Automatic classification of lexical stress errors for German CAPT

Anjana Sofia Vakil

Department of Computational Linguistics & Phonetics
Saarland University, Saarbrücken, Germany
anjanav@coli.uni-saarland.de

Abstract

(200 word limit)

Index Terms: computer-assisted pronunciation training, CAPT, word prosody, German

1. Introduction

For adult learners of a second language (L2), the phonological system of the L2 can pose a variety of difficulties. For certain L2s, such as German or English, one important difficulty involves the accurate prosodic realization of lexical stress, i.e. the accentuation of certain syllable(s) in a given word, with the placement of stress within a word varying freely and carrying a contrastive function in such languages **[TODO cite]**. Lexical stress is an important part of German word prosody, and has been found to have an impact on the intelligibility of non-native German speech **[TODO cite]**. Coping with this phenomenon in German is especially challenging for native (L1) French speakers, because lexical stress is realized very differently (or perhaps not at all) in the French language **[TODO cite]**.

To overcome this difficulty and improve their L2 word prosody, learners typically need to have their pronunciation errors pointed out and corrected by a language instructor; unfortunately, the lack of attention typically given to pronunciation in the foreign language classroom, along with other factors such as high student-to-teacher ratios, make this level of individualized attention not always feasible in a classroom setting [1, 2, 3]. Fortunately, advances in Computer-Assisted Pronunciation Training (CAPT) over recent decades have made it possible to automatically provide highly individualized analysis of learners' prosodic errors, as well as feedback on how to correct them, and thus to help learners achieve more intelligible pronunciation in the target language. However, while much research has gone into the creation and improvement of CAPT systems for English (see e.g. [4, 5]), relatively little work has been done on the development of CAPT systems for German, especially on those targeting errors in German prosody.

This paper describes work that advances the state of German CAPT by applying machine learning methods to the task of diagnosing lexical stress errors in non-native German speech, a necessary prerequisite for delivering individualized corrective feedback on such errors in a CAPT system. The paper is organized as follows: Section 2 provides background on the phenomenon of lexical stress as it is realized in German and French word prosody, motivates the creation of CAPT systems that address this error specifically, and summarizes some past work related to this topic. Section 3 describes the manual annotation of lexical stress errors in a small corpus of L2 German speech, carried out to create labeled training and test data for the classification experiments explained in section 4. Section 5 presents and analyzes the results of these experiments. Finally, section 6 offers some concluding remarks and outlines possible directions for future work.

2. Background and related work

Broadly speaking, lexical stress is the phenomenon of how a given syllable is accentuated within a word [6], i.e. how a syllable is given a more prominent role such that this syllable is perceived as "standing out" [7]. This perceived prominence of a syllable is a function not merely of the segmental characteristics of the uttered syllable, i.e. the speech sounds it contains, but rather of its (relative) suprasegmental properties, namely:

- duration, which equates on the perceptual level to length;
- fundamental frequency (F0), which corresponds to perceived pitch; and
- intensity (energy or amplitude), which perceptually equates to loudness.

In variable-stress languages, such as German and English, the location of lexical stress in a word is not always predictable, and therefore knowing a word requires, in part, knowing its stress pattern. This allows lexical stress to serve a contrastive function in these languages, e.g. distinguishing *UMfahren* (to drive around) from *umFAHRen* (to run over with a car) in German. Furthermore, in German, misplaced stress can disrupt understanding even in cases where there is no stress-based minimal pair [8].

However, in fixed-stress languages, stress is completely predictable, as it always falls on a certain position in the word (e.g. the final syllable), making the lexical stress pattern less crucial to the knowledge of a word than in variable-stress languages. Furthermore, in fixed-stress languages there may be a weaker distinction between stressed and unstressed syllables. While French has often been categorized as a fixed-stress language, given that word-final syllables are given prominence when a French word is pronounced in isolation, some argue that it may be more properly considered a language without lexical stress, in that speakers do not seem to accentuate any syllable within the word, with word-final lengthening effects explained by interactions with the realization of phrasal accent (lengthening of the final syllable in each prosodic group or phrase) [9, 10]. Regardless, French has no contrastive word-level stress [10, p. 89], and in this respect differs considerably from German.

This difference between the languages leads us to expect French learners of German to have difficulties with both perception and production of lexical stress prosody, although the literature contains little, if any, empirical research on the nature of lexical stress errors for this particular L1-L2 pair. However, work on French speakers' perception of Spanish, another contrastive-stress language, has revealed that these speakers

seem to be "deaf" to lexical stress, i.e. seem to have significant and lasting difficulty perceiving and remembering stress contrasts [9]. With respect to production, studies of L2 Dutch have shown that French speakers, especially beginners, make systematic errors with lexical stress, exhibiting a tendency to stress the final syllable of Dutch words even when stress should be placed on the initial or medial syllable [11, 10]. Similar findings have also been reported for French learners of English [12].

