

Book of Extended Abstracts

Workshop on

Phonetic Learner Corpora

12 August 2015 in Glasgow

Satellite workshop of the

18th International Congress of Phonetic Sciences

Workshop organised and book of abstracts edited by

Jürgen Trouvain (Saarland University, Saarbrücken)

Frank Zimmerer (Saarland University, Saarbrücken)

Mária Gósy (Hungarian Academy of Sciences, Budapest)

Anne Bonneau (LORIA, Nancy)

Preface

Most of the data used to explore and explain phonetic variation in the speech of foreign language learners was recorded on a small-scale in experimental research. Likewise language learner corpora are usually based on written rather than on spoken data whereas phonetically annotated corpora of various speech styles do not explicitly consider language learners as speakers. However, we notice a growing number of research using large-scale collections of phonetic learner data as well as the development and investigation of full-fledged phonetic learner corpora.

The aim of this half-day workshop is to bring together researchers working with large-scale data sets of speech in a background of foreign and second language learning. The target audience comprises colleagues setting up learner corpora, experts in phonetic corpora, researchers with phonetic and phonological experiments in L2 acquisition, and those with an interest in phonetic aspects of L2 teaching and non-native speech in general.

At this meeting 17 papers from 10 countries will be presented to nearly 40 participants. Our thanks go to all authors and presenters for their work. Particularly we would like to thank very much our invited speakers for their keynotes: Anke Lüdeling (Humboldt University Berlin, Germany), as well as Sylvain Detey (Waseda University, Japan) and Isabelle Racine (Geneva University, Switzerland). We are also grateful to the organisers of ICPH2015 who provided the room and helped with the infrastructure. Last but not least we are happy that this workshop is supported by the Deutsche Forschungsgemeinschaft (DFG) and the Agence Nationale de Recherche (ANR) through the project IFCASL where also the homepage of the workshop is located at <www.ifcasl.org>.

We wish all participants a fruitful and interesting meeting and we hope that it will be a kick-off for further meetings and publications.

The organisers

Programme

Location of the workshop is the room *Boisdale 2* in SECC Glasgow, the ICPHS Congress site.

Registration

13.45 - 14.00

Welcome

14.00-14.10

Invited talk

14.10 - 14.50

Sylvain Detey, Isabelle Racine

The Interphonology of Contemporary French (IPFC): An international corpus-based L2 phonology research programme

Regular talk

14.50 - 15.10

Jacques Koreman, Olaf Husby, Keivan Hedayatfar & Øyvind Bech

Learning from L2 learners to improve pronunciation training

Poster session 1

15.10 - 15.15

Flash demonstrations of posters of session 1

15.15 - 16.00

Coffee/tea break with poster session 1

Helene N. Andreassen, Nadine Herry-Bénit, Takeki Kamiyama & Véronique Lacoste

The ICE-IPAC project: Testing the protocol on Norwegian and French learners of English

Nicolas Ballier

Determining learner pronunciation models with the K-NN algorithm

Mario Carranza

The influence of intelligibility, comprehensibility and degree of foreign accent in evaluating and categorizing non-native pronunciation errors

Christoph Draxler

BAS resources for phonetic learner corpora

Frank Zimmerer, Jürgen Trouvain & Anne Bonneau

One corpus, one research question, three methods - German vowels produced by French speakers

Invited talk

16.00 - 16.40

Anke Lüdeling, Simon Sauer, Malte Belz & Christine Mooshammer

Error annotation in spoken learner corpora

Regular talks

16.40 - 17.00

Rosemary Orr, David van Leeuwen, Jacky Zoë de Rode, Georg Lohfink & Hugo Quené

L1 phonetic drift in Dutch L2 speakers of English

17.00 - 17.20

Calbert Graham, Andrew Caines & Paula Buttery

Phonetic and prosodic features in automated spoken language assessment

Poster session 2

17.20 - 17.25

Flash demonstrations of posters of session 2

17.25 - 18.10

Refreshment break with poster session 2

Taekehiko Makino

English read by Japanese phonetic corpus

Einar Meister & Lya Meister

Development and use of the Estonian L2 corpus

Adrien Méli

Addressing phonemic acquisition: A normalization-dependent procedure?

Riikka Ullakonoja, Hannele Dufva, Mikko Kuronen, Maria Kautonen & Elina Tergujeff

Using language testing corpora to investigate L2 pronunciation

Elisabeth Delais-Roussarie, Fabián Santiago & Hi-Yon Yoo

The extended COREIL corpus: first outcomes and methodological issues

Regular talks

18.10 - 18.30

Mária Gósy, Dorottya Gyarmathy & András Beke

The development of a Hungarian-English learner speech database and a related analysis of filled pauses

18.30 - 18.50

Malte Belz, Simon Sauer, Anke Lüdeling & Christine Mooshammer

Repair behaviour of advanced German learners in the Berlin Map Task Corpus

Closing

18.50 - 19.00

THE INTERPHONOLOGY OF CONTEMPORARY FRENCH (IPFC): AN INTERNATIONAL CORPUS-BASED L2 PHONOLOGY RESEARCH PROGRAMME

Sylvain Detey, Isabelle Racine

Waseda University (SILS), Geneva University (ELCF)
detey@waseda.jp, Isabelle.Racine@unige.ch

