French Learners Audio Corpus of German Speech (FLACGS)

Jane Wottawa, Martine Adda-Decker

French Learners Audio Corpus of German Speech (FLACGS)

Jane Wottawa, Martine Adda-Decker
LPP, UMR 7018 CNRS - U. Paris 3 / Sorbonne Nouvelle
19, rue des Bernardins, 75005 Paris
jane.wottawa@univ-paris3.fr, martine.adda-decker@univ-paris3.fr

Abstract

The *French Learners Audio Corpus of German Speech* (FLACGS) was created to compare German speech production of German native speakers (GG) and French learners of German (FG) across three speech production tasks of increasing production complexity: repetition, reading and picture description. 40 speakers, 20 GG and 20 FG performed each of the three tasks, which in total leads to approximately 7h of speech. The corpus was manually transcribed and automatically aligned. Analysis that can be performed on this type of corpus are for instance segmental differences in the speech production of L2 learners compared to native speakers. We chose the realization of the velar nasal consonant /ŋ/. In spoken French, /ŋ/ does not appear in a VCV context which leads to production difficulties in FG. With increasing speech production complexity (reading and picture description), /ŋ/ is realized as [ŋg] by FG in over 50% of the cases. The results of a two way ANOVA with unequal sample sizes on the durations of the different realizations of engma indicate that duration is a reliable factor to distinguish between [ŋ] and [ŋg] in FG productions compared to the [ŋ] productions in GG in a VCV context. The FLACGS corpus allows to study L2 production and perception.

Keywords: L2 German, L1 German, French learners, realization of [ŋ]

1. Introduction

Mastering a second language (L2) means to be able to express oneself and to be understood by native speakers of that language. On one hand, native speakers are able to adapt their perception to accented speech but the learners' intelligibility has to be high in order to permit native listeners to achieve utterance recognition close to 100% (Bradlow and Bent, 2008). On the other hand, the pronunciation of a foreign language is conditioned by the phonological system of the L1. Mastering the phonological system of the L2 improves the communication with native speakers. Flege et al. (1997) state that L2 speech production is linked to the phonemic inventories of both the L1 and L2. The study highlights that erroneous speech production is a result of how close L1 and L2 sounds are in their acoustic properties and that the production skills of L2 learners do not only depend on perception skills in the L2.

We recorded the *French Learners Audio Corpus of German Speech* (FLACGS) to investigate where pronunciation of German differs between German native (GG) speakers and French learners of German (FG). This resource was created to identify pronunciation difficulties of French learners in German. Results of this corpus are used to develop a pronunciation training for French learners of German.

The long term aim of our research is to develop a training method that improves both pronunciation quality and perception accuracy in FG.

At the time, we created the FLACGS corpus (2014/2015), we found corpora for German native speech only (Kohler, 1996; Kohler, 2000) and investigations of French speakers' productions in English (Tortel and Hirst, 2010; Shoemaker, 2014; Grosbois, 2014; Horgues and Scheuer, 2014). In 2013, researchers from the Universities of Saarland and Nancy started the collection of the *Bilingual speech corpus for French and German language learners* (Fauth et al., 2014; Zimmerer et al., 2014; Zimmerer and Trouvain, 2015). This corpus is not yet publicly available.

In the following, first the FLACGS Corpus is presented and second a case study is carried out on the realization of engma in a VCV context in GG and FG across the three production tasks of the FLACGS Corpus.

2. The FLACGS Corpus

NAME
French Learners Audio Corpus of German Speech (FLACGS)
LANGUAGE
German
SPEAKERS
40 speakers (20 male and 20 female)
- 20 L1 German
- 20 L1 French, L2 German (Level of competence: A2-C2)
VOLUME
ca.7 h of speech (ca. 3.30h & ca. 3.30h) (35 250 words)
CONTENT
repeated, read and semi-spontaneous speech
TRANSCRIPTION
manual using the German orthography
ALIGNMENT
webMAUS (automatic) and manual checking

Table 1: Summary of the German FLACGS corpus

2.1. Participants

All participants were recruited in Paris, France. Participation was on a voluntary basis. The recordings took approximately 45 minutes per participant.