[TODO transition]

3. Data

3.1. The IFCASL corpus

3.2. Annotation of lexical stress realizations

3.3. Inter-annotator agreement

3.4. Error distribution

4. Evaluation method

4.1. Feature sets

4.2. Datasets for training and testing

5. Results

5.1. Feature performance

5.2. Performance on unknown words

5.3. Performance on unknown speakers

6. Conclusions and future work

7. References

[1] A. Neri, C. Cucchiarini, H. Strik, and L. Boves, "The pedagogy-technology interface in computer assisted pronunciation training," *Computer Assisted Language Learning*, 2002. [Online]. Available: http://www.tandfonline.com/doi/abs/10.1076/call.15.5.441.13473

[2] T. M. Derwing and M. J. Munro, "Second Language Accent and Pronunciation Teaching: A Research-Based Approach," *TESOL Quarterly*, vol. 39, no. 3, pp. pp. 379–397, 2005. [Online]. Available: http://www.jstor.org/stable/3588486 http://onlinelibrary.wiley.com/doi/10.2307/3588486/abstract

[3] U. Hirschfeld and J. Trouvain, "Teaching prosody in German as foreign language," in *Non-Native Prosody: Phonetic Description and Teaching Practice*, J. Trouvain and U. Gut, Eds. Walter de Gruyter, 2007, pp. 171–187. [Online]. Available: http://books.google.com/books?hl=en&lr=&id=cZ2St1QhV1cC&oi=fnd&pg=PA1
b&sig=TpN9xHbMtwL-r5kBRi5dpQb_28w

[4] M. Eskenazi, "An overview of spoken language technology for education," *Speech Communication*, vol. 51, no. 10, pp. 832–844, Oct. 2009. [Online]. Available: http://linkinghub.elsevier.com/retrieve/pii/S0167639309000673

[5] S. M. Witt, "Automatic error detection in pronunciation training: Where we are and where we need to go," in *Proceedings of the International Symposium on Automatic Detection of Errors in Pronunciation Training (IS ADEPT)*, 2012, pp. 1–8. [Online]. Available: http://www.researchgate.net/publication/250306074_Automatic_Error_Detection_in

[6] A. Cutler, "Lexical Stress," in *The Handbook of Speech Perception*, D. B. Pisoni and R. E. Remez, Eds., 2005, pp. 264–289. [Online]. Available: http://books.google.com/books?hl=en&lr=&id=EwY15naRiFgC&oi=fnd&pg=PA2

[7] G. Dogil and B. Williams, "The phonetic manifestation of word stress," in *Word Prosodic Systems in the Languages of Europe*, H. van der Hulst, Ed. Berlin: Walter de Gruyter, 1999, ch. 5, pp. 273–334. [Online]. Available: http://books.google.com/books?hl=en&lr=&id=gDiqDQqO2hwC&oi=fnd&pg=PA
https://books.google.com/books?hl=en&lr=&id=gDiqDQqO2hwC&oi=fnd&pg=P

[8] U. Hirschfeld, *Untersuchungen zur phonetischen Verständlichkeit Deutschlernender*, ser. Forum Phoneticum, 1994, vol. 57.

[9] E. Dupoux, N. Sebastián-Gallés, E. Navarette, and S. Peperkamp, "Persistent stress 'deafness': The case of French learners of Spanish," *Cognition*, vol. 106, pp. 682–706, 2008. [Online]. Available: http://www.sciencedirect.com/science/article/pii/S0010027707001102

[10] M.-C. Michaux and J. Caspers, "The production of Dutch word stress by Francophone learners," in *Proceedings of the Prosody-Discourse Interface Conference 2013 (IDP-2013)*, 2013, pp. 89–94. [Online]. Available: https://lirias.kuleuven.be/bitstream/123456789/417450/1/Mertens_Simon_2013_Pro

[11] M.-C. Michaux, "Exploring the production and perception of word stress by French-speaking learners of Dutch," in *Workshop on Crosslinguistic Influence in Non-Native Language Acquisition*, 2012. [Online]. Available: http://www.fcsh.unl.pt/clunl/pl2/wcil2/pdfs/Marie-Catherine Michaux.pdf

[12] A. Bonneau and V. Colotte, "Automatic Feedback for L2 Prosody Learning," in *Speech and Language Technologies*, I. Ipsic, Ed. InTech, 2011, no. 1977. [Online]. Available: http://www.intechopen.com/books/speech-and-language-technologies/automatic-feedback-for-l2-prosody-learning http://hal.inria.fr/inria-00579255/