The IFCASL Corpus of French and German Non-native and Native Read Speech

Jürgen Trouvain[1], Anne Bonneau[2], Vincent Colotte[2], Camille Fauth[2,3], Dominique Fohr[2], Denis Jouvet[2], Jeanin Jügler[1], Yves Laprie[2], Odile Mella[2], Bernd Möbius[1], Frank Zimmerer[1]

[1]Computational Linguistics and Phonetics, Saarland University, Germany;
[2]Speech Group, LORIA (Inria; Université de Lorraine, LORIA, UMR 7503; CNRS, LORIA, UMR 7503)
Villers-lès-Nancy, France;
[3]Phonetics, Strasbourg University, France
Corresponding author: trouvain [at] coli.uni-saarland.de

Abstract

The IFCASL corpus is a French-German bilingual phonetic learner corpus designed, recorded and annotated in a project on individualized feedback in computer-assisted spoken language learning. The motivation for setting up this corpus was that there is no phonetically annotated and segmented corpus for this language pair of comparable of size and coverage. In contrast to most learner corpora, the IFCASL corpus incorporate data for a language pair in both directions, i.e. in our case French learners of German, and German learners of French. In addition, the corpus is complemented by two sub-corpora of native speech by the same speakers. The corpus provides spoken data by about 100 speakers with comparable productions, annotated and segmented on the word and the phone level, with more than 50% manually corrected data. The paper reports on inter-annotator agreement and the optimization of the acoustic models for forced speech-text alignment in exercises for computer-assisted pronunciation training. Example studies based on the corpus data with a phonetic focus include topics such as the realization of /h/ and glottal stop, final devoicing of obstruents, vowel quantity and quality, pitch range, and tempo.

Keywords: learner corpus, phonetics, French, German, non-native speech, native speech

1. Introduction

The IFCASL corpus is a French-German bilingual phonetic learner corpus designed, recorded and annotated in the project IFCASL (Individualized Feedback in Computer-Assisted Spoken Language Learning <www.ifcasl.org>) (Trouvain et al. 2013; Fauth et al. 2014).

1.1 Motivation and aims for the corpus

The motivation for setting up this corpus was that there is no phonetically annotated and segmented corpus for this language pair of comparable of size and coverage. Generally speaking, most learner corpora exist for the written language and the majority of spoken learner corpora has English as the target language. In addition, only few learner corpora incorporate data for a language pair in both directions, i.e. in our case French learners of German, and German learners of French.

To our knowledge existing spoken learner corpora for the French-German language pair are restricted to the HABLA Corpus (Hamburg Adult Bilingual LAnguage), with recordings of early French and German bilinguals (Kupisch al. 2012), and the German part of the phonological platform "*Interphonologie du Français Contemporain*" (IPFC-allemand), with well advanced German learners of French (Chervinski & Pustka 2010).

The purpose of the IFCASL corpus is to have a reliable empirical foundation to investigate phonetic and phonological deviations of both learner groups. In the past these aspects were either based on personal and anecdotal experience or on purely theoretical assumptions based on contrastive comparisons of the phonological systems.

The aims for constructing this corpus are as follows:

1. to perform analyses for phonetic and phonological research with respect to the prediction of the types of errors made by French and German learners,
2. to exploit the data for exercises in computer-assisted pronunciation training (CAPT) with a focus on feedback methods for the individual learner,
3. to provide training and test data for the improvement of automatic recognition of non-native speech, which is notoriously difficult,
4. to bring to the research community two non-native and two native phonetic corpora for the French-German language pair.

Thus, the design of the corpus was based on in-depth phonetic knowledge to predict the types of errors made by French and German learners. The research community will benefit from the automatic annotation of the entire corpus and the hand-labeling of more than 50% of the corpus, with a special emphasis placed on highlighting non-native pronunciation variants.

1.2 Possible phonetic and phonological interferences

Non-native speech shows general features such as a reduced pitch range, reduced speech fluency, slower articulation rate and an increased number of pauses and disfluencies.

