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

Anjana Sofia Vakil

UNIVERSITÄT
DES
SAARLANDES

Department of Computational Linguistics and Phonetics
University of Saarland, Saarbrücken, Germany

Master's Thesis Colloquium
16 April 2015

Lexical stress

Some syllable(s) in a word more accentuated/prominent[1]

um·FAHR·en	vs.	UM·fahr·en
to run over		*to drive around*

- German: variable stress placement, contrastive stress[1]
- French: no word-level stress, final syllable lengthening[2]

Goal: Computer-Assisted Pronunciation Training (CAPT) for lexical stress errors for French learners of German

---

[1] A. Cutler. "Lexical Stress". In: *The Handbook of Speech Perception*. Ed. by D. B. Pisoni and R. E. Remez. 2005, pp. 264–289.

[2] M.-C. Michaux and J. Caspers. "The production of Dutch word stress by Francophone learners". In: *Proc. of the Prosody-Discourse Interface Conference (IDP)*. 2013, pp. 89–94.

Outline

Motivation

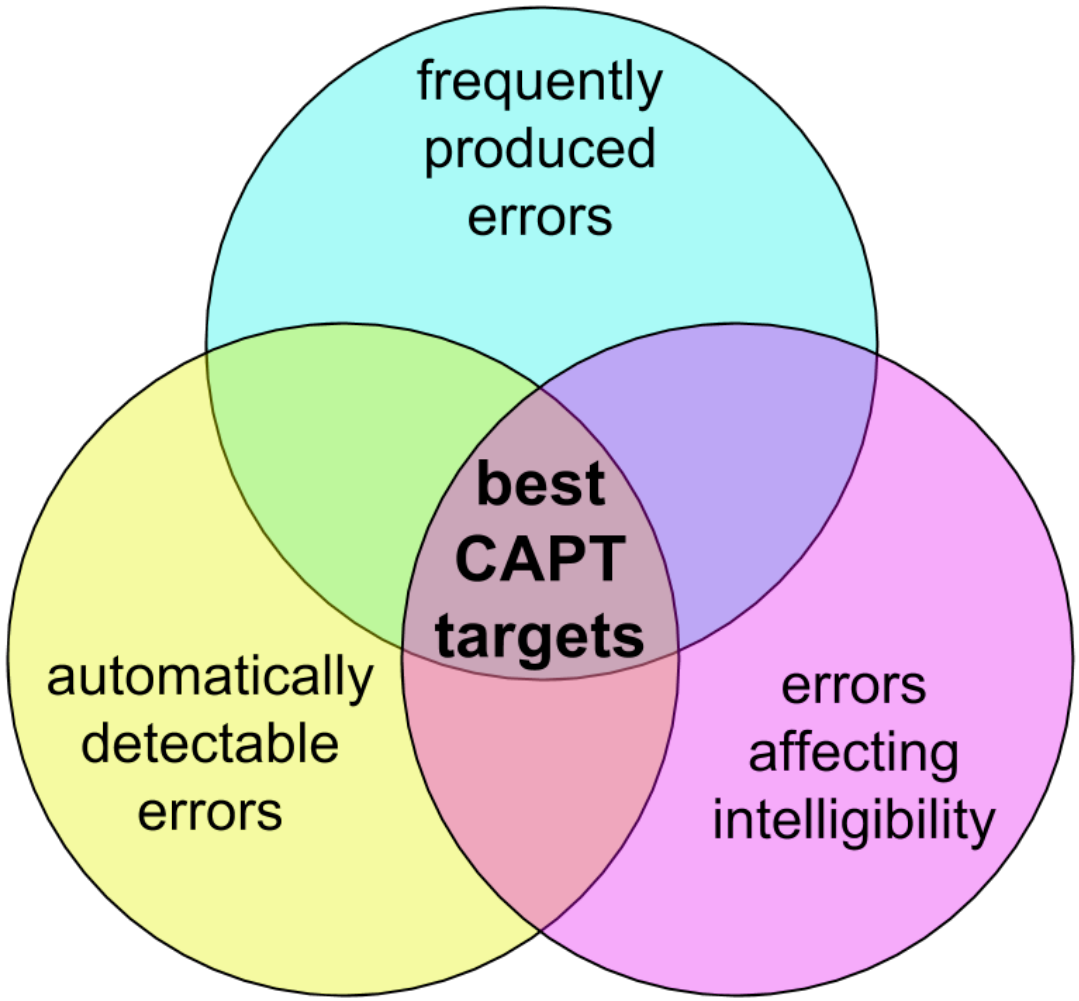
Lexical stress errors by French learners of German
- Annotation of a learner speech corpus
- Inter-annotator agreement
- Errors

Error diagnosis
- Word prosody analysis
- Diagnosis by comparison
- Diagnosis by classification

Feedback
- Implicit
- Explicit
- Self-assessment

The de-stress CAPT tool

Figure: Criteria for selecting errors to target in a CAPT system.