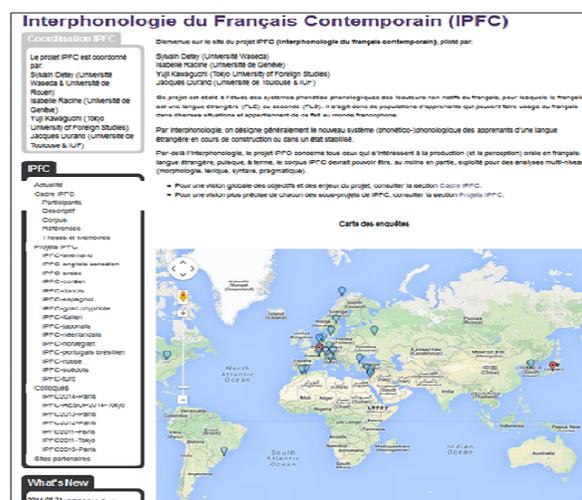
Keywords: corpus, French, interphonology, IPFC

1. ORIGINS

In 2014, the first *Handbook of Corpus Phonology* [10] was published by Oxford University Press, including a chapter describing the field of L2 corpus-based phonology written by Ulrike Gut [14], five years after the publication of her monograph about the LeaP corpus [13], which can be considered as a landmark corpus in the area. The inclusion of such a chapter in the handbook is an acknowledgement of the need to extend the methods and principles of corpus phonology to the field of non-native speakers, at a time when L2 pronunciation research is getting over the past theoretical or disciplinary boundaries (phonology vs phonetics; linguistics vs psycholinguistics; L2 acquisition studies vs L1 variationist sociolinguistic studies) and benefits from a new momentum. Unsurprisingly, L2 English has been the focus of most L2-oriented corpus-based phonology-phonetics projects (e.g. LeaP [13], AESOP [21]), but L2 Dutch has also been a proficient domain in connection with educational objectives in the field of Computer-Assisted Pronunciation Training programmes [16]. It is only recently that similar ventures have been launched for French [2, 15], with the *Interphonologie du Français Contemporain* (IPFC) project at its forefront. The IPFC project originally stemmed from the *Phonologie du Français Contemporain* (PFC) research programme [11]: a large-scale phonological survey across the French-speaking world launched in the late 1990s by three French phonologists, Jacques Durand, Bernard Laks and Chantal Lyche, who had become wary of the often limited set of data on which most French phonology had been relying. Standing against ‘armchair’ linguistics and in tune with the technical developments of oral databases and advances in sociophonetics, they designed a recording protocol inspired by William Labov’s work which could be used with most speakers to ensure data comparability. The PFC corpus (www.projet-pfc.net) turned out to include multilingual speakers

(in Africa, Canada, Louisiana), whose status as ‘native’ speakers of French could sometimes be debatable. Hence the launch of a dedicated ‘non-native’ avatar of PFC in 2008, the first of its kind for L2 French, with a Japanese team [5], immediately followed by a Spanish team, trying to combine the principles of PFC with the lessons learned from other projects, particularly in L2 English [13, 14]. The IPFC project now includes different teams representing 16 first languages (German, Alemannic, English, Arabic, Korean, Danish, Spanish, Greek, Italian, Japanese, Dutch, Norwegian, Portuguese, Russian, Swedish, Turkish) (Fig. 1), whose members (linguists and language educators) share an interest in L2 French pronunciation and are keen on exploring non-native varieties as part of the sociolinguistic continuum on which native, plurilingual and non-native speakers can be examined. It is important to note that the educational perspective is an integral part of the project, which partly explains some of its methodological orientations [9].

Figure 1: The website of the IPFC project with the different teams and survey points.



2. METHODOLOGY

2.1. Recording protocol

The protocol is made up of 6 tasks, two of which (b and d) are common to the PFC project for comparability purposes: a) repetition of an L1-specific wordlist¹, b) reading out of the PFC wordlist, c) reading out of the L1-specific wordlist², d) reading out of the PFC text, e) interview with a native speaker, f) interaction between two non-native speakers. A sociolinguistic questionnaire as well as a consent form are also included.

2.2. Data processing

Following the rationale adopted in PFC regarding data interoperability on the one hand and the pitfalls of phonetic transcriptions for large data-sets on the other hand, the audio recordings are orthographically transcribed with text-to-sound alignment in Textgrid files used with *Praat* [1]. Specific transcription conventions were designed to handle the characteristics of non-native speech [20].

2.3. Data analyses

Since one of the objectives of the corpus was to process as automatically as possible large sets of data, and following PFC's stance on variation, with a strong educational perspective in the case of IPFC, we decided to adopt and extend the coding system used in PFC for schwa and liaison, as an intermediate step between rough perceptual categorization (correct/incorrect) and fine-grained acoustic analysis (with its limits) [4]. The overall structure of each code is divided into four sections: 1) target structure, 2) left context, 3) right context, 4) perceptual assessment, primarily in terms of target-likeness (e.g. for nasal vowels: nasality, quality, postvocalic consonantal excrescence [8]). Alphanumeric codes were designed for consonants, oral vowels, nasal vowels, liaison and consonant clusters, and human coders, on the basis of their perceptual assessment of the production, insert the code in the orthographic transcription right after the structure under scrutiny, using separate tiers for each phenomenon (Fig. 2). The files are then analyzed with *Dolmen*, a phonological concordancer developed for the project by Julien Eychenne [12] (Fig. 3), which allows users to perform queries in the coded corpus and recover the requested items with several options. *Dolmen* provides descriptive statistics for code-based queries, and digs out the corresponding occurrences in concordance lines with the possibility of opening the sound files in *Praat*. A multiple-blind assessment option is included in the system.

Figure 2: An example of a sequence from the PFC text coded for nasal vowels by two different coders, and for liaison by one coder in a TextGrid file opened with Praat.

