# Individualized feedback for lexical stress errors

## Towards a CAPT system for French learners of German

---

Anjana Sofia Vakil

M.Sc. Thesis proposal

Language Science and Technology

*Supervised by*

Prof. Dr. Bernd Möbius

Dr. Jürgen Trouvain

UNIVERSITÄT
DES
SAARLANDES

Saarland University

Department of Computational Linguistics & Phonetics

10 Never, 2014

# Contents

# 1 Introduction

In the field of second-language education, pronunciation has traditionally been given less attention than other areas such as grammar or vocabulary (Derwing and Munro, 2005). One reason for this may be that pronunciation is best taught through one-on-one instruction, which is not often possible in the traditional classroom setting. Hence the attraction of Computer-Assisted Pronunciation Training (CAPT) systems, which have the potential to automatically provide highly individualized analysis of learner errors, and feedback on how to correct them and achieve more intelligible and native-like pronunciation in the target language (Witt, 2012).

For students with French as their first language (L1) who are learning German as a second language (L2), the sound system of the L2 can pose a variety of difficulties, one of the most important and interesting of which is the way in which certain syllables in German words are accentuated more than others, a phenomenon referred to as lexical stress. Learning to navigate German lexical stress is especially challenging for L1 French speakers, because this phenomenon is realized differently, or perhaps does not occur at all, in the French language.

With these motivations in mind, this thesis project aims to advance the state of the art of German CAPT by creating a tool which will diagnose and offer feedback on lexical stress errors in the L2 German speech of L1 French speakers, in the hopes of helping these learners become more sensitive to the lexical stress patterns of German and develop the ability to accurately realize these patterns in their speech.

## 1.1 Context: The IFCASL project

The work reported here has been conducted in the context of the ongoing research project “Individualized Feedback in Computer-Assisted Spoken Language Learning (IFCASL)” at the University of Saarland (Saarbrücken, Germany) and LORIA (Nancy, France).

The ultimate goal of the project is to take initial steps toward the development of a CAPT system targeting, on the one hand, native (L1) French speakers learning German as a foreign language (L2), and on the other, L1 German speakers learning French as their L2. To this end, a bidirectional learner speech corpus has been recorded, comprising phonetically diverse utterances in French and German spoken

by both native speakers and non-native speakers with the other language as L1 (Fauth et al., 2014; Trouvain et al., 2013).

This thesis will focus exclusively on French L1 speakers learning German as L2. The German-language subset of the IFCASL corpus will be instrumental in training and testing the automatic diagnosis and feedback systems which this work aims to develop. Futhermore, those systems will be designed with a view to contributing to the overall set of software developed in the context of the IFCASL project, such that they will be as compatible as possible with the other tools developed and used by the IFCASL team.

## 1.2 Objectives

The main objective of this work is to investigate the automatic treatment of lexical stress errors in the context of a CAPT system for French learners of German. This includes, on the one hand, an examination of the ways in which lexical stress errors of the type made by French L1 speakers when speaking German as L2 can be reliably detected and measured automatically, and on the other, an exploration of the types of multimodal feedback on such errors that can be automatically delivered based on the aforementioned error detection.

The intented outcome of these investigations is a prototype CAPT tool, illustrated in 1.1, which can diagnose lexical stress errors in different ways and present learners with different types of feedback on these errors, such that researchers can use this modular system to study the impact of various assessment and feedback types on learner outcomes, user engagement, and other factors impacting the success of a CAPT system.

Once more is known about which diagnosis/feedback types should be delivered to which learners in which situations, this tool could become a useful component to a fully-fledged CAPT system, in which learner models and other intelligent components automatically decide which modules of the tool to activate.

## 1.3 Thesis overview

**Chapter 2** introduces Computer-Assisted Pronunciation Training in the context of pronunciation teaching in foreign-language education and computer-based and intelligent tutoring systems, and outlines the differences in the lexical stress systems of French and German as well as the motivation for focusing on lexical stress errors in this work.

