"Comparing Online and Offline Methodologies for Spoken Corpus Annotation. The case of Filled Pauses and Hesitation-Related Lengthening in French"

Duvivier, Emilie ; Grosman, Iulia ; Christodoulides, George

ABSTRACT

In this contribution we compare the results of two methodologies for annotating phenomena related to hesitation in speech (e.g. Martin & Strange, 1968), namely filled pauses and lengthening. The first method is the off-line annotation of a spoken corpus by annotators who have access to additional information (e.g. a visual representation of the signal) and the possibility to listen multiple times to segments; most manual annotations in speech corpora are produced using this methodology. The second method is using software that allows for the online annotation of a phenomenon in real-time, without the possibility of stopping the playback or repeatedly listening to segments. In both cases, annotators were “experts” (they had training and experience in phonetics and prosody); we can envisage using the second methodology with “naïve” listeners (participants in an annotation campaign not having prior knowledge of the underlying research question). We also compare these two method...

CITE THIS VERSION

Duvivier, Emilie ; Grosman, Iulia ; Christodoulides, George ; et. al. Comparing Online and Offline Methodologies for Spoken Corpus Annotation. The case of Filled Pauses and Hesitation-Related Lengthening in French. ExLing 2017, 8th Tutorial and Research Workshop on Experimental Linguistics (Heraklion, Crete, Greece, du 19/06/2017 au 22/06/2017). http://hdl.handle.net/2078.1/185770

Le dépôt institutionnel DIAL est destiné au dépôt et à la diffusion de documents scientifiques émanant des membres de l'UCLouvain. Toute utilisation de ce document à des fin lucratives ou commerciales est strictement interdite. L'utilisateur s'engage à respecter les droits d'auteur lié à ce document, principalement le droit à l'intégrité de l'oeuvre et le droit à la paternité. La politique complète de copyright est disponible sur la page Copyright policy.

DIAL is an institutional repository for the deposit and dissemination of scientific documents from UCLouvain members. Usage of this document for profit or commercial purposes is strictly prohibited. User agrees to respect copyright about this document, mainly text integrity and source mention. Full content of copyright policy is available at Copyright policy.
Comparing Online and Offline Methodologies for Spoken Corpus Annotation
The case of Filled Pauses and Hesitation-Related Lengthening in French

Emilie Duvivier¹, Iulia Grosman¹, George Christodoulides¹
¹Centre VALIBEL, Université catholique de Louvain, Belgium

Objective

In this contribution we compare the results of two methodologies for annotating phenomena related to hesitation in speech (e.g. Martin & Strange, 1968), namely filled pauses and lengthening. The first method is the off-line annotation of a spoken corpus by annotators who have access to additional information (e.g. a visual representation of the signal) and the possibility to listen multiple times to segments; most manual annotations in speech corpora are produced using this methodology. The second method is using software that allows for the online annotation of a phenomenon in real-time, without the possibility of stopping the playback or repeatedly listening to segments. In both cases, annotators were “experts” (they had training and experience in phonetics and prosody); we can envisage using the second methodology with “naïve” listeners (participants in an annotation campaign not having prior knowledge of the underlying research question). We also compare these two methodologies with the results of automatic annotation, which is based on statistical models trained on a different speech corpus. These comparisons are a form of triangulation, i.e. “cross-checking data from multiple sources to search for regularities in the research data” (O’Donoghue and Punch, 2003).

Methodology

All annotations were performed on the LOCAS-F corpus (Degand et al., 2014) which is a balanced corpus of spoken French with samples of 14 different speaking styles (discourse genres). The duration of the corpus is 3.5 hours and it contains 43,000 tokens. The corpus is aligned at the phone level (manually verified) and also contains an expert prosodic annotation (two levels of boundaries and four prosodic contours) as well as a part-of-speech and syntactical annotation. The transcription conventions used in the corpus readily identify autonomous filled pauses, and allow us to identify syllables ending in schwa (possibly perceived as epenthetic).

With respect to off-line annotation methods, we used two different sets of data: the entire corpus has been annotated by the second and third authors, independently from one another, and using different annotation schemes. Both annotation protocols, however, include markers for filled pauses and for hesitation-related lengthening at the syllable level.

With respect to the on-line annotation method, we used the Tapping Annotator software (previously used to study the perception of prosodic boundaries, cf. Simon & Christodoulides, 2016). A simple interface (Figure 1) allows the user to listen to corpus audio samples and press a key as soon as they perceive the phenomenon under study. All keystrokes (press and release times) are recorded for analysis. The first author used this interface to annotate the entire LOCAS-F corpus, in non-contiguous sessions of six corpus samples; she repeated the process twice.

Analysis

The annotation campaigns described above result in four datasets (two offline and two online). In the case of on-line annotation, we used a simple method to correlate a keypress to the syllables that triggered the perception of a hesitation (based in the average reaction time of the annotator to pure tones). An additional dataset was the result of an automatic detection of hesitation-related lengthening using the statistical models of the DisMo annotator (Christodoulides & Avanzi, 2015).
We will present a comparison of the five datasets, along with an analysis of the correlates of syllables that triggered the perception of hesitation. These correlates include syllable duration, intonation contour, presence and duration of a subsequent silent pause, local speech rate, position within token and within syntactic constituent.

Figure 1 The Tapping Annotator interface. The user selects a corpus sample (left) and launches playback. A progress bar (bottom, in green) indicates remaining playback time and each keystroke recorded is also displayed as a vertical blue line (middle). The results are saved in an XML file for further analysis.

References