German and French show marked differences in the systems of vowels and consonants. Most notably, German uses both tenseness and length to differentiate vowels, whereas in French, vowel length is not distinctive. In French, fricatives and plosives at the end of words can also occur as voiced consonants – in contrast to German where final devoicing applies. Therefore, interferences on the segmental level can be expected when French native speakers learn to speak German and vice versa.

The segmental level is also affected by the orthography.

Incorrect correspondences of spelling to pronunciation can lead to phonemic errors. Further sources of errors are cognate words which exist in both languages (often with the same spelling but different pronunciation).
On the suprasegmental level it can be expected that French learners have problems with the location of the lexical stress, which is not fixed in German words. Regarding sentence prosody the realization of pitch accents can be different in both languages, particularly when a contrast is expressed.
The IFCASL corpus can provide substantial empirical evidence for most of the conjectures regarding the phonological interferences in this language pair as listed in Table 1.

French speakers of German	*German speakers of French*
Realization of /h/ and glottal stop	Liaison and enchaînement consonantique
Missing aspiration of /p t k/	Suppression of aspiration of /p t k/
Realisation of final devoicing	Realization of final voicing
Consonant clusters and affricates	
Realization of [ç, x]	
Postvocalic /r/ as lowered schwa	Postvocalic /r/ as consonant
Reductions, elision, assimilations	
Vowel quality	
Vowel quantity	
Oral vowel + nasal consonant	Nasal vowels
Location of word stress	
Realization and location of pitch accents	
Location of contrastive accents	
Mistakes induced by orthography	
Mistakes induced by cognates	

Table 1: Main phenomena of expected phonetic and phonological interferences in the German-French language pair.

2. Description of the corpus

2.1 Speakers

The main part of the corpus contains read speech of about 100 persons: about 50 speakers with French as their first language (L1) and about 50 with German as L1. The pool of subjects includes learners at the beginner, intermediate and advanced level (balanced for gender). For each language, in addition to 40 adults (between 18 and 30 years) we recorded 10 teenagers (15/16 years of age, beginners).
The recruitment of the speakers was handled via advertisements at the campuses of the universities, and direct contact to secondary schools for the teenage learners.

2.2 Questionnaire

Each speaker was asked to complete a questionnaire. This included landmarks of the individual linguistic biography such as L1, age (residence in first 16 years and in school time), and highest educational degree. For each foreign language (L2) we asked for school time, stay abroad, and certificates. The subjects also gave a self-assessment of language skills, especially pronunciation, their motivation to learn this L2, their general attitude towards language learning, and their opinion on language learning with a computer.
Although the selection was not balanced for regional variation, the accompanying questionnaire revealed a great diversity of origin for our subjects. Most of the subjects hold a high school degree and had English as their dominant L2.
In addition to the questionnaire the subjects signed an agreement that their spoken data recorded for the corpus can be used for scientific purposes. For the teenagers their parents signed the agreement.

2.3 Sub-corpora

All speakers produced the entire material in their respective L2 as well as in their respective L1. An advantage of this corpus compared to other corpora is thus that we have L1 *and* L2 data of each of the 100 speakers for both languages. For this reason it can be considered as a "symmetric" corpus. As a result we have four sub-corpora annotated at the word and the phone level:

1. GF: German learners speaking French
2. FG: French learners speaking German
3. FF: French native speech
4. GG: German native speech

2.4 Design of reading material

Linguistic coverage for both languages comprises:

1. a phonetically rich design covering all phonemes, to support a reliable assessment of the entire phonemic inventory for each speaker,
2. the most important phenomena in the phonetics and prosody of French and German as a foreign language, respectively (e.g., vowel quantity, consonantal articulation and lexical stress),
3. phonological processes and alternations (e.g., vocalization of /r/ after vowels in the same syllable in German),
4. minimal pairs.

Moreover, cognates (e.g. "chance"), proper names (e.g. "Berlin", "Paris"), numbers and abbreviations were integrated in some sentences.