2.1.1. French learners of German (FG)

20 FG (10 women and 10 men) were recorded. The women were aged between 20 and 30 years, the men between 24 and 32. All FG as well as their parents had only French as a first language (L1). They auto-evaluated their competences in German based on the *Common European Framework of Reference for Languages* (CEFRL). In both gender groups, all levels were represented: A1/A2 up to C2. In English, all participants had an equal or superior competence compared to their German level.

2.1.2. German native speakers (GG)

20 GG (10 women and 10 men) were recorded. The women were aged between 22 and 47 years, the men between 30 and 45. All GG as well as their parents had only German as L1. Except for one female and one male participant who had no knowledge of the French language, all GG were high proficient in French (B1/B2 up to C2+ according to the CEFRL). Their knowledge of English was equal or inferior to their competences in French. The great majority lived in France for several years.

Even if the GG were born and raised in different regions of Germany, their productions can be considered as standard German as none of them had a noticeable regional accent.

2.2. Tasks

The participants performed three tasks of increasing production complexity:

1. **Repetition task** (audio input)**:**
   Participants heard small sentences over headphones they repeated immediately.

2. **Reading task** (text input)**:**
   Participants read aloud the texts *Nordwind und Sonne* and *Die Buttergeschichte*.

3. **Picture description:**
   The picture description task was the only task without linguistic input (see Figure 1).

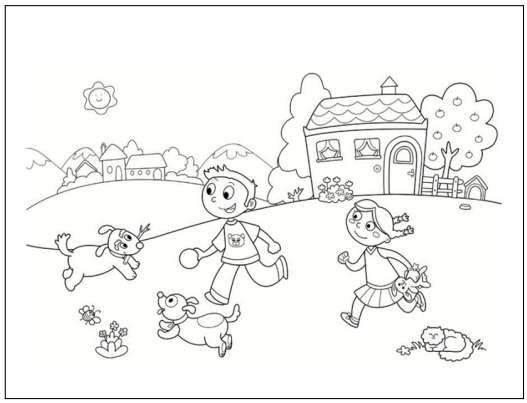

Figure 1: Picture description task

The repetition task aimed to investigate whether FG successfully produce lexical stress in different word positions, long and short vowel contrasts as well as consonants and consonant clusters that are unusual or different in the French language.

Carrier sentences (*Er sagt ... klar und deutlich* and *Ich sage ... klar und deutlich*) including 55 distinct words in central position were recorded by a female German native speaker. The participants listened to all the spoken utterances in a randomized order over headphones and repeated them.

The material of the repetition task was composed of words with lexical stress in different positions (first syllable, last syllable, penultimate syllable and ante-penultimate syllable), minimal pairs with long and short vowels e.g. *Hüte* /hyːtə/ and *Hütte* /hʏtə/, minimal pairs with a voiced or unvoiced plosive *glauben* /glaʊ̯bən/ and *klauben* /klaʊ̯bən/. Words that are difficult to pronounce because of their phonotactics for French natives: challenging consonants, clusters and glottal stops between vowels in adjacent syllables were included as well e.g. *Schächtelchen* /ʃɛçtəlçən/, *erobernde* /ɛrʔoːbɛʁndə/.

The participants were asked to read two texts *Nordwind und Sonne* and *Die Buttergeschichte*. These two texts have also been recorded in the Kiel corpus (Simpson et al., 1997). Both languages, French and German, use the Latin alphabet. But the letters and letter combinations do not code the same sounds e.g. *Mantel* is produced as [ˈmantl̩] by GG. FG are more likely to say [mɑ̃ˈtɛl] as the letter combination <an> corresponds to a nasal vowel in written French. Conflicting orthographic conventions are possible sources of pronunciation difficulties. For instance the graphic <z> is pronounced /z/ in French but /ts/ in German. Another example is the grapic <au> which is pronounced as the vowel /o/ in French but as the diphtong /aʊ̯/ in German.

The aim of the reading task is twofold:

(i) check overall FG pronunciation difficulties, when reading;

(ii) focus on difficulties which may arise due to conflicting orthographic conventions between German and French.

The reading task also allows us to compare prosodic patterns in different places of the utterance e.g. to compare how word stress is realized in the beginning, the middle and the end of an utterance with respect to prosody.

The description task aims to collect semi-spontaneous speech. All participants described the same picture. We concentrated our analysis on isolated words like *Haus, Mädchen, Junge* and *Sonne*. The image description is the only task where the participants did not have a linguistic support (written sentences, spoken utterances) to help them with their speech production. Before the participants started the image description, we made sure they knew the names of the items and actions represented on the picture.