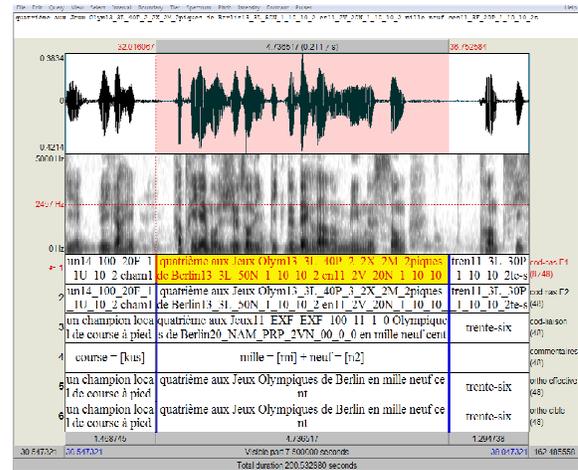
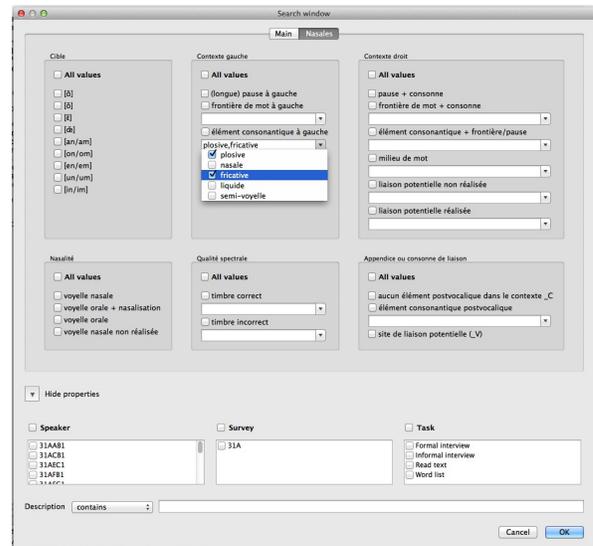


Figure 3: The Dolmen interface for the nasal vowels in the IPFC project.



Despite its obvious limitations in terms of phonetic description accuracy, this approach has proved successful so far (congruency between multiple subject psycholinguistic tests and coded results for isolated words), offering several advantages (especially for continuous speech): it can delineate specific data subsets for extensive analyses, and be used in the framework of perceptual studies, with a clear connection to perceptual norms. Last but not least, it really fits, in our view, with the overall objective and method of corpus phonology (but not laboratory phonetics at this stage).

3. ILLUSTRATIONS

Among the common objects of interest for all IPFC teams, nasal vowels and *liaison*, two typically difficult structures to be mastered by learners of French, have been extensively studied in the project [4, 7, 17, 18]. For several reasons (e.g. the relative difficulty of acoustical measures), the French nasal vowels have been a good benchmark to test and develop our approach, both with beginners and advanced learners. As for *liaison*, which has been a central object of study in the PFC project, it is also of particular interest in the case of L2 learners since it must be analyzed in a multidimensional manner: both from a segmental and a suprasegmental perspectives, but also at the interface between phonology, morphosyntax, lexis and orthography. Other elements have also been studied (high rounded vowels, voiced plosives, liquid consonants, lexical stress), and each team has its own focus (e.g. final consonant devoicing among Germanic languages speakers and vocalic epenthesis among Japanese learners). Having a common coding system for all L1-specific surveys, with the *Dolmen* application to perform cross-corpus queries, is one of the main methodological assets of the IPFC project to carry out comparative analyses between different groups of learners (e.g. *liaison* production by Italian and Spanish learners of French, with 4788 coded *liaison* sites [19]).

4. PERSPECTIVES³

Building up an international database such as the one we are striving to achieve in IPFC takes time. Even though most of the methodological features of the project are now set, we are still in the process of developing: (i) a full-fledged searchable database, (ii) automatic functions in *Dolmen* to provide richer descriptions of the learners' productions, (iii) guidelines to evaluate our data with semi-manual acoustic analyses on the one hand and automatic machine assessment on the other hand, (iv) pedagogical applications for syllabus design and pronunciation training. For more information about the IPFC project, see: <http://cblle.tufs.ac.jp/ipfc/>.

