema SAGAS, proporciona una herramienta que genera de manera automática subtítulos, adecuados a la normativa española, a partir de un audio y una transcripción con información del audio.

Esta investigación se enmarca en el proyecto Sistema "Avanzado de Generación Automática

ISSN 1135-5948

de Subtítulos, SAGAS" (TSI-020100-2010-184)[1] cofinanciado por el Ministerio de Industria, Energía y Turismo, dentro del Plan Nacional de Investigación Científica, Desarrollo e Innovación Tecnológica 2008-2011.

Este proyecto supone una oportunidad para los agentes involucrados en la comercialización de contenidos digitales, con el objetivo de mejorar sus procesos productivos, simplificando y optimizando los procesos de subtitulado. El consorcio responsable del proyecto es un equipo multidisciplinar de expertos compuesto por: DAEDALUS[2], empresa referente en el sector de la producción audiovisual que coordina el proyecto y se encarga del desarrollo técnico, grupo de investigación LaBDA[3] de la Universidad Carlos III de Madrid encargado de la validación que permite asegurar los mejores resultados en cuanto a la calidad del proceso y la adecuación a normas de subtitulado y estándares de accesibilidad y por último, RTVE con el valioso rol de usuario. RTVE aporta al proyecto su experiencia y archivo de contenidos multimedia, que mantiene como fruto de todos los años de actividad en el sector, su colaboración y visión resulta imprescindible a la hora abordar el proyecto.

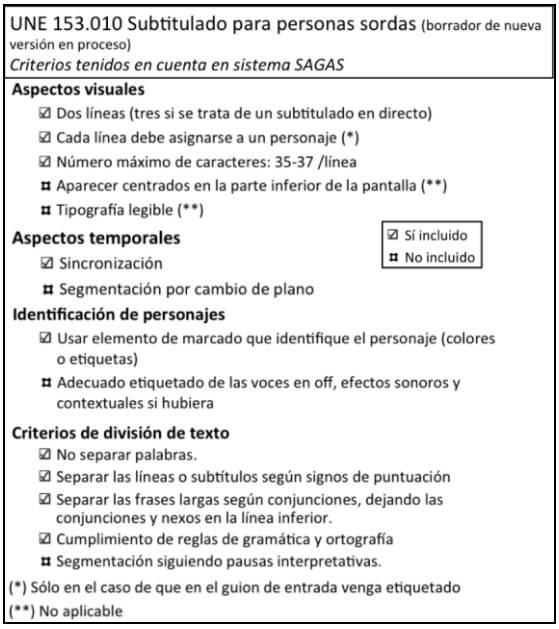
2 Trabajos relativos

Se encuentran sistemas de subtitulado asistidos por motores de reconocimiento automático de audio, algunos de estos parcialmente asistidos por operadores humanos (Boulianne et al., 2006), y otros que obtienen una transcripción del audio sin ningún tipo de asistencia humana (Neto et al, 2008). Estos últimos conllevan problemas como una alta tasa de error en el Módulo de Reconocimiento automático del Habla y retardo en la generación de los subtítulos (Meinedo, Viveiros y Neto, 2008).

En el panorama de TV en España, se encuentran varios sistemas automáticos de generación de subtitulado (Álvarez, del Pozo y Arruti, 2010) (Bordel et al, 2011). Cada uno de ellos utiliza distintos recursos de Tecnologías del Lenguaje para obtener mejores resultados en el Módulo de Reconocimiento. Al igual que SAGAS, se encuentran trabajos relativos que parten de transcripciones de textos más o menos equivalentes a la trascripción del audio, donde se hace un reconocimiento automático del habla en tiempo real para alinear temporalmente el texto y la voz (García et al, 2009).

3 Aplicación de Norma UNE 153.010 de subtitulado

La generación automática de los subtítulos en España no se puede llevar a cabo con otras herramientas internacionales ya que, siguiendo la norma UNE 153.010, hay requisitos propios que las distinguen. La adecuación a la normativa es un valor añadido de este trabajo. Se ha tenido en cuenta la versión del 2003, así como la nueva versión en borrador, que será el nuevo estándar a seguir próximamente. En esta nueva versión se han tenido en cuenta otros escenarios de aplicación y se han incorporado requisitos nuevos. Así se distingue entre aspectos visuales, temporales, de identificación de locutores y criterios de división de texto a cumplir.