Motivation

Lexical stress errors seem to be:

- Frequently produced by French learners of variable-stress languages[1,2]
- More important for intelligibility in L2 German than other types of errors[3]
- Possible to identify automatically by comparison[1] or classification[4]

---

[1] A. Bonneau and V. Colotte. "Automatic Feedback for L2 Prosody Learning". In: *Speech and Language Technologies*. Ed. by I. Ipsic. InTech, 2011.

[2] 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.

[3] U. Hirschfeld. *Untersuchungen zur phonetischen Verständlichkeit Deutschlernender*. Vol. 57. Forum Phoneticum. 1994.

[4] Y.-J. Kim and M. C. Beutnagel. "Automatic assessment of American English lexical stress using machine learning algorithms". In: *SLaTE*. 2011, pp. 93–96.

Lexical stress errors in learner speech

UNIVERSITÄT DES SAARLANDES

- How reliably can human annotators identify errors in learner utterances?
- How frequently are errors actually produced by French learners of German?

Error annotation

Data: IFCASL corpus of French-German L1/L2 speech[1]

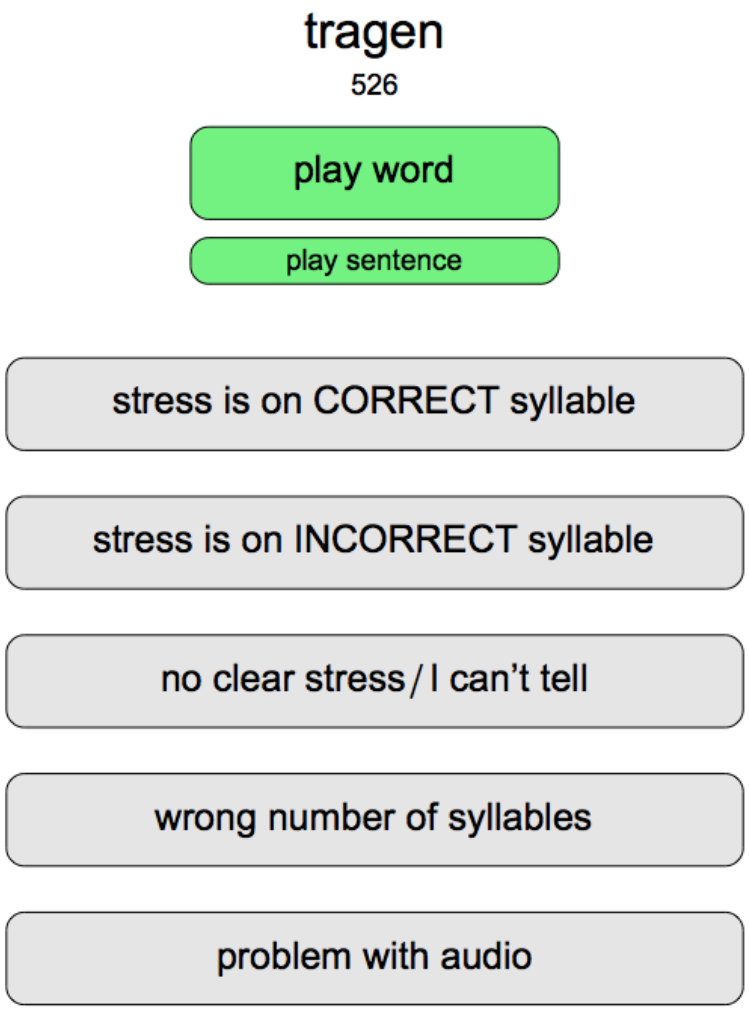
- German utterances by French and German speakers
- Word- and phone-level segmentations (syllable level added automatically)
- Selected 12 word types (bisyllabic, initial stress)
- Dataset: 668 word utterances by 55-56 speakers

Annotators (15 in total):

- L1: 12 German speakers, 2 English, 1 Hebrew
- Expertise: 2 Experts, 10 Intermediates, 3 Novices
- Annotated 3 word types in 1 $\sim$15 min. session