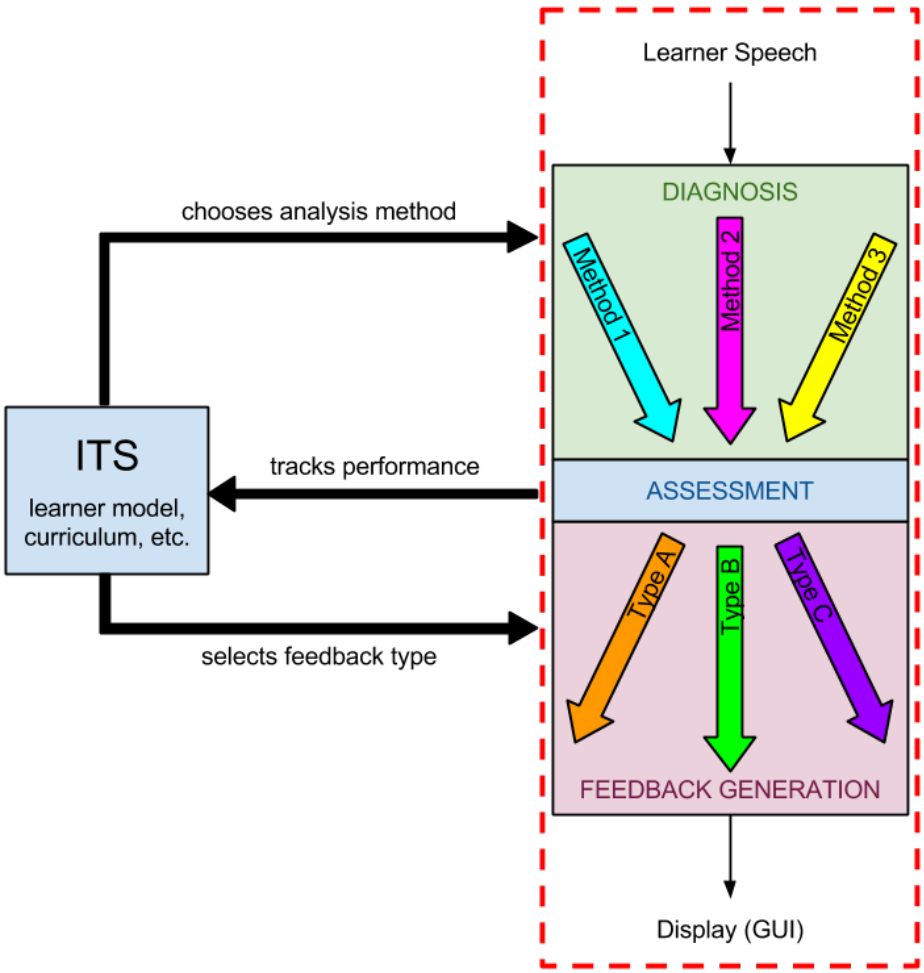


**Figure 1.1:** Conceptual diagram of the prototype lexical stress CAPT tool (demarcated by dotted line) and its possible function in the context of a more comprehensive Intelligent Tutoring System.

**Chapter 3** will briefly introduce the tools that have been developed, and the technology used to build them.

**Chapter 4** will detail the system for assessing learner speech in terms of lexical stress errors. It will describe the methods used to automatically segment the learner's utterance, analyze the prosody of this utterance in terms of the relative pitch, duration, and intensity of the relevant syllables, and compare this analysis to one or more models of native pronunciation to produce a diagnosis.

**Chapter 5** will describe the multimodal feedback options that the system can deliver, and how these feedback types are generated based on the analysis of the learner's speech described in the previous chapter.

**Chapter 6** will summarize the contributions of this work and outlines some interesting future directions to build on these contributions.

# Background and related work

# 2

## 2.1 Computer-Assisted Pronunciation Training

In the field of second-language education, pronunciation has traditionally been given less attention than other areas such as grammar or vocabulary (Derwing and Munro, 2005). One reason for this may be that pronunciation is best taught through one-on-one instruction, which is not often possible in the traditional classroom setting. Hence the attraction of Computer-Assisted Pronunciation Training (CAPT) systems, which have the potential to automatically provide highly individualized analysis of learner errors, and feedback on how to correct them and achieve more intelligible and native-like pronunciation in the target language (Witt, 2012).

