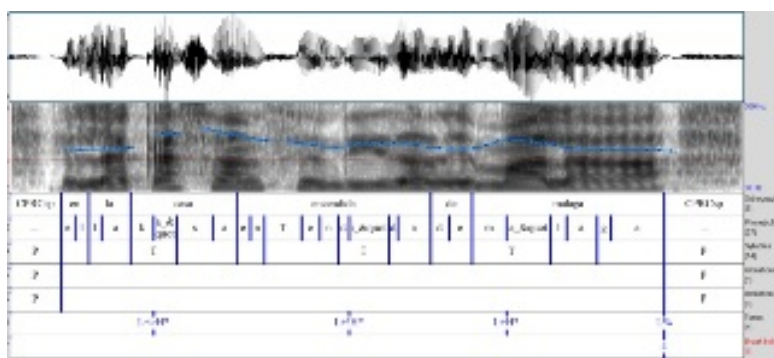


05/2014

Glissando: un corpus para estudios prosódicos multidisciplinares en español y catalán



La bibliografía actual sobre prosodia revela la ausencia de corpus de habla extensos en lengua catalana y castellana que posibiliten el estudio empírico de la entonación de estos idiomas. El corpus Glissando pretende precisamente llenar ese vacío y poner al alcance de todos los públicos grabaciones de voz que permitan caracterizar la entonación de cada lengua, establecer diferencias según distintos criterios (género, estilo, registro, etc.) y ayudar en tareas docentes relacionadas con la lingüística, la expresión oral y la comunicación.

El corpus Glissando se ha desarrollado en el marco del proyecto “Glissando, un corpus de habla anotado para estudios prosódicos en catalán y español: aplicaciones en tecnologías del habla” (FFI2008-04982-C03-02/FILO), proyecto de carácter coordinado entre las universidades de Valladolid (grupo ECA-SIMM, IP: D. Escudero), la Universitat Pompeu Fabra (Departament de Lingüística i Comunicació, IP: J.M. Garrido) y la Universitat Autònoma de Barcelona (Departamento de Filología Española, IP: Lourdes Aguilar). Este corpus comprende distintos estilos de habla: un subcorpus de noticias y un subcorpus de diálogos que, a su vez, se divide en diálogos dirigidos a la consecución de una tarea específica (obtener información sobre transportes, viajes o trámites de gestión universitaria) y conversaciones espontáneas.

Las grabaciones fueron realizadas en alta calidad por dos perfiles de hablantes: profesionales

de la radio y de la publicidad y estudiantes pregraduados nativos. Así, las veinticinco horas de grabaciones abarcan distintos estilos de lectura (radio, publicidad y neutro), registros (lectura de noticias y diálogos formales e informales), voces (masculinas y femeninas) y lenguas (catalán central y español europeo estándar). La inclusión de estas variables tiene por objeto facilitar estudios comparativos de diverso alcance.

La imagen es un ejemplo de visualización de lo que se obtiene fácilmente con un programa de tratamiento de habla de distribución libre (i.e. *praat*). Las grabaciones han sido transcritas ortográficamente y fonéticamente para que puedan reconocerse las palabras en relación con la señal acústica, alineándose con el principio y el final de cada palabra. También se dispone de la separación silábica, la situación de los acentos léxicos (identificados por “_"”) y las prominencias tonales (simbolizado con *T*). Los silencios entre fragmentos se representan entre líneas y con la letra *P* (pausa). De esta forma, el usuario puede practicar la transcripción fonética de los sonidos, aislando aquellos que le presenten mayor dificultad de identificación; puede comparar niveles de acentuación en las sílabas (por ejemplo, acentos primarios en contraposición a acentos secundarios) o puede observar cómo la duración de las pausas incide en una mejor o peor comprensión del mensaje.

Para representar la entonación, se ha aplicado el modelo de la fonología entonativa (i.e. *ToBI*), desarrollado en trabajos anteriores para el castellano y el catalán (*Sp_ToBI* http://prosodia.upf.edu/sp_tobi/en/ y *Cat_ToBI* http://prosodia.upf.edu/cat_tobi/en/). Este modelo permite describir los movimientos melódicos a partir de la diferencia entre tonos bajos (i.e. *L*) y altos (i.e. *H*), así como distinguir entre segmentos entonativos que conllevan un significado completo (i.e. fronteras caracterizadas con el número 4) y segmentos intermedios (i.e. fronteras caracterizadas con el número 3). Gracias a este sistema de representación tonal, se caracteriza desde un punto de vista fonológico la entonación de cada texto oral. Los archivos de audio con sus etiquetas asociadas son de libre acceso para fines de investigación, en la página web del proyecto: <http://veus.glicom.upf.edu>.

La aplicación más directa del corpus Glissando se relaciona con la docencia en el sector universitario pero también es común que en la educación secundaria y el bachillerato se den créditos relacionados con la comunicación, con el objetivo de introducir a los alumnos en terrenos como el periodismo, las nuevas tecnologías, etc. En este ámbito, conocer corpus orales de habla espontánea o dirigida puede ayudar a los alumnos a mejorar sus capacidades comunicativas, ya que, entre otras cosas, les permite formarse como buenos oradores o, simplemente, diferenciar entre una entonación coloquial y otra formal a partir de prácticas orales. Como consecuencia, es más fácil que incorporen en su vida diaria diferentes registros comunicativos.

Por último, además de las investigaciones teóricas y descriptivas que pueden desarrollarse gracias al corpus Glissando, la anotación prosódica de corpus permite crear herramientas útiles para las nuevas Tecnologías de la Información y la Comunicación (TIC). Por ello, sus posibilidades de aplicación cada vez son más extensas. En concreto, una muestra anotada del corpus Glissando ha servido para crear una herramienta semiautomática de transcripción prosódica y se está evaluando su utilidad como material de apoyo para la creación de herramientas de producción y/o comprensión de textos y el aprendizaje del español y del catalán como lenguas extranjeras.

Yurena María Gutiérrez

Lourdes Aguilar

yurenagg@gmail.com, Lourdes.Aguilar@uab.cat

Referencias

Garrido, J. M.; Escudero, D.; Aguilar, L.; Cardeñoso, V.; Rodero, E.; De-La-Mota, C.; González, C.; Rustullet, S.; Larrea, O.; Laplaza, Y.; Vizcaíno, F.; Cabrera, M.; Bonafonte, A. [Glissando: a corpus for multidisciplinary prosodic studies in Spanish and Catalan](#). Language Resources and Evaluation 47(4): 945-971. 2013.

[View low-bandwidth version](#)