5. REFERENCES

- [1] Boersma, P., Weenink, D. 2015. *Praat: doing phonetics by computer*. Version 5.4.08.
- [2] Delais-Roussarie, E., Yoo, H. 2011. Learner corpora and prosody: from the COREIL corpus to principles on data collection and corpus design. *Poznań Studies in Contemporary PSCIL*, 471: 28-39.
- [3] Detey, S. 2005. *Interphonologie et représentations orthographiques. Du rôle de l'écrit dans l'enseignement / apprentissage du français oral chez des étudiants japonais*. PhD dissertation, University of Toulouse-Le Mirail.
- [4] Detey, S. 2012. Coding an L2 phonological corpus: from perceptual assessment to non-native speech models – an illustration with French nasal vowels. In: Tono, Y., Kawaguchi, Y., Minegishi, M. (eds), *Developmental and Crosslinguistic Perspectives in Learner Corpus Research*. Amsterdam: John Benjamins, 229-250.
- [5] Detey, S., Kawaguchi, Y. 2008. Interphonologie du Français Contemporain (IPFC) : récolte automatisée des données et apprenants japonais. *Phonologie du français contemporain : variation, interfaces, cognition*. Paris, 11-13 Dec. 2008.
- [6] Detey, S., Nespoulous, J.-L. 2008. Can orthography influence L2 syllabic segmentation? Japanese epenthetic vowels and French consonantal clusters. *Lingua. International Review of General Linguistics*, 118(1), 66-81.
- [7] Detey, S., Racine, I., Eychenne, J., Kawaguchi, Y. 2014. Corpus-based L2 phonological data and semi-automatic perceptual analysis: the case of nasal vowels produced by beginner Japanese learners of French. *Proc. 15th Interspeech* Singapore, 539-544.
- [8] Detey, S., Racine, I., Kawaguchi, Y. 2014. Des modèles prescriptifs à la variabilité des performances non-natives : les voyelles nasales des apprenants japonais et espagnols dans le projet IPFC. In: Durand, J., Kristoffersen, G., Laks, B. (eds), *La phonologie du français : normes, périphéries, modélisation. Hommage à Chantal Lyche*. Paris: Presses Universitaires de Paris Ouest, 197-226.
- [9] Detey, S., Racine, I., Kawaguchi, Y., Zay, F. in press. Variation among non-native speakers: the InterPhonology of Contemporary French. In: Detey, S., Durand, J., Laks, B., Lyche, C. (eds), *Varieties of Spoken French: a Source Book*. Oxford: Oxford University Press.
- [10] Durand, J., Gut, U., Kristoffersen, G. (eds) 2014. *The Oxford Handbook of Corpus Phonology*. Oxford: Oxford University Press
- [11] Durand, J., Laks, B., Lyche, C. 2014. French phonology from a corpus perspective: the PFC programme. In: Durand, J., Gut, U., Kristoffersen, G. (eds), *The Oxford Handbook of Corpus Phonology*. Oxford: Oxford University Press, 486-497.
- [12] Eychenne, J., Paternostro, R. in press. Analyzing transcribed speech with *Dolmen*. In: Detey, S., Durand, J., Laks, B., Lyche, C. (eds), *Varieties of Spoken French: a Source Book*. Oxford: Oxford University Press.
- [13] Gut, U. 2009. *Non-native Speech: a Corpus-based Analysis of Phonological and Phonetic Properties of L2 English and German*. Bern: Peter Lang.
- [14] Gut, U. 2015. Corpus phonology and second language acquisition. In: Durand, J., Gut, U., Kristoffersen, G. (eds), *The Oxford Handbook of Corpus Phonology*. Oxford: Oxford University Press, 286-301.
- [15] Landron, S., Paillereau, N., Nawafleh, A., Exare, C., Ando, H., Gao, J. 2010. Le corpus PhoDiFLE : un corpus commun de français langue étrangère pour

- une étude phonétique des productions de locuteurs de langues maternelles plurielles. *Cahiers de Praxématique*, 54/55, 73-86.
- [16] Neri, A., Cucchiarini, C., Strik, H. 2006. Selecting segmental errors in L2 Dutch for optimal pronunciation training. *IRAL*, 44, 357-404.
- [17] Racine, I., Detey, S. (eds) to appear. L'apprentissage de la liaison en français par des locuteurs non natifs : éclairage des corpus oraux. Special issue of *Bulletin VALS-ASLA*, 102.
- [18] Racine, I., Detey, S. Buehler, N., Schwab, S., Zay, F., Kawaguchi, Y. 2010. The production of French nasal vowels by advanced Japanese and Spanish learners of French: a corpus-based evaluation study. *Proc. 6th New Sounds Poznan*, 367-372.
- [19] Racine, I., Paternostro, R., Falbo, C., Janot, P., Murano, M. 2014. La liaison chez les hispanophones et les italophones : du texte lu à la conversation. *Rencontres FLORAL 2014 : corpus oraux et enseignement de la prononciation en FLE & interphonologie et corpus oraux*. Paris, 8-9 Dec. 2014.
- [20] Racine, I., Zay, F., Detey, S., Kawaguchi, Y. 2011. De la transcription de corpus à l'analyse interphonologique : enjeux méthodologiques en FLE. *Travaux Linguistiques du CerLiCO*, 24, 13-30.
- [21] Visceglia, T., Tseng, C.-Y., Kondo, M., Meng, H., Sagisaka, Y. 2009. Phonetic aspects of content design in AESOP (Asian English Speech cOrpus Project). *Oriental-COCOSDA*, Urumuqi.

¹ The lists include items common to all surveys (e.g. words with nasal vowels, since these vowels are difficult to acquire for most learners of French) and words specifically chosen for specific groups of learners (e.g. items with consonantal clusters for Japanese learners).

² The L1-specific wordlist repetition-reading tasks included in the protocol aim at taking into account the impact of the orthographic factor in the elicitation process [3, 6].

³ The research presented here has been partly supported by the Japanese Society for the Promotion of Science through two Grants-in-Aid for Scientific Research (B) n°2332012 and n°15H03227 to S. Detey, as well as by the University of Geneva and by the Fonds National Suisse de la Recherche Scientifique (subside n° 100012_1321441 to Isabelle Racine). We wish to thank Yuji Kawaguchi (co-director of the IPFC project), Julien Eychenne, Jacques Durand, Chantal Lyche, Bernard Laks, Mariko Kondo, Françoise Zay, Roberto Paternostro, as well as all the IPFC colleagues and participating students.

Learning from L2 learners to improve pronunciation training

Jacques Koreman, Olaf Husby, Keivan Hedayatfar, Øyvind Bech

Department of Language and Literature, NTNU, Trondheim, Norway
{jacques.koreman; olaf.husby; keivan.hedayatfar; oyvind.bech}@ntnu.no

Keywords: CALST, language typology, generalization, multi-lingual, learning trajectory.