2.3. Transcription and Alignment

First, manual checking and correction, if necessary, of the orthographical text of the repeated and read material, and an orthographical transcription of the spontaneous speech part were realized. Second, the webMAUS (Munich AUtomatic Segmentation web service) (Schiel, 1999; Kisler et al., 2012) performed the alignment of the speech sig-

nal with its transcription. This aligner generates a TextGrid file that can be opened with Praat (Boersma and Weenink, 2001).
MAUS uses orthographic transcriptions to segment the speech signal. The orthographic transcriptions were realized manually and took hesitations, disfluencies and false starts into account. The TextGrids generated by MAUS comprise three tiers: the orthographic word, the canonical pronunciation of the word and the aligned phones.
The automatic alignment of each sound file was checked manually for boundaries and labelling. Phone boundaries of targeted words were manually corrected if necessary. We also checked some aligned pronunciations, for instance when MAUS had to perform a graph to phone conversion for words that were not included in its dictionary. Performing those adjustments took about two minutes for one minute of automatically aligned speech.

3. Case study - the consonant /ŋ/

3.1. The engma in German and French

The velar nasal consonant /ŋ/ is part of the German phonemic inventory, whereas it is missing in French. This consonant appears at the end of German syllables, before stop consonants or in intervocalic position if the second vowel is unstressed, e.g. *schwing* /ʃvɪŋ/, *schwingt* /ʃvɪŋt/, *schwingen* /ʃvɪŋən/, *Schwingung* /ʃvɪŋʊŋ/. The German orthography codes the consonant with <ng>.
In French, the engma only exists in loan words that are often borrowed from the English language e.g. *parking* [parkiŋ]. In those loan words, the /ŋ/ only appears in word endings. The English orthographic coding is <ng> as well. In French native productions of such loan words, one may often perceive an additional homorganic plosive after the engma proper.

3.2. Production of /ŋ/ in German words

In the following, we are investigating acoustic differences in engma productions between GG and FG speakers in a VCV context in the repetition task. Repetition is the less complex task in respect to speech production.
Figures 2 and 3 show the respective productions of the German word *singen* by a native speaker [zɪŋən] (e.g. Fig. 2) and by a French learner of German [zɪŋgən] (e.g. Fig. 3) The word *singen* presents the nasal consonant /ŋ/ in a VCV context. The /ŋ/ is realized as a smooth voiced segment, as shown in the spectrogram of the GG speaker (e.g. Fig. 2).
The spectrogram of the FG shows a different realization of the expected /ŋ/. First, we observe that the portion coded as /ŋ/ in the third tier of the spectrogram is longer than the /ŋ/-portion of the GG although the global word duration is shorter for the FG than for the GG. The longer engma duration of the FG speaker can thus not be explained by a lower speech rate.
Second, the labelled [ŋ] segment in FG's spectrogram shows two distinct parts which could be more precisely described as a nasal consonant [ŋ] followed by a voiced plosive consonant [g]. In French, the /ŋ/ sound between vowels does not exist, and French speakers tend to add an homorganic plosive before the next vowel.

In a reading task, one could explain the [g] insertion by the speaker's difficulties with the writing convention <ng> , which might trigger a [g] insertion. However, the repetition task only provided audio input to the speakers which suggests that the FG's realization of /ŋ/ in a VCV context is not only linked to the graphic coding of the /ŋ/-sound but that French learners might "repair" the invalid sound structure with respect to French phonotactics by inserting a plosive consonant before the next vowel after the /ŋ/. Studies of Polish learners of English show similar plosive insertion after the sound /ŋ/ (Szpyra-Kozłowska, 2014; Bryla-Cruz, 2014).

