Automatic diagnosis and feedback for lexical stress errors in non-native speech: Towards a CAPT system for French learners of German

May 14, 2015

Abstract

This is some text.

1 Introduction

2 Related work

3 Lexical stress errors by French learners of German

In an effort to shed light on the nature of lexical stress errors in the speech of L1 French learners of German as L2, **[TODO this chapter has described]** original efforts to annotate and analyze such errors in a small corpus of learner speech.

[TODO As described in ??????,] lexical stress realizations in utterances of bisyllabic, initial-stress words (**??**) taken from the IFCASL corpus (Fauth et al., 2014; Trouvain et al., 2013; **[TODO see also ??)]** were evaluated by multiple annotators from different L1 and phonetic training backgrounds **[TODO (??)]**. Annotators were asked to use a graphical annotation tool to label each recorded word utterance as correctly or incorrectly realizing lexical stress (i.e. the speaker clearly stressed the correct or incorrect syllable), failing to clearly realize stress (i.e. the speaker did not seem to stress either syllable), or having technical or other problems which prevented the assessment of lexical stress **[TODO (??)]**.

Analysis of the labels assigned by different annotators to the same utterances (**??**) revealed that inter-annotator agreement was **[TODO NUMBERS only "fair" (Landis and Koch, 1977),]** on average, between each pair of annotators who labeled the same utterances. Considerable

variation was observed among individual annotators [**TODO (??)**], which did not seem to be explained by differences in their L1 [**TODO (??)**] or level of phonetics/phonology expertise [**TODO (??)**]. However, it was observed that L2 German speakers annotated a higher proportion of utterances as having unclear stress compared to L1 speakers [**TODO (??)**], and that expert annotators judged substantially higher proportion of utterances as correctly realizing stress compared to intermediate or novice annotators [**TODO (??)**]. [**TODO As described in ??,**] there also seemed to be variability in inter-annotator agreement with respect to the different word types in the dataset, and further work is needed to discern the factors responsible for this observation (see **??**).

The multiple, often conflicting, error annotations from different annotators were consolidated into a single gold-standard annotation for each utterance in the dataset [**TODO (see ??)**], which served as the basis for an analysis of the frequency and type of errors produced by learners [**TODO (??)**]. On average, approximately two-thirds of learners' utterances were deemed to realize lexical stress correctly, confirming the expectation that French learners of German frequently make errors with respect to lexical stress [**TODO (see ????)**]. The observed frequency of such errors was considerably lower in the speech of advanced learners than that of beginners [**TODO (??)**], and children seemed to make more errors than adult beginners [**TODO (??)**]; no substantial difference was observed between speakers of different genders. As in the case of inter-annotator agreement, considerable variation was observed in the frequency of errors in utterances of different word types [**TODO (??)**], though once again the factors underlying this variability are not immediately evident and should be investigated in future work [**TODO (see ??)**].

The error annotation and analysis described in this chapter thus contribute considerably to our understanding of the difficulties L1 French speakers may have realizing lexical stress in German, and the fact that learners, especially beginners and children, seem to struggle with lexical stress production justifies the selection of lexical stress errors as the focus of this thesis project. Additionally, the analysis of inter-annotator agreement presented in this chapter, specifically the finding that the observed agreement was generally rather low, constitutes an important discovery with respect to the task of identifying such errors in learner speech: though further research is needed to determine why and under which conditions this is the case (see **??**), it would seem that diagnosing lexical stress errors may be a challenging task for at least some L1 and L2 German speakers. If true, this has important implications for the development and evaluation of automatic error diagnosis systems, which [**TODO will be discussed further in ??.**]

4 Diagnosis of lexical stress errors

5 Feedback on lexical stress errors

References

Fauth, Camille, Anne Bonneau, Frank Zimmerer, Jürgen Trouvain, Bistra Andreeva, Vincent Colotte, Dominique Fohr, Denis Jouvet, Jeanin Jügler, Yves Laprie, Odile Mella, and Bernd Möbius (2014). "Designing a Bilingual Speech Corpus for French and German Language Learners: a Two-Step Process". In: *9th Language Resources and Evaluation Conference (LREC)*. Reykjavik, Iceland, pp. 1477–1482 (cit. on p. 1).

Landis, J. Richard and Gary G. Koch (1977). "The measurement of observer agreement for categorical data." In: *Biometrics* 33.1, pp. 159–174 (cit. on p. 1).

Trouvain, Jürgen, Yves Laprie, Bernd Möbius, Bistra Andreeva, Anne Bonneau, Vincent Colotte, Camille Fauth, Dominique Fohr, Denis Jouvet, Odile Mella, Jeanin Jügler, and Frank Zimmerer (2013). "Designing a bilingual speech corpus for French and German language learners". In: *Corpus et Outils en Linguistique, Langues et Parole: Statuts, Usages et Mésuages*. ii. Strasbourg, France, pp. 32–34 (cit. on p. 1).