2.5 Recording conditions

There are four different recording conditions in which we recorded material in each language:

1. SR (**S**entences **R**ead): sentences to be read aloud,
2. SH (**S**entences **H**eard): sentences to be read aloud after listening to a native speaker,
3. FC (**F**ocus **C**ondition): sentences to be read aloud with a different word in focus,
4. CT (***ConTe***): a short story to be read aloud.

The SR part consists of 31 sentences presented item by item which had to be read aloud. The items were displayed orthographically on a computer screen. The SH part also consists of 29 sentences per language. Likewise the sentences to be produced were displayed orthographically on the screen but here the subjects produced each sentence after listening to the sentence read by a model native speaker. One purpose of this condition is to exclude or at least minimise spelling-induced errors. The sentences in the SR and the SH parts have a length between 3 and 16 words, and also include questions. These sentences contain all phonemes of the given language and selected minimal pairs.

In the focus condition (FC) there are two sentences per language that vary with respect to the word in focus. For instance "*Yvonne amène un ami.*" (Engl.: "Yvonne brings a friend.") varied between "***Yvonne*** *amène un ami.*", "*Yvonne* ***amène*** *un ami.*", "*Yvonne amène* ***un ami.***", and a broad focus condition. The subjects first listened to a question and then read aloud the answer. The focused word was indicated by capitalised letters. The purpose of the focus condition sentences was to elicit variable locations of sentence accents which can be realised in different ways in both languages.

The CT part is a narrative text which was selected to investigate prosodic phenomena such as speech fluency and prosodic phrasing beyond single sentences. The English fairy tale "The three little pigs" was translated into short versions in both languages of about 200 words. Both, the French and the German versions each contain 13 sentences.

2.6 Recording procedure

The recordings took place in quiet offices in Nancy (France) and Saarbrücken (Germany) as it would be the case in applications of computer-aided pronunciation training (CAPT). The mean duration of a recording session was about 50 minutes per speaker. Another 10 minutes were needed for the above mentioned questionnaire.

The recordings were performed with the JCorpusRecorder software (Colotte 2015) in both locations. In the parts SR and FC, each sentence was first displayed on a laptop. The recording started when the subject pressed a "record" button before being ready to speak, and a "stop" button to end the recording. Pressing the button "next" after a recording automatically prompted the next sentence. In the SH condition there was an additional button "listening to the golden speaker" which had to be pressed before recording. Each sentence production could be repeated as many times as wished. For this purpose buttons with the function of "play back" of the own recording and "delete" were displayed. For the CT production the subjects had the entire text displayed on the screen and additionally as a print-out. The subjects wore head-mounted close-talk microphones. The microphone was calibrated before each session and the signal intensity was automatically checked during the recordings (to avoid clicks or too weak signals).

2.7 Pilot corpus

In addition to the main corpus described in the previous sub-sections we also recorded a preliminary pilot corpus with 14 subjects (5 adults and 2 teenagers for each language). These recordings were used to test the technical performance, the designed reading material, the usability of the questionnaire, and the duration of the recording procedure (see Fauth et al 2013). For the recordings of the main corpus some changes were applied to the reading material, e.g. leaving out a second short story text and some focus sentences to reduce the duration of the session.

3. Annotation and segmentation

The quality of the corpus data heavily depends on the quality of the annotation, which was performed by means of forced speech-text alignment and subsequent corrections by human annotators.

So far, more than half of all data was manually re-labelled: the non-native sub-corpora (FG, GF) were manually re-annotated to 80% each, whereas the sub-corpora with native speech were corrected by hand to 60% for the French part and to 25% for the German part.

3.1 Labelling procedure

The procedure of segmentation and annotation took place in two steps:

1. Forced speech-text alignment was used to determine phone and word boundaries.
2. Manual re-annotation of phone labels and phone boundaries where necessary.

Step 2 was performed by trained student annotators (only for their own L1 which was either French or German). In total, 8 annotators for French and 9 annotators for German worked on the manual re-annotation. For the phone labels insertions, deletions and substitutions were marked along with labels for incorrect (de)voicing of stops and fricatives.