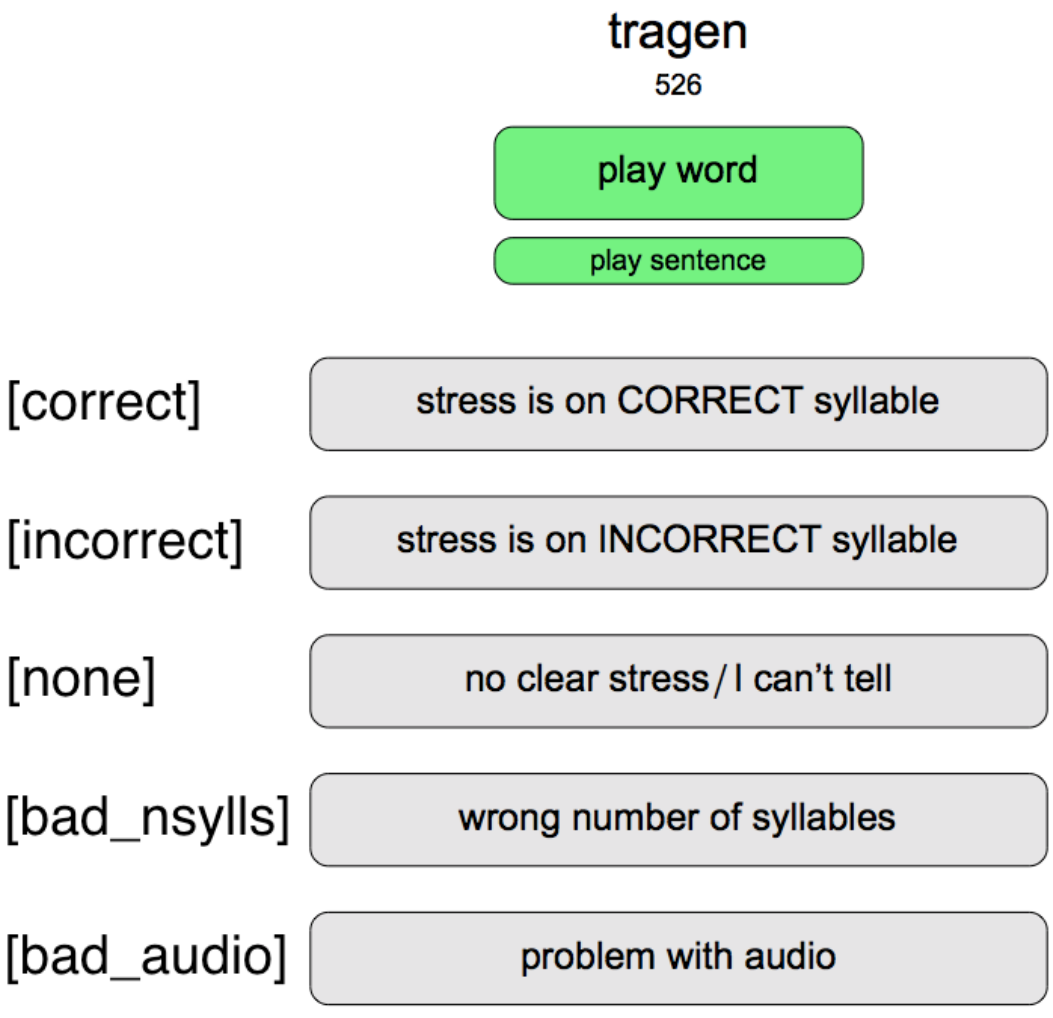
[1] C. Fauth et al. "Designing a Bilingual Speech Corpus for French and German Language Learners: a Two-Step Process". In: *9th Language Resources and Evaluation Conference (LREC)*. Reykjavik, Iceland, 2014, pp. 1477–1482.

Figure: Praat annotation tool



Error annotation

Figure: Praat annotation tool



Inter-annotator agreement

How reliably can human annotators identify errors in learner utterances?

- Agreement calculated for each overlapping pair
- Quantified by:
  - Percentage agreement: N agreed/N both annotated
  - Cohen's Kappa[1] (κ): accounts for chance agreement
- Overall agreement represented by mean, minimum, median, and maximum of all pairwise values

[1] J. Cohen. "A Coefficient of Agreement for Nominal Scales". In: *Educational and Psychological Measurement* 20.1 (Apr. 1960), pp. 37–46.

Inter-annotator agreement

Table: Overall pairwise agreement between annotators

	% Agreement	Cohen's κ
Mean	54.92%	0.23
Maximum	83.93%	0.61
Median	55.36%	0.26
Minimum	23.21%	-0.01

- Generally low agreement ("fair"[1] mean κ)
- Large variability between annotators
- Not explained by L1/expertise groups

[1] J. R. Landis and G. G. Koch. "The measurement of observer agreement for categorical data." In: *Biometrics* 33.1 (1977), pp. 159–174.

[TODO]

UNIVERSITÄT DES SAARLANDES

[TODO]

UNIVERSITÄT DES SAARLANDES

[TODO]

UNIVERSITÄT DES SAARLANDES

[TODO]

[TODO]

[TODO]

[TODO]

UNIVERSITÄT DES SAARLANDES

[TODO]