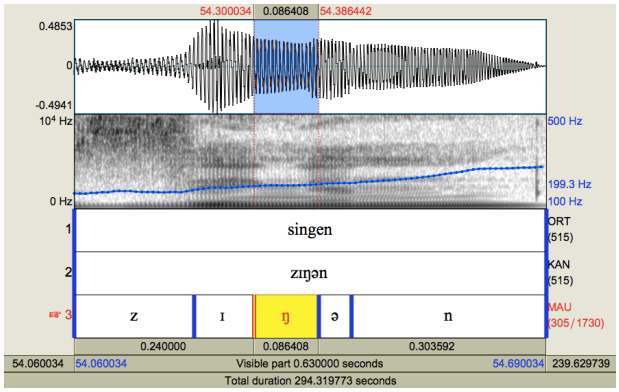

Figure 2: Spectrogram *singen* GG, female speaker, velar nasal

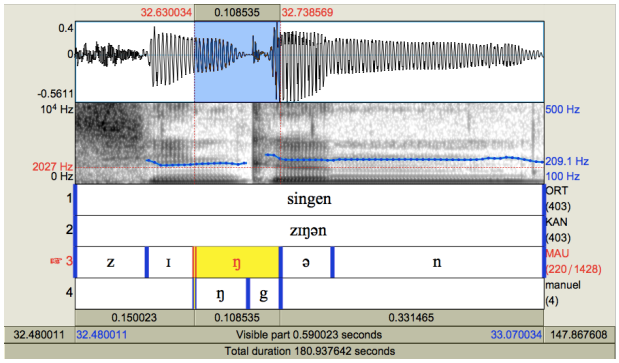

Figure 3: Spectrogram *singen* FG, female speaker, velar nasal followed by an obstruent [g]

3.3. Results

We took all the cases where /ŋ/ appeared in a VCV context in the oral productions of FG. Table 2 shows the percentages of [ŋ] and [ŋg] productions respective to each tasks. GG speakers do not appear in the table because they produce [ŋ] exclusively. Only one of the French speakers (female, trained linguist) was able to produce a German like engma over all three tasks. The other 19 French native speakers produced the [ŋg] at least once during their recording.
In Table 2, we can observe that FG produce most [ŋg] in the reading task. This result could be explained by the written input. As the /ŋ/ is coded as <ng> by the German orthography, the plosive production after the velar nasal could be due to the decoding efforts by the FG.
In the picture description task, FG production of the word *Junge* was either [juŋə] or [juŋgə]. Even if multiple occur-

rences of the same word were uttered by the same speaker, there was no alternation of [ŋ] and [ŋg].

TASK	(Tokens)	[ŋ]	[ŋg]
repetition	(80)	55%	45%
reading	(60)	32%	68%
description	(78)	45%	55%

Table 2: Overview of /ŋ/ realizations in FG speakers

We are also interested in the differences between GG and FG realizations of /ŋ/ reagrding their acoustics. In Figure 4, the duration means of the engma realizations are plotted across the three speech production tasks: GG are plotted in blue and FG are plotted in yellow. For FG, we found two different realizations [ŋ] and [ŋg]. In the German natives, only [ŋ] productions are observed.
Statistical analyses were carried out using a two-way ANOVA with unequal sample sizes. First, we can exclude an effect of the task on the durations of [ŋ] and [ŋg] for both GG and FG. Second, across all tasks, we observe a significant duration difference between the engma of GG and the [ŋg] of FG whereas there is no significant difference between the [ŋ] of GG and FG, except for the repetition task. GG show rather a great variability in the reading task regarding segment duration. This is due to different individual performances on the task. GG speakers are used to read in German some of them privileged a fast reading in order to finish quickly others chose to read aloud carefully in order to interpret the stories. As a consequence, in reading, speech rate varies a lot in GG speakers what leads to varying segment durations.

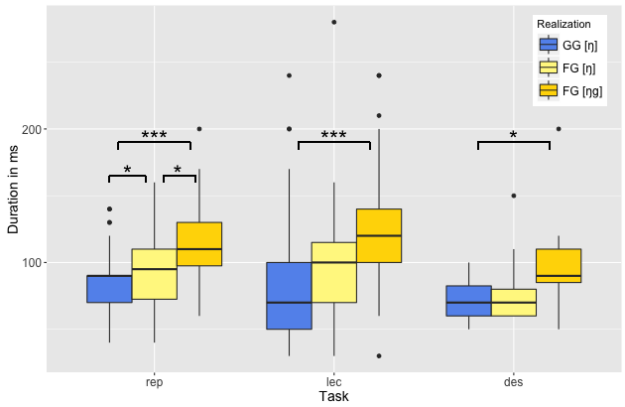

Figure 4: Duration of [ŋ] and [ŋg] in respect to L1
$^{*}\ p \leq 0,05$; $^{**}\ p \leq 0,01$; $^{***}\ p \leq 0,001$