The annotation contains information on the levels of sentence, word and phone. For each audio file of the corpus, a Praat Textgrid file exists with the following six tiers (see Figure 1):

1. *Text*: The sentence in its orthographic transcription.
2. *Canon*: The sentence in a broad phonetic transcription (based on the French and German versions of SAMPA). The canonical form and main pronunciation variants were specified manually.
3. *Word*: Words in their orthographic transcription. Word boundaries were determined by forced alignment and corrected manually, if necessary.
4. *Align*: Phones taken from the canonical form. Phone boundaries were determined by forced alignment.
5. *Real*: Manually corrected annotations and boundaries of phones based on the actual realizations. (This tier is available only in the manually corrected data.)

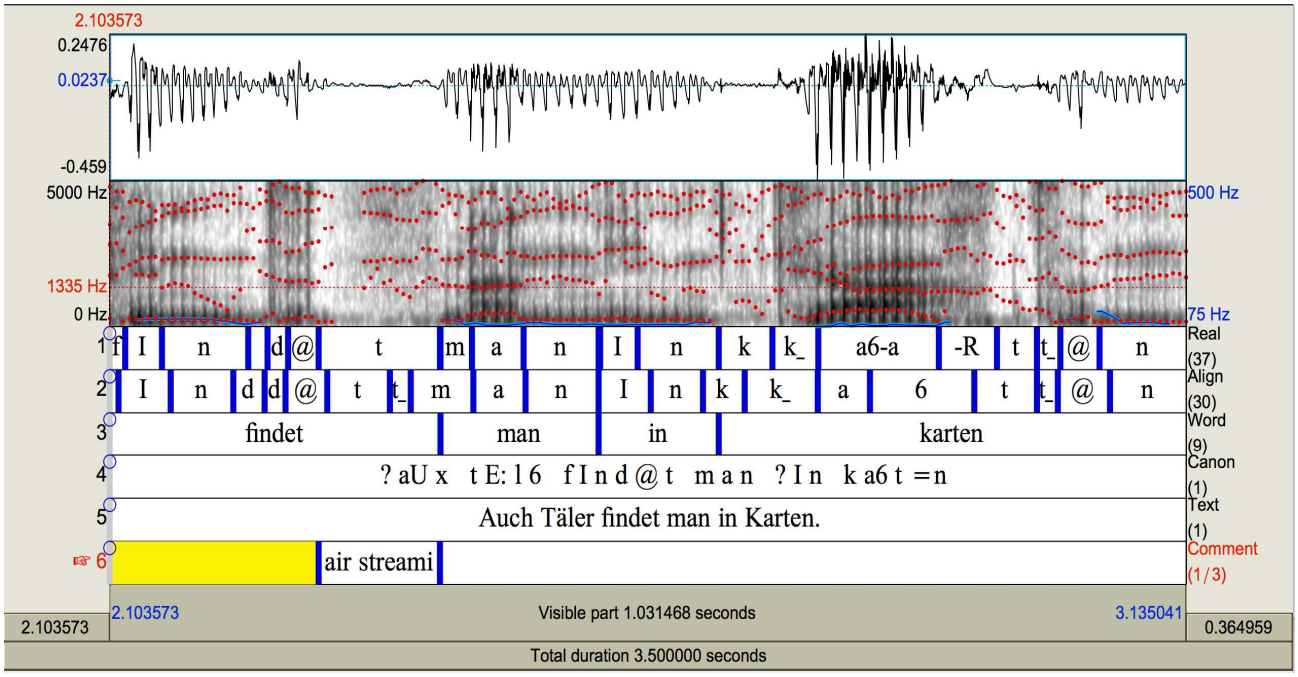

Figure 1: Example for a section of a German sentence with the six annotation tiers.

6. *Comment*: Comments on unusual sounds and noises but also on irregularities of prosody, e.g. incorrect lexical stress, or unusual intonation contour.