1. INTRODUCTION

The Computer-Assisted Listening and Speaking Tutor (CALST) is a multi-lingual platform for listening and pronunciation training. It both makes predictions about mistakes that learners can be expected to make and learns from actual, observed mistakes: 1) It uses typological databases to perform a contrastive analysis, which is subsequently used to select exercises depending on the learner's native language (L1). 2) For linguistic properties for which no typological data can be used to predict possible mistakes, learner data are logged. Exercises which were unproblematic for previous learners with a given native language are deleted from the learning trajectory of a new user with the same L1. 3) The rationale in 2 is also extended across target languages: If a linguistic property in one target language was unproblematic for speakers of a given L1, it is assumed that the same or a similar property will also be unproblematic for the same speaker group in another target language (L2).

The collection of an L2 learner corpus is thus an integral part of our multi-lingual computer-assisted pronunciation training (CAPT) platform. The data collection includes perception test results, voice recordings and orthographic transcriptions of auditorily presented words. The data can be analysed phonetically and will also be used directly to provide efficient learning trajectories for L2 learners.

2. PREDICTING L2 CHALLENGES

When learning a new language, L1-L2 differences are an important cause of pronunciation challenges [1]. Instead of presenting all learners with the same pronunciation exercises, information about the learner's L1 can thus be used to select L2 pronunciation exercises which can be assumed to be particularly relevant to each specific learner. Doing just that, CALST offers tailored listening, speaking and writing exercises [5]. Based on the phoneme inventories of a large number of languages in the UCLA Phonetic Segment Inventory Database (UPSID) and the Lyon-Albuquerque Phonological Systems Database (LAPSyD), the L1-L2*map* tool for

contrastive analysis is used to determine all unfamiliar sounds in the L2 [3]. Only these sounds are linked to exercises for any given learner. This makes the learning trajectory in CALST more efficient than in CAPT systems where the exercises are fixed.

L2 learners not only have to learn to recognize and pronounce unfamiliar sounds, they also have to learn to pronounce them in unusual positions and unfamiliar clusters [4]. In L1-L2*map* we have provided a list of consonant clusters which can occur in the syllable onset or coda of ten languages, including Norwegian. The exhaustive lists enable us to perform a contrastive analysis and use the result to offer tailored exercises for learners speaking those languages (at present only for Norwegian as a target language). The criteria for choosing the languages were that they belong to diverse language families and are spoken by large immigrant groups in Norway.

Unfortunately, comparable consonant cluster information is not available for other languages in a form which is suitable for contrastive analysis. Although LAPSyD contains information about the maximum syllable templates in a large number of languages, it is not possible to use this information to create tailored exercises for complex syllable clusters. This would require that we can expand the syllable templates into an exhaustive list of syllables of consonant clusters which can be used for contrastive analysis, but this is not possible [4].

Learners with native languages for which no complete list of consonant clusters is available are guided through *all* consonant cluster exercises. In sections 4 and 5, we shall explain how corpus data obtained from these learners can make the learning trajectory in CALST more efficient for other learners.

Other language typological information, such as the information about word stress in the StressTyp2 database can in principle also be used for a contrastive analysis, but there are still many open questions with regard to the effect of differences between L1 and L2 stress systems on language learning [2]. Similar problems in the predictability of learning challenges exist when it comes to lexical tone, where the presence or absence of lexical tones in L1 does not predict problems with the production and perception of lexical tones in L2 [6]. Since it is not immediately clear how typological information can be used in these cases, a more pragmatic approach must be adopted. Such an approach is presented in sections 4

and 5. But first we describe the collection of a phonetic learner corpus in the next section.

3. PHONETIC DATA COLLECTION

The phonetic corpus collection is an integral part of the user logging in the CALST platform.

The user logging helps learners to evaluate their L2 language skills shown in the learner progress statistics (exercise scores presented in charts) and is at the same time used to gather phonetic learner data. The data include the results from several exercise types:

- results from perception tests using minimal pairs (see AXB and minimal pair exercises in CALST)
- voice recordings of single words (pronunciation exercises)
- orthographic transcriptions of single words (spelling exercises)

Firstly, the corpus data can be analysed linguistically to further develop L2 acquisition theory. Secondly, they can be used directly to create more efficient learning trajectories, as is explained in the next two sections. In section 4, we explain how phonetic data obtained from learners with the same L1 can be used to reduce the number of exercises for a given L2, while section 5 describes a parallel application across different target languages.

4. LEARNING FROM OTHER L1 LEARNERS

In cases where language typological information cannot be used to make a selection from all available pronunciation exercises (cf. section 2), logged phonetic learner data collected about actual, *observed* problems for each learner can be used to tailor the learning trajectory.

CALST logs all the results from exercises (cf. section 3) together with user data, esp. the learner's L1. If there is no typological information available which can be used to select a subset of exercises, learners must complete the full set of exercises (cf. section 2). Exercises with close to 100 % correct responses for a set of learners with a shared L1 (and no learners with the same L1 having problems with the exercise) are taken out of the learning trajectory for learners with the same L1, reducing time spent on unnecessary exercises and leaving more time to focus on other properties of the L2 which do need attention.

5. LEARNING ACROSS LANGUAGES

CALST is a truly multilingual CAPT platform, not only because it takes into consideration learner's L1 in generating the learning trajectory, but also because any new language can be implemented on the CALST

platform. By supplying language content, all the exercises in CALST automatically become available for the new target language. CALST is available for Norwegian and is presently being extended to British English.

Any user data logged for a given L2 not only contributes to a more effective learning trajectory for learners with the same L1 who want to learn the same target language (section 4), it can also be used to create an effective learning trajectory in other languages. If learners with a shared L1 do not have problems with a particular sound in a given target language (e.g. Norwegian), we can assume that the same sound is equally unproblematic in another L2. Consequently, the exercise for that sound in a different L2 can also be taken out of the learning trajectory for learners with the given L1.