4. Discussion and perspectives

To summarize, we created the *French Learners Audio Corpus of German Speech* (FLACGS). The corpus contains repeated, read and semi-spontaneous speech of 20 German native speakers and 20 French learners of German. This corpus of ca. seven hours of speech and was created to improve pronunciation teaching by detailed phonetic explanations and spectrograms. We plan to distribute this resource at the end of the PhD. The corpus could then be used not only for analyses regarding second language learning but also for automatic accented speech recognition, for instance.
In the presented case study on the engma production of FG, we found high rates (ca. 50%) of homorganic plosive insertion (higher in reading task) if the engma appears in a VCV context in German words. Durations between German natives [ŋ] and French learners [ŋg] are significantly different in all three speech production tasks. Duration can thus be used as a cue to decide whether FG produced [ŋ] or [ŋg] compared to a German native control population. Duration for FG [ŋ] and [ŋg] are not significantly different within the group means.
Further studies on the FLACGS corpus will investigate the acoustic differences between the fricatives /ʃ/ and /ç/, vowel quality in GG and FG and lexical stress realization in FG.

5. Acknowledgements

This work was made possible through a Sorbonne Nouvelle University PhD funding to the first author. It was also supported by the French Investissements d'Avenir - Labex EFL program (ANR-10-LABX-0083).

6. Bibliographical References

Boersma, P. and Weenink, D. (2001). Praat, a system for doing phonetics by computer.

Bradlow, A. R. and Bent, T. (2008). Perceptual adaptation to non-native speech. *Cognition*, 106(2):707–729.

Bryla-Cruz, A. (2014). Priorities in teaching english pronunciation to poles: Empirical evidence from scottish native listeners. In *Language Skills: Traditions, Transitions and Ways Forward*, pages 196–213. Cambridge Scholars Publishing.

Fauth, C., Bonneau, A., Mella, O., Colotte, V., Fohr, D., Jouvet, D., Laprie, Y., and Trouvain, J. (2014). Constitution d'un corpus de français langue etrangère destiné aux apprenants allemands. In *SHS Web of Conferences*, volume 8, page 14.

Flege, J. E., Bohn, O.-S., and Jang, S. (1997). Effects of experience on non-native speakers' production and perception of english vowels. *Journal of phonetics*, 25(4):437–470.

Grosbois, M. (2014). Mobilité internationale et production orale en anglais l2. réflexion, médiation, dénativisation. *Les dossiers des sciences de l'éducation*, (32):65–80.

Horgues, C. and Scheuer, S. (2014). "i understood you, but there was this pronunciation thing…": L2 pronunciation feedback in english/french tandem interactions. *Research in Language*, 12(2):145–161.

Kisler, T., Schiel, F., and Sloetjes, H. (2012). Signal processing via web services: the use case webmaus. In *Digital Humanities Conference 2012*.

Kohler, K. J. (1996). Labelled data bank of spoken standard german: the kiel corpus of read/spontaneous speech. In *Spoken Language, 1996. ICSLP 96. Proceedings., Fourth International Conference on*, volume 3, pages 1938–1941.

Kohler, K. J. (2000). Investigating unscripted speech: Implications for phonetics and phonology. *Phonetica*, 57(2-4):85–94.

Schiel, F. (1999). Automatic phonetic transcription of non-prompted speech.

Shoemaker, E. (2014). The exploitation of subphonemic acoustic detail in l2 speech segmentation. *Studies in Second Language Acquisition*, 36(04):709–731.

Simpson, A. P., Kohler, K. J., and Rettstadt, T. (1997). The kiel corpus of read/spontaneous speech: Acoustic data base, processing tools, and analysis results.

Szpyra-Kozłowska, J. (2014). *Pronunciation in EFL instruction: a research-based approach*, volume 82. Multilingual Matters.

Tortel, A. and Hirst, D. (2010). Rhythm metrics and the production of english l1/l2. In *Proceedings of Speech Prosody*.

Zimmerer, F. and Trouvain, J. (2015). Productions of/h/in german: French vs. german speakers. In *Sixteenth Annual Conference of the International Speech Communication Association*.

Zimmerer, F., Jügler, J., Andreeva, B., Möbius, B., and Trouvain, J. (2014). Too cautious to vary more? a comparison of pitch variation in native and non-native productions of french and german speakers. In *Proceedings of the 7th Speech Prosody Conference, Dublin*.