3.2. Inter-annotator agreement

In Mella, Fohr & Bonneau (2015) we investigated the inter-annotator agreement for the non-native and native French part of the IFCASL corpus. The agreement was evaluated by comparing the manual alignments by seven annotators to the manual alignment of an expert, for 18 sentences. The software CoALT (Comparing Automatic Labelling Tool) (Fohr & Mella 2012) was used to compare the results of the annotators to those of the expert annotator. Whereas results for the presence of the devoicing diacritic show a certain degree of disagreement between the annotators and the expert, there is a good consistency between annotators and the expert for segment boundaries as well as for insertions and deletions. We find a good overall agreement for boundaries between annotators and expert with a mean deviation of 7.6 ms and 93% of boundaries within 20 ms (see Table 2 for details).

	Native speech	Non-native speech
French	93.1%	90.5%
German	93.9%	90.4%

Table 2: Percentage of labels whose boundaries are within 20ms from those of the expert labeler regarding native and non-native speech.

3.3. Optimization of phonetic segmentation

An important use of the corpus is the optimization of the acoustic models for forced speech-text alignment in CAPT exercises. In Fohr & Mella (2015) we evaluated different HMM phone models for detecting accurate phone boundaries. The optimal parameters were determined by minimizing the number of phones in the non-native speech corpus whose boundaries are shifted by more than 20 ms compared to the manual boundaries. It was observed that the best performance was obtained by combining a French native HMM model with an automatically selected German native HMM model for each phone.

4. Phonetic and phonological analysis of interferences

The manual annotation is the starting point for various studies, from analyzing phonological interferences to deriving diagnosis and feedback.

The sub-corpora can be used to perform an analysis of native and non-native speech either in mono-lingual or in cross-linguistic studies. The following examples focus on interferences that could be expected on the basis of anecdotal and contrastive comparisons of the sound inventories of the two languages.

4.1. /h/ and glottal stop

An important problem on the segmental level is the production of /h/ by French speakers assuming the deletion of /h/ when speaking German. In Zimmerer et al. (2015) we showed that complete deletion was quite rare in our corpus. Although beginners sometimes omit /h/, they predominantly realize it as a glottal stop or other forms of glottalization. Advanced learners are more successful in producing /h/ native-like, that is as a voiceless or voiced glottal fricative depending on the voicing status of the left context.

4.2 Final devoicing

Voiced obstruents at the end of words are difficult for German speakers. In recent studies (Bonneau, 2015; Bonneau & Cadot, 2015) the realizations of the French voiced fricatives /z, Z/ by German non-native and French native speakers of our corpus were analyzed. Results show that the non-native realizations are strongly influenced by L1 (German) final devoicing, particularly for learners at a lower proficiency level. There was also an influence of spelling that leads in cases of "e" as in "*neige*" to a rather intense schwa at the end of words, a phenomenon that is important to consider for CAPT exercises.

4.3 Vowel quality and quantity

An experiment (Zimmerer & Trouvain 2015b) in which German listeners judged vowels in minimal pairs produced by the French learners of German indicate that these learners indeed have problems producing German vowels correctly. Beginners and advanced learners show lengthening and shortening errors. Furthermore, rounded vowels seem to pose more severe problems in L2 acquisition than unrounded vowels. These results were backed up by another study with a different method (Jouvet et al. 2015) in which the manually corrected annotation on the phone level allowed a detailed comparison of the realized sounds with the expected sounds. The analysis of phone confusion matrices for selected error-prone classes of sounds revealed that, for instance, French learners of German show complex interferences with the vowel contrasts for length and quality. This also refers to vowels like /E/ and /9/ which are also phonemes in French.

4.4 Pitch range

As an example for a suprasegmental topic, Zimmerer et al. (2014) analyzed the short story productions of the pilot corpus for pitch range. The results indicate that most speakers produce a smaller pitch range when they speak an L2 compared to their L1. In a follow-up study (Zimmerer et al. 2015) examining the pitch profiles for sentences taken from the SR and the SH part, both French and German speakers did not show pitch range differences in their native production, neither did they reduce their pitch range when speaking L2. A possible explanation for the difference to the prior study is that the length of the production (single sentences with c. 5 words vs. text with 200 words) influences the pitch range variability, and that the first study was based on the production of 7 speakers per language, whereas the second one had 40 German and 44 French native speakers as basis for comparison.

4.5 Pauses, tempo and fluency

Another example for a prosodic analysis of the short story productions (CT) is the investigation of pausing details of audible breathing, particularly in disfluent phases (Trouvain, Fauth & Möbius 2016). As expected, there were more frequent pauses and more frequent disfluencies in L2, as well as longer durations of pauses filled with breath noise than those without. However, the analysis also reveals that in fluent phases the vast majority of pauses contains an audible inhalation - which suggests a reinterpretation of the terms "unfilled" and "silent" pauses. Most disfluent phases are marked by genuinely silent pauses (i.e. without breathing noises), which are also shorter than those in fluent phases. So-called "filled pauses" are virtually absent. Surprisingly, French speakers use more but shorter pauses as an L2 pausing strategy than the Germans.

5. Applications for computer-assisted pronunciation training

The results of the phonetic analyses have important implications for language learning and teaching, particularly for individualized CAPT. Providing feedback faces the difficulty of segmenting non-native utterances, which deviate from the expected sequence of speech sounds. Missing/added phones, or incorrect acoustic features, substantially complicate the segmentation task. We are exploiting our manually annotated corpus to design more robust forced speech-text alignment algorithms by anticipating possible errors made by learners.

On the other hand, the non-native realizations are studied in order to investigate efficient acoustic feedback provided to learners, and how feedback can robustly interact with automatic segmentation provided by automatic speech recognition. We are especially considering vowel duration, lexical stress (Vakil & Trouvain 2015), F0, energy levels and voicing to help German learners master the voiced obstruents at the end of words in French. For instance the insufficiently realized contrast of vowel length and/or quality in L2 German is presumably combined with a perceptual deficit for this contrast. In order to help learners improve both perception and production of acoustically similar vowels, a prototype of a visual feedback tool was proposed (Carroll, Trouvain & Zimmerer 2015) that illustrates the differences between the sounds in listening and repetition exercises. The audio samples are accompanied with graphic representations of the first two formants and duration. The idea is that with repeated use, the learners can adjust their production of vowels to approximate spectrally and duration based targets derived from the native German productions.

6. Conclusion

We have presented a "symmetric" phonetic learner corpus with speech read aloud by speakers in their L2 and their L1 for the under-studied language pair French/German. It provides spoken data by many speakers with comparable productions, annotated and segmented on the word and the phone level, with a substantial amount of hand-correction.

One important general observation from our studies is that there is a significant degree of individual variation on top of more general L1-L2 interference patterns and that the

interference patterns are by no means symmetrical for the two languages.

Although the spoken data in the IFCASL corpus is restricted to scripted speech, there is plenty of material to explore for phonetic research, and to exploit for purposes in speech technology such as ASR of non-native speech. It will be the scope of future corpora to focus on dialogues with learners and other forms of unscripted speech.

Ongoing studies include topics such as using the material for designing exercises for automatic feedback and training, phonetic studies on pausing, phonological questions like the realization of contrastive focus, and perceptual tests on intelligibility, comprehensibility and foreign accentedness. Moreover, the IFCASL corpus has a big potential for many researchers to come up with their own research questions. The corpus will be made available for the interested scientific public for non-commercial research at the end of the year 2016.

7. Acknowledgements

This work was supported by the *Agence Nationale de Recherche* (ANR) and *Deutsche Forschungsgemeinschaft* (DFG) to the project IFCASL (PIs: Yves Laprie, Bernd Möbius, Jürgen Trouvain).

8. Bibliographical References

Carroll, P., Trouvain J., Zimmerer, F. (2015). A visual feedback tool for German vowel production. *Proc. 26. Konferenz Elektronische Sprachsignalverarbeitung* (ESSV), Eichstätt, pp. 177-184.

Chervinski, J. & Pustka, E. (2010). IPFC-allemand : une pré-enquête auprès de quelques étudiants munichois. *Journée IPFC2010: Interphonologie, corpus et français langue étrangère*, Paris, MSH.

Colotte, V. and Casano E. (2015) JCorpusRecorder. Technical Report, Université de Lorraine.

Fauth, C., Bonneau, A., Zimmerer, F., Trouvain, J., Andreeva, B., Colotte, V., Fohr, D., Jouvet, D., Jügler, J., Laprie, Y., Mella, O., Möbius,B. (2014). Designing a bilingual speech corpus for French and German language learners: a two-step process. *Proc. 9th Language Resources and Evaluation Conference* (LREC), Reykjavik, pp. 1477-1482.

Fohr, D. & Mella, O. (2012). CoALT; A Software for comparing automatic labelling tools. *Proc. 9th Language Resources and Evaluation Conference* (LREC),, Istanbul.

Fohr, D., Mella, O. (2015). Detection of phone boundaries for non-native speech using French-German models. *Proc. Workshop on Speech & Language Technology in Education* (SLaTE), Leipzig, pp. 181-182.

Jouvet, D., Bonneau, A., Trouvain, J., Zimmerer, F., Laprie, Y., Möbius, B. (2015). Analysis of phone confusion matrices in a manually annotated French-German learner corpus. *Proc. Workshop on Speech and Language Technology for Education* (SLaTE), Leipzig, pp. 107-112.

Kupisch, T., Barton, D., Bianchi, G. & I. Stangen, I. (2012). The HABLA-Corpus (German-French and German-Italian). In: T. Schmidt & K. Wörner (eds). *Multilingual Corpora and Multilingual Corpus Analysis*. Amsterdam: Benjamins, pp.163-179.

Mella, O., Fohr, D., Bonneau, A. (2015). Inter-annotator agreement for a speech corpus pronounced by French and German language learners. Proc. Workshop on *Speech and Language Technology in Education* (SLaTE), Leipzig, pp. 143-148.

Trouvain, J., Laprie, Y., Möbius, B., Andreeva, B., Bonneau, A., Colotte, V., Fauth, C., Fohr, D., Jouvet, D., Mella, O., Jügler, J., Zimmerer, F. (2013). Designing a bilingual speech corpus for French and German language learners. *Proc. Confer. on Corpus et Outils en Linguistique, Langues et Parole: Statuts, Usages et Mésuages*, Strasbourg, pp. 32-34.

Trouvain, J., Fauth, C. & Möbius, B. (2016). Breath and non-breath pauses in fluent and disfluent phases of German and French L1 and L2 Read Speech. *Proc. 8th Conference on Speech Prosody*, Boston

Vakil, A. & Trouvain, J. (2015). Automatic classification of lexical stress errors for German CAPT. *Proc. Workshop on Speech and Language Technology for Education* (SLaTE), Leipzig, pp. 47-52.

Zimmerer, F., Jügler, J., Andreeva, B., Möbius, B. Trouvain, J. (2014). Too cautious to vary more? A comparison of pitch variation in native and non-native productions of French and German speakers. *Proc. 7th Confer. on Speech Prosody*, Dublin, pp. 1037-1041.

Zimmerer, F., Andreeva, B., Jügler, J., Möbius, B. (2015). Comparison of pitch profiles of German and French speakers speaking French and German. *Proc. 18th International Congress of Phonetic Sciences* (ICPhS), Glasgow, 5 pp.

Zimmerer, F., Trouvain, J. (2015a). Perception of French speakers' German vowels. *Proc. Interspeech*, Dresden, pp. 1720-1724.

Zimmerer, F., Trouvain, J. (2015b). Productions of /h/ in German: French vs. German speakers. *Proc. Interspeech*, Dresden, pp. 1922-1926.