### 2.1.1 Pronunciation in foreign language education

### 2.1.2 Computer-based and intelligent tutoring systems

### 2.1.3 Survey of existing CAPT systems?

## 2.2 Lexical stress

Lexical stress is the phenomenon of how syllables are accentuated within a word (Cutler, 2005). This relates not to the segmental characteristics of a syllable, i.e. the speech sounds it contains, but rather to its (relative) suprasegmental properties, namely:

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

### 2.2.1 German vs. French

As Cutler points out, different languages make use of this suprasegmental information in different ways. In what are termed free- or variable-stress languages, such

as German, Spanish, and English, it is not always possible to predict which syllable in a word will carry the stress, and therefore knowing a word requires, in part, knowing its stress pattern. This allows stress to serve a contrastive function in these languages, such that two words may share exactly the same sequence of phones and nevertheless be distinguished exclusively by their stress pattern, as is the case with X and Y in German. Because stress carries meaning thus, native speakers of such languages are sensitive to stress patterns, and readily able to perceive differences in stress.

However, in the so-called fixed-stress languages, stress is completely predictable, as it always falls on a certain position in the word; in Czech and Hungarian, for example, stress always falls on the word-initial syllable. Therefore, lexical stress may not be as crucial to the knowledge of a word in these languages as in the free-stress languages. Furthermore, although lexical stress is realized in these languages, the distinction between stressed and unstressed syllables may be weaker than in free-stress languages. French has often been placed into this category of fixed-stress languages, although it may be more properly considered a language without lexical stress, insofar as there is no systematic way in which speakers distinguish a certain syllable from others in the word, aside from the fact that French exhibits phrasal accent, i.e. lengthening of the final syllable in each prosodic group or phrase (Dupoux et al., 2008).

Therefore, native speakers of French may lack the sensitivity to stress patterns possessed by native speakers of German. Indeed, this has been borne out by research by Dupoux et al. (Peperkamp and Dupoux, 2002; Dupoux, 2001; Dupoux et al., 2008), which demonstrated that native French speakers were "deaf" to differences in stress patterns, such that they have great difficulty discriminating between Spanish words which contrast only at the level of stress. This difficulty should therefore also exist for French speakers when they are presented with German words in which the stress pattern is crucial to the word's meaning, as in the minimal pair above.

### 2.2.2 Lexical stress in foreign language teaching

## 2.3 Targeting lexical stress errors in CAPT

Learners of a foreign language typically make a wide variety of pronunciation errors, at both the segmental level (e.g. errors in producing certain individual phones of the target language) and the prosodic level (e.g. errors in the speaker's intonation contour or the duration of certain syllables or words). As it is not possible to address all of these in an automated system, one of the first aims of this work is to identify

a single type of error which is well suited to being addressed via a CAPT system targeting French L1 learners of German as the L2.

To guide this selection, we may consider a set of three criteria that such an error must meet. First, the error must be produced with a some degree of frequency by French L1 speakers in their production of L2 German, as it would be a misuse of resources to design a system which addresses an error that is seldom made by learners. Secondly, the given error must have a significant impact on the perceived intelligibility of the learner's speech; as the ultimate goal of the system is to help learners communicate more effectively in the L2, an error which is commonly made but nevertheless does not impede understanding of the learner's L2 speech, and thus does not hinder communication in the L2, is not an ideal target. Finally, in order for the CAPT system to provide any meaningful diagnosis of and feedback on the error, it must lend itself to reasonably accurate and reliable detection through automatic processing.


**Figure 2.1:** Criteria for selecting errors to target in a CAPT system.

As illustrated in 2.1, the best error to target with the CAPT system will fulfill all of these criteria, rather than only one or two of the three. For example, vowel quality errors (e.g. an L1 French speaker producing a German /ə/ as [œ]) may occur frequently in the L2 speech and may be relatively easy to detect automatically, but may not have a great impact on the intelligibility of the L2 German speech. On the other hand, equally frequent vowel quantity errors (e.g. the L1 French speaker producing a German long /eː/ as [e]) may have a greater impact on intelligibility in some cases, but may be more difficult to reliably identify automatically.

Analysis of the typical and expected errors described in **??** in terms of these criteria reveals that lexical stress errors are a strong candidate for treatment via CAPT, and will therefore be the focus of the prototype CAPT system described in this thesis. The

remainder of this section justifies the selection of this type of error by describing how it fulfills the aforementioned criteria as well or better than any other error type.