This is not only true for unproblematic sounds, but also for any other language property. CALST offers exercises for different repair strategies which learners may adopt when confronted with an unfamiliar consonant cluster: reduction, substitution, metathesis or epenthesis. It is not possible to predict on linguistic grounds which repair strategies learners with a given L1 will use, but exercises for consonant clusters which are unproblematic as well as exercises for repair strategies that other learners with the same L1 do not apply can be deleted from the learning trajectory. This helps to reduce the large number of exercises for consonant clusters.

6. SUMMARY AND CONCLUSION

CALST uses large language typological databases to adapt the learning trajectory to the learner's native language. At the same time, there are many phonetic properties for which we cannot predict whether learners will have problems with them. In these cases, CALST generalizes across learners, deleting exercises from the learning trajectory when other learners with the same L1 have high scores for them.

Languages can share phonetic properties. Consequently, it is possible to generalize information from one language to another: If learners with a shared L1 do not have problems with a given sound, sound cluster, stress pattern, etc. in one target language, we can assume that it will also be unproblematic in another language. This generalization strategy makes the CALST concept truly multi-lingual.

CALST is restricted to simple drill exercises using single words or short word sequences. We hope that knowledge obtained from these exercises can be a useful complement to more spontaneous phonetic learner corpora with L2 speech recordings, which are likely to display more variation in learner behaviour.

REFERENCES

- [1] Ellis, R. 1994. *The Study of Second Language Acquisition*. Oxford University Press.
- [2] Kijak, A. 2009. *How stressful is L2 stress? A cross-linguistic study of L2 perception and production of metrical systems*. Utrecht: LOT (Ph.D. thesis).
- [3] Koreman, J., Bech, Ø., Husby, O. & Wik, P., 2011. L1-L2map: a tool for multi-lingual contrastive analysis, *Proc. ICPHS2011*, Hong Kong.
- [4] Koreman, J., Martínez-Paricio, V., Abrahamsen, J.E., Husby, O. 2015. A systematic approach to the pronunciation training of phonotactics, *Proc. ICPHS2015*, Glasgow.
- [5] Koreman, J., Wik, P., Husby, O. & Albertsen, E. 2013. Universal contrastive analysis as a learning principle in CAPT, *Proc. SLaTE 2013*, 172-177.
- [6] van Dommelen, W. & Husby, O. 2009. Native and non-native identification of Norwegian word tones, *Proc. Xth Nordic Prosody Conf.*, Helsinki, 237-247.

Notes:

THE ICE-IPAC PROJECT: TESTING THE PROTOCOL ON NORWEGIAN AND FRENCH LEARNERS OF ENGLISH

Helene N. Andreassen¹, Nadine Herry-Bénit², Takeki Kamiyama², Véronique Lacoste³

1. UiT The Arctic University of Norway
2. Linguistique Empirique: Cognition, Société et Langage (LECSel), EA 1569, Université Paris 8
3. Universität Freiburg

helene.n.andreassen@uit.no, {nadine.herry, takeki.kamiyama}@univ-paris8.fr,
veronique.lacoste@anglistik.uni-freiburg.de

Keywords: interphonology, English as a Foreign Language (EFL), English as a Second Language (ESL), inter- & intra-speaker variation, L1 diversity.

1. INTRODUCTION

Different learner corpora of English exist already, but they typically tend to focus on various aspects of the grammar, vocabulary, and written forms, e.g. the *International Corpus of Learner English, ICLE* [7]; the *Longitudinal Database of Learner English, LONGDALE* [13], see also [12]. There has been less emphasis on interphonology, but see the *Asian English Speech Corpus Project, AESOP* [10] [15], and projects which focus on prosody in L2 English, e.g. the *LeaP Corpus* [8].

English interphonology has hitherto been fairly understudied and, still, many questions merit attention: Which factors influence interlanguage phonology the most? Do interphonological phenomena surface in a similar fashion, and to the same extent, in situations of EFL vs. ESL? In other words, do learners develop the same features, regardless of whether they are EFL learners or ESL learners? If not, how different are they and which extra-linguistic factors are involved? To what extent is such comparison tenable altogether, both theoretically and empirically? What is the role of input in contexts where different varieties of English are spoken?

2. ICE-IPAC AND ITS MOTHER PROJECTS

ICE-IPAC is inspired by two well-established corpus projects, namely IPFC (*Interphonologie du Français Contemporain* [3]), which focuses on learners of French as a foreign language, and PAC (*Phonologie de l'Anglais Contemporain* [2] [5]), which focuses on variation in contemporary L1 English.

The originality of the PAC research project lies in the stability of the research protocol, open access to the sound files and the transcribed data, and a variety of tools and coding conventions adapted for

PAC and/or its sister project PFC (*Phonologie du Français Contemporain* [4]). Taken together, the protocol and the analytic tools offer an initial treatment of the data, as well as inter-speaker and inter-variety comparability, with possibilities of extending the linguistic analysis on several levels.