UNE 153.010 Subtitulado para personas sordas (borrador de nueva versión en proceso)

Criterios tenidos en cuenta en sistema SAGAS

Aspectos visuales
- ☑ Dos líneas (tres si se trata de un subtitulado en directo)
- ☑ Cada línea debe asignarse a un personaje (*)
- ☑ Número máximo de caracteres: 35-37 /línea
- ⌗ Aparecer centrados en la parte inferior de la pantalla (**)
- ⌗ Tipografía legible (**)

Aspectos temporales

☑ Sí incluido
⌗ No incluido

- ☑ Sincronización
- ⌗ Segmentación por cambio de plano

Identificación de personajes
- ☑ Usar elemento de marcado que identifique el personaje (colores o etiquetas)
- ⌗ Adecuado etiquetado de las voces en off, efectos sonoros y contextuales si hubiera

Criterios de división de texto
- ☑ No separar palabras.
- ☑ Separar las líneas o subtítulos según signos de puntuación
- ☑ Separar las frases largas según conjunciones, dejando las conjunciones y nexos en la línea inferior.
- ☑ Cumplimiento de reglas de gramática y ortografía
- ⌗ Segmentación siguiendo pausas interpretativas.

(*) Sólo en el caso de que en el guion de entrada venga etiquetado
(**) No aplicable

Figura 1. Criterios UNE 153.010 aplicados en Sistema SAGAS

Tal como indica la figura 1, el sistema SAGAS se ha implementado para que los subtítulos generados cumplan con la mayoría de los requisitos de esta nueva versión de la norma.

[1] http://labda.inf.uc3m.es/sagas/esp/
[2] http://www.daedalus.es/
[3] http://labda.inf.uc3m.es/

4 Sistema SAGAS

La contribución consiste en el desarrollo de un prototipo que da soporte al subtitulado automático de contenido audiovisual en diferido para diversos medios: TV, Internet y dispositivos móviles. El contexto es el del subtitulado en diferido, *off-line* (fuera de línea) o enlatado, en cuyo caso no se trata de un proceso en tiempo real sino que se realiza previamente, sin necesidades temporales.

4.1 Arquitectura del sistema

El prototipo de herramienta de subtitulado que aquí se presenta puede ser utilizada bien como una herramienta aislada (*standalone)* o bien como un módulo en una herramienta comercial actualmente utilizada en la producción de subtítulos (como FAB Subtitler, por ejemplo).

A través de la interfaz de usuario, la herramienta toma como entrada el guión del material audiovisual en cuestión y el vídeo de dicho material. La salida está formada por el vídeo de entrada en el que se habrán integrado los subtítulos sincronizados, construidos a partir del guión, en los instantes de tiempo correspondientes, almacenados en formato estándar como EBU o formatos XML.

La arquitectura tiene un diseño modular, de forma que el sistema comprende de manera desacoplada el Módulo de Reconocimiento de Habla y el Módulo de Alineación, además de tener una interfaz para la entrada de datos y otra para la salida de subtítulos. Este diseño permite que cada módulo pueda ser utilizado por separado por aplicaciones externas, simplificando además el mantenimiento de los mismos.

Para el **Módulo de Reconocimiento del habla**, se han generado nuevos modelos en los motores de reconocimiento o ASR que no lo incorporan de manera nativa para obtener mejores resultados en el proceso de reconocimiento. En el prototipo actual, se han utilizado como motores la API del motor de reconocimiento de *MMIndexer* y *Dragon Naturally Speaking*. Este módulo da cómo salida un archivo de índices.

El **Módulo de Alineación** es mostrado en la Figura 2. En este módulo se da un proceso automático de alineación del guión o trascripción con el audio mediante la adición de marcas de tiempo, así como el tratamiento de revisión de errores y la segmentación del texto en subtítulos conformes a la norma en España y criterios de calidad vistos (ver apartado 3).

Tal como muestra la figura 2, como entrada se tiene la transcripción y el archivo de índices obtenido tras el reconocimiento de audio (Resultado Módulo de Reconocimiento).

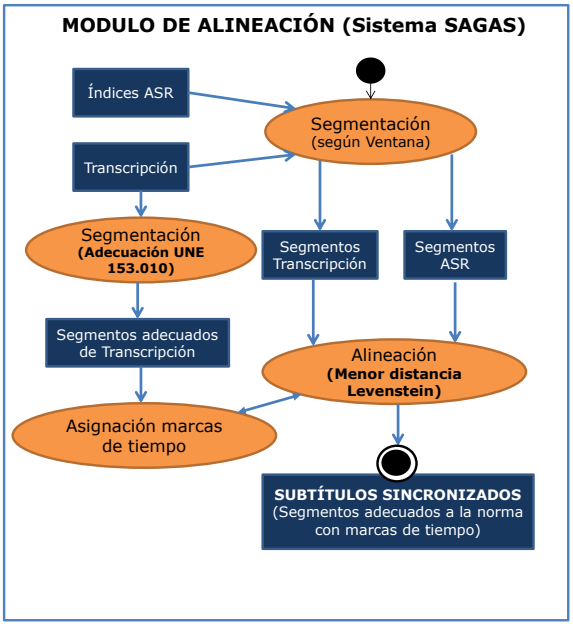


Figura 2: Diagrama de actividades de Módulo de alineación de Sistema SAGAS

De ambos recursos se hace una agrupación de las palabras o segmentación hasta llegar a un determinado tamaño de ventana de comparación. A continuación, se hace una comparación de las ventanas obtenidas por palabras de la transcripción con el de palabras reconocidas del ASR. Estos segmentos se alinean cuando la comparación entre ventanas da una distancia menor, utilizando la distancia de Levenshtein (Levenshtein, 1966).

A esta alineación se le incorporan marcas de tiempo para que los segmentos de subtítulos sigan la UNE 153.010. Para llevar a cabo este último proceso, la transcripción previamente ha sido segmentada siguiendo criterios de la norma como que el número de caracteres de cada segmento no supere el tamaño máximo fijado en 37, aunque se puede configurar con otros valores distintos a la norma.

Como resultado del módulo, se obtiene una estructura de datos que contiene en cada posición: los subtítulos a mostrar, el instante temporal asociado, y el locutor asignado.

Por último, tomando toda esta información, se genera la salida o subtítulos sincronizados en un **Interfaz de salida** tal como se muestra en la figura 3 que reproduce los resultados (reproducción de vídeo con los subtítulos

generados), y permite la descarga de los archivos de subtitulado en distintos formatos.

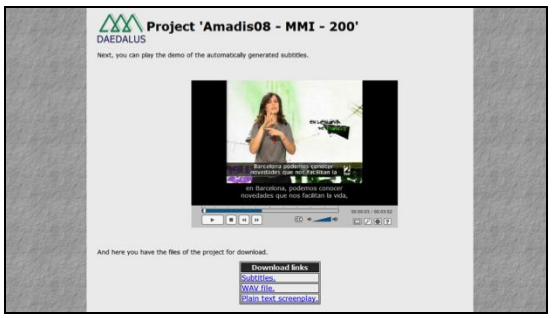


Figura 3: Interfaz de salida con demostrador de los subtítulos generados

5 Conclusiones y lineas futuras

El sistema SAGAS ofrece una herramienta de utilidad a los profesionales involucrados en el proceso de creación de subtítulos siguiendo normativa. Sobre el prototipo actual se ha hecho una evaluación preliminar, detectando algunos aspectos a mejorar en los que se trabaja en la actualidad para incrementar los niveles de calidad del subtitulado generado.

Como líneas futuras, además de mejorar el sistema actual, se quiere integrar una herramienta de autor para que el profesional o subtitulador pueda editar el subtitulado en tiempo de proceso de generación.

Bibliografía

AENOR, 2003. UNE 153.010 Subtitulado para personas sordas y personas con discapacidad auditiva. Subtitulado a través del teletexto.

Álvarez, A., del Pozo, A. and Arruti, A. 2010 APyCA: Towards the automatic subtitling of television content in Spanish. Computer Science and Information Technology (IMCSIT), 567- 574

BOE, 2010. Ley General Audiovisual. Ley 7/2010, de 31 de Marzo, General de la Comunicación Audiovisual se regula la comunicación audiovisual de cobertura estatal y establece las normas básicas en materia audiovisual sin perjuicio de las competencias reservadas a las Comunidades Autónomas y a los Entes Locales en sus respectivos ámbitos.

Bordel, G., Nieto, S., Penagarikano, M., Rodriguez-Fuentes, LJ., Varona, A. 2011. Automatic Subtitling of the Basque Parliament Plenary Sessions Videos. INTERSPEECH 2011. Florence, Italy.

Boulianne, G., Beaumont, F.F., Boisvert, M., Brousseau, J., Cardinal, P., Chapdelaine, C., Comeau, M., Ouellet, P., Osterrath, F. 2006. Computer-assisted closedcaptioning of live TV broadcasts in French, Interspeech 2006, Pittsburgh, USA.

Garcia, JE, Ortega, A, Lleida, E, Lozano, T, Bernues, E and Sanchez, D. 2009. Audio and Text Synchronization for TV news Subtitling based on Automatic Speech Recognition. Broadband Multimedia Systems and Broadcasting, 2009. BMSB '09.

Levenshtein, V. I. 1966. Binary codes capable of correcting deletions, insertions, and reversals. Soviet Physics Doklady 10 (1966):707710.

Meinedo, H., Viveiros, M. and Neto, J. 2008. Evaluation of a live broadcast news subtitling system for Portuguese. Interspeech 2008, Brisbane, Australia, Sep. 2008.

Neto, J.; Meinedo, H.; Viveiros, M.; Cassaca, R.; Martins, C.; Caseiro, D. 2008. Broadcast news subtitling system in Portuguese. 2008. Acoustics, Speech and Signal Processing, 2008. ICASSP 2008. IEEE International Conference on, vol., no., pp.1561-1564