### 2.3.1 Impact on intelligibility

(Warren et al., 2009)

(Magen, 1998)

Stress errors may affect perception of segmental errors; errors in stressed syllables are more noticeable (Cutler, 2005)

### 2.3.2 Frequency of production

(Cutler, 2005)

(Peperkamp and Dupoux, 2002; Dupoux, 2001; Dupoux et al., 2008)

### 2.3.3 Feasibility of automatic detection

(Engwall, 2012; Delmonte, 2011)

(Bonneau and Colotte, 2011)

(Shahin et al., 2012)

## 2.4 Summary

# 3 System overview?

# 4 Diagnosis of lexical stress errors

## 4.1 Related work

Lots of Nancy references

Duong et al., 2011

Probst et al., 2002

## 4.2 Automatic segmentation of nonnative speech

### 4.2.1 Segmentation via forced alignment

### 4.2.2 Evaluation of system accuracy

### 4.2.3 Coping with segmentation errors

Mesbahi et al., 2011, Orosanu et al., 2012

## 4.3 Prosodic analysis

"acoustic differences between stressed and unstressed syllables are relatively large in spontaneous speech. With laboratory-read materials, however, such differences do not always arise" (Cutler, 2005, p. 275)

### 4.3.1 Pitch

"...in German utterances, stressed syllables could be signaled by any F0 obtrusion from the overall contour, so that a stressed syllable could be either higher or lower in pitch than its neighbors" (Cutler, 2005, p. 267)

### 4.3.2 Duration

### 4.3.3 Intensity

least important of the three (Cutler, 2005)

## 4.4 Comparison of native and nonnative speech

Probst et al., 2002

### 4.4.1 Using a single reference speaker

#### Manually selecting a reference

#### Automatically selecting a reference

### 4.4.2 Using multiple reference speakers

### 4.4.3 Using no reference speaker?

Duong et al., 2011

## 4.5 Summary

# 5 Feedback on lexical stress errors

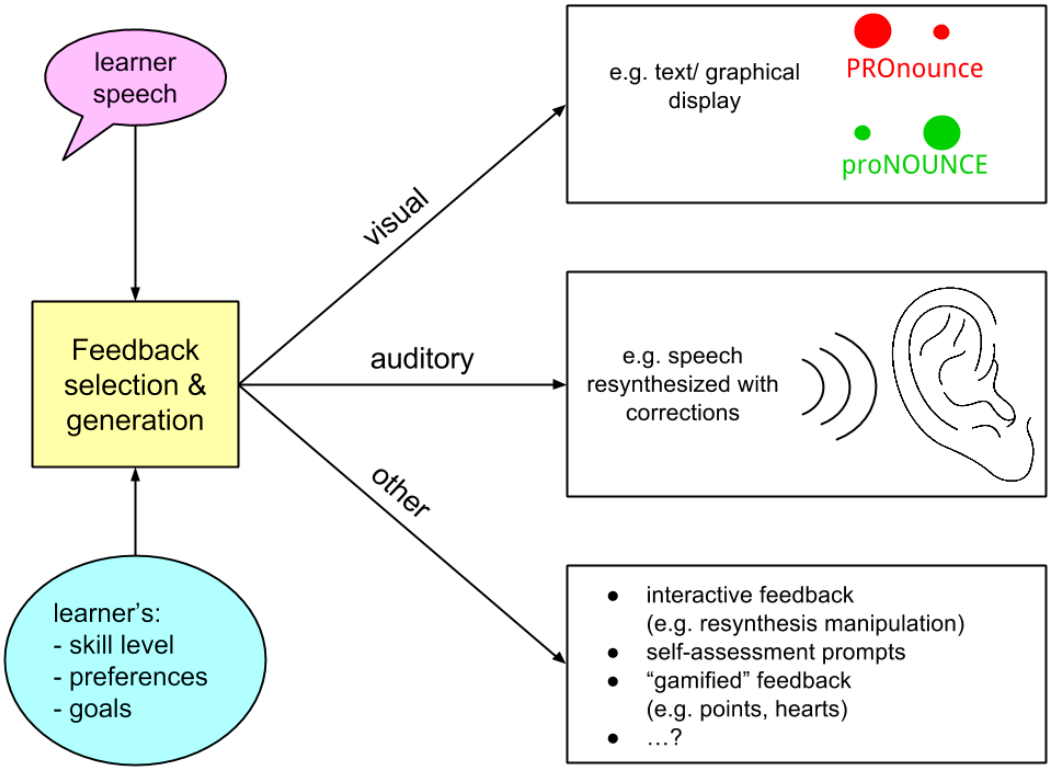


**Figure 5.1:** Delivery of prosody feedback in different modalities.

## 5.1 Related work

Sitaram et al., 2011 Bonneau and Colotte, 2011 (Hattie and Timperley, 2007)