The originality of the ICE-IPAC research project stems from the fact that it provides an international database of L2 English phonology and phonetics, consisting of learners with different L1s (Spanish, Norwegian, etc.), learners with different L1 varieties (Hexagonal French, Canadian French, etc.), and learners with different L2 varieties as their target/model (and not simply traditional models of English such as RP or American English). The corpus includes different tasks within the same dataset (e.g. word lists and conversation), emphasizing both inter-speaker and intra-speaker variation, all while taking into account the social background of learners. The ICE-IPAC protocol further allows for a comparison between learners and native speakers of English, as it shares some aspects of the PAC protocol. As for the structure of the datasets, the ICE-IPAC protocol provides recording of various speech styles for each speaker, that is 1 repetition task (word lists), 2 reading tasks (word lists and text), and 2 conversations (formal interview and informal conversation). The 2 word lists include: 1) a *common list* applied to all learners in the different corpora; 2) an *L1 specific list* adapted to verify difficulties frequently observed in the production of the learners of a given L1.

The general proficiency level of learners who are invited to participate in the study ranges from A1-B1 to B2-C1 of the *CEFR (Common European Framework of Reference for Languages)*.

3. PHENOMENA TO BE STUDIED

3.1. Segmental features

The common word list includes, among other things, the tense vs. lax vowel contrast, low front vs. back vowels, and the lenis-fortis contrast in word-initial

and word-final stops. As for the last phenomenon, French speakers tend to produce pre-voiced vs. unaspirated voiceless stops; neutralization of word-final voicing is observed in most West Germanic and Slavic languages; no voicing or aspiration contrast for word-final obstruents, or no word-final obstruents at all in many East Asian languages. See Section 4 for Norwegian speakers.

The specific word list includes, among other things, voiced (or lenis) fricatives for the Norwegian learners, and the /s/-/tʃ/, /r/-/w/ and /h/-/Ø/ contrasts for the French learners.

3.2. Word stress-related and suprasegmental features

Some features related to word stress will be analyzed from the two word lists, as follows: vowel reduction in unstressed position (lettER, commA), vowel quality in the stressed position of disyllabic words containing particular digraphs (e.g. *aunt*, *awful*), vowel quality in the stressed position of polysyllabic words containing two distinct vowels (e.g. *continuity*), and stress-imposing endings with a diphthong (*-ize*, *-ify*, *-ate*). As far as suprasegmental features are concerned, the learners' rhythm (i.e. coefficient of variation of consonants and vowels, pauses, speech tempo, speech rate, etc.) and their intonation system (range and slope of F0, etc.) will also be studied.

4. PRELIMINARY RESULTS

During the academic year 2014/2015, a pilot study has been conducted with 2 L1 Norwegian and 2 L1 French learners of English following the PAC protocol. The two female Norwegian learners (NW1 and NW2) are enrolled at UiT The Arctic University of Tromsø, while the two female French learners (FR1 and FR2) are enrolled at the University of Lyon. All four have performed two tasks, i.e. reading of two word lists and an informal conversation between the two.

Some phonological phenomena observed in the L2 English of the Norwegian learners:

1) devoicing of fricatives /z/ /ʒ/ and affricate /dʒ/, not part of the Norwegian sound system. First, /z/ is replaced by [s]: e.g. *seal-zeal* are near-homophonous; *graze* is produced [gɹæis] or even [gɹæs]. Second, /ʒ/ is replaced by [ʃ]: e.g. *leisure* is produced [ˈleɪʃə]. Third, /dʒ/ is replaced by [tʃ]: e.g. *batch-badge* are both produced with [tʃ].

2) partial neutralization of the lenis-fortis contrast in word-final stops. Even though pre-fortis clipping is observed, voicing ceases well before the release of the lenis coda. *fat-fad* are near-homophonous, i.e. [fæt]-[fæt]; [p] is used for /b/, making *lap-lab* near-

homophonous [læp]-[læp]. In some cases, however, the underlying stop contrast is clearly preserved by the target voiceless member being aspirated, e.g. *sack* [sæk^h] vs. *sag* [sæk]; the latter observation is anticipated by the Norwegian L1 system, in which the underlying contrast is primarily expressed through aspiration and not voicing on the edges of words [11].

3) post-vocalic rhoticity (although NW1 is more consistent with it); *carter* [ˈkɑɪtə].

4) production of [v] for phoneme /w/, not part of the Norwegian sound system: *word* [vɔɪd], *which* [vɪtʃ], *worthy* [ˈvɔɪdi]; hypercorrected production of [w] for phoneme /v/ in *vexed* [wekst]; epenthesis of labial [v] or [w] in front of rounded, single onset, [ɹ]: *written* [ˈvɪɪtn], *run* [vɹʌn].

5) interdental fricatives, not part of the Norwegian sound system, replaced by coronal plosives: [t] for /θ/ in *farther* [ˈfɑɪðə] and [d] for /ð/ in *worthy* [ˈvɔɪdi].

Some phonological phenomena observed in the L2 English of the French learners:

1) no systematic difference in the quality of production between /i:/ and /ɪ/, or /ɔ:/ and /ɒ/ [6] [9].

2) <i> is pronounced [aɪ] in *sinner* or *simmer*. This may be the result of an overgeneralization of one of the L2 pronunciation rules for <i> [16].

3) /h/ is deleted in words like *here*, *hurry*, etc. (FR2 is more consistent with it) [1].

4) <ch> in *witch* or *China* is pronounced [ʃ] instead of [tʃ], especially while reading the word list [17].

5) postvocalic rhoticity; *carter* and *garter* are pronounced [kaɪtə] and [gɑɪtə] respectively. [14]

Most interphonological phenomena observed at the segmental level seem to be in correlation with the spelling, which is also visible in the conversation task.

5. CONCLUSION

The phenomena observed in the preliminary recordings will be compared with future recordings of speakers of other languages in foreign and second language contexts.

6. ACKNOWLEDGEMENTS

We would like to thank Audrey Bonfiglioli and Ioana Trifu for their help in the collection of the French learner data.