## 5.2 Visual feedback

### 5.2.1 Stylized text

### 5.2.2 Graphical representations of prosody

### 5.2.3 Visualizations of the speech signal

## 5.3 Auditory feedback

### 5.3.1 Enhanced reference utterance

### 5.3.2 Resynthesized learner speech

## 5.4 Alternative feedback types

### 5.4.1 Metalinguistic feedback

### 5.4.2 Interactive feedback

### 5.4.3 Implicit feedback

## 5.5 Summary

# 6 Conclusion and outlook

## 6.1 Thesis summary

## 6.2 Future work

# Bibliography


Bonneau, Anne and Vincent Colotte (2011). “Automatic Feedback for L2 Prosody Learning”. In: *Speech and Language Technologies*. Ed. by Ivo Ipsic. 1977. InTech.

Cutler, Anne (2005). “Lexical Stress”. In: *The Handbook of Speech Perception*. Ed. by David B Pisoni and Robert E Remez, pp. 264–289.

Delmonte, Rodolfo (2011). “Exploring Speech Technologies for Language Learning”. In: *Speech and Language Technologies*. Ed. by Ivo Ipsic. InTech.

Derwing, Tracey M and Murray J Munro (2005). “Second Language Accent and Pronunciation Teaching: A Research-Based Approach”. In: *TESOL Quarterly* 39.3, pp. 379–397.

Duong, Minh, Jack Mostow, and Sunayana Sitaram (2011). “Two methods for assessing oral reading prosody”. In: *ACM Transactions on Speech and Language Processing* 7.212, pp. 1–22.

Dupoux, E (2001). “A robust method to study stress “deafness””. In: *The Journal of the . . .*

Dupoux, Emmanuel, Núria Sebastián-Gallés, Eduardo Navarette, and Sharon Peperkamp (2008). “Persistent stress ’deafness’: The case of French learners of Spanish”. In: *Cognition* 106, pp. 682–706.

“Proceedings of the International Symposium on Automatic Detection of Errors in Pronunciation Training (IS ADEPT)” (2012). In: ed. by Olov Engwall. Stockholm, Sweden: KTH, Computer Science and Communication.

Fauth, Camille, Anne Bonneau, and Frank Zimmerer (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.

Hattie, J and H Timperley (2007). “The power of feedback”. In: *Review of Educational Research* 77.1, pp. 81–112.

Magen, Harriet S (1998). “The perception of foreign-accented speech”. In: *Journal of Phonetics* 26.4, pp. 381–400.

Mesbahi, Larbi, Denis Jouvet, Anne Bonneau, and Dominique Fohr (2011). “Reliability of non-native speech automatic segmentation for prosodic feedback.” In: *SLaTE*.

Orosanu, Luiza, Denis Jouvet, Dominique Fohr, Irina Illina, and Anne Bonneau (2012). “Combining criteria for the detection of incorrect entries of non-native speech in the context of foreign language learning”. In: *SLT 2012 - 4th IEEE Workshop on Spoken Language Technology*.

Peperkamp, S and E Dupoux (2002). “A typological study of stress ’deafness’”. In: *Laboratory phonology*.

Probst, Katharina, Yan Ke, and Maxine Eskenazi (2002). “Enhancing foreign language tutors – In search of the golden speaker”. In: *Speech Communication* 37.3-4, pp. 161–173.

Shahin, Mostafa Ali, Beena Ahmed, and Kirrie J. Ballard (2012). “Automatic classification of unequal lexical stress patterns using machine learning algorithms”. In: *2012 IEEE Spoken Language Technology Workshop (SLT)*. IEEE, pp. 388–391.

Sitaram, S, J Mostow, Y Li, A Weinstein, D Yen, and J Valeri (2011). “What visual feedback should a reading tutor give children on their oral reading prosody?” In: *SLaTE*.

Trouvain, Jürgen, Yves Laprie, and Bernd Möbius (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.

Warren, Paul, Irina Elgort, and David Crabbe (2009). “Comprehensibility and prosody ratings for pronunciation software development”. In: *Language Learning & Technology* 13.3, pp. 87–102.

Witt, Silke M (2012). “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)*, pp. 1–8.