7. REFERENCES

- [1] Capliez, M. 2011. Typologie des erreurs de production d'anglais des francophones : segments vs. Suprasegments. *Recherche et pratiques pédagogiques en langues de spécialité* [En ligne] 30(3), mis en ligne le 12 avril 2012, consulté le 23 juin 2015. URL : <http://apliut.revues.org/1645> ; DOI : 10.4000/apliut.1645
- [2] Carr, P., Durand, J., Pukli, M. 2004. The PAC project: principles and methods. *La tribune internationale des langues vivantes* 36, 24-35.
- [3] Detey, S., Racine, I., Kawaguchi, Y., Zay, F. in press. Variation among non-native speakers: the InterPhonology of Contemporary French. In: Detey, S., Durand, J., Laks, B., Lyche, C. (eds.), *Varieties of Spoken French: a Source Book*. Oxford: Oxford University Press.
- [4] Durand, J., Laks, B., Lyche, C. 2009. *Phonologie, variation et accents du français*. Paris: Hermès.
- [5] Durand, J., Przewozny, A. 2012. La Phonologie de l'Anglais Contemporain: usages, variétés et structure. Les approches de corpus à orientation phonologico-phonétique pour l'anglais. *Revue Française de Linguistique Appliquée* 17(1), 25-37.
- [6] Ginésy, M. 1995. *Mémento de phonétique anglaise*. Paris : Nathan.
- [7] Granger, S. 1993. The International Corpus of Learner English. *The European English Messenger* 2(1), 34. <http://www.essenglish.org/mess/mestoc21.html>.
- [8] Gut, U. 2009. *Non-native Speech. A corpus-based analysis of the phonetic and phonological properties of L2 English and L2 German*. Frankfurt: Peter Lang.
- [9] Herry-Bénit, N. 2011. *Didactique de la phonétique anglaise*. Rennes : Presses Universitaires de Rennes.
- [10] Kondo, M. 2012. Design and analysis of Asian English speech corpus —How to elicit L1 phonology in L2 English data—. In: Tono, Y., Kawaguchi, Y., Minegishi, M. (eds.), *Developmental and Crosslinguistic Perspectives in Learner Corpus Research*. Amsterdam & Philadelphia: John Benjamins, 251-278.
- [11] Kristoffersen, G. 2000. *The Phonology of Norwegian*. Oxford: Oxford University Press.
- [12] Learner Corpus Association <http://www.learnercorpusassociation.org/>
- [13] Longitudinal Database of Learner English, LONGDALE www.uclouvain.be/en-cecl-longdale.html
- [14] Roach, P. 2000. *English Phonetics and Phonology*. Cambridge: Cambridge University Press.
- [15] Visceglia, T., Tseng, C., Kondo, M., Meng, H., Sagisaka, Y. 2009. Phonetic aspects of content design in AESOP (Asian English Speech Corpus Project). *Proc. 2009 Oriental COCOSA Ürümqi*, 60–65.
- [16] Wells, J. C. 2007. *Longman Pronunciation Dictionary* (3rd edn). London: Longman.
- [17] Wells, J. C. 2014. *Sounds Interesting*. Cambridge: Cambridge University Press.

Notes:

DETERMINING LEARNER PRONUNCIATION MODELS WITH THE K-NN ALGORITHM

Nicolas Ballier

Univ. Paris Sorbonne Cité (Paris Diderot) / CLILLAC-ARP (EA3967)
nicolas.ballier@univ-paris-diderot.fr

Keywords: learner pronunciation models, vowels, French-English interphonology, K-NN algorithm,

1. INTRODUCTION

Most phonetic learner corpora are dictionary-based and transcribe a target / pronunciation model for learners [1], which is usually an American pronunciation model, because the electronic resources for British English are scarce. This paper proposes to investigate which reference pronunciation model learners aim to emulate, by comparing their realisations with two reference formant datasets, which represent pronunciation models for British and American speakers.

Reference varieties (and their putative rejection or adoption in the name of ELF) have been the topic of a heated debate, sometimes laden with political and economical considerations ([9], [12]). The whole debate ([5], [7], [10], [12]) and the economic issues surrounding prestige [9], as well as ideological and political implications have been commented upon, as well as the potential need for a “Lingua Franca Core” [12].

Various methods have been used to determine a variety of English on segmental criteria. Wieling et al. [13] have adopted another methodology to “evaluate the suitability of a computational pronunciation comparison method” and have used ready-made transcriptions from the ACCENT project archives and have shown the Levenshtein distance to be a good metrics for the measure of pronunciation distance, congruent with on-line native judgements.

2. METHOD

Our method is signal-based, analysing F1 and F2 learner performance for variables.

2.1. Reference datasets (training sets)

For our case study, we have focused on the two main competing pronunciation models for French students. French students are advised in the prescriptivist Agrégation reports to avoid a ‘MidAtlantic’ mix of British and American features

(see also Cruttenden [6] discussing ‘Amalgam English’ and ‘International English’).

The reference value for American English used in this experiment were the ones included in the Phontool R package [3]. Data from [8, personal communication] were converted from barks to hertz to ensure comparability.

For the following assumed realisations of the nine vowels of the words “*had*”, “*head*”, “*heard*”, “*heed*”, “*hid*”, “*hod*”, “*hood*”, “*hud*” and “*who’d*”, datapoints were collected, yielding 273 comparable observations from [8] and 1,390 data points for [11].

Vowel formants were measured in the middle position of the vowel, the vowel interval boundaries were automatically determined by automatic pitch tracking of F0. This means that diphthongs were excluded from comparison.

2.2 The test set

The learner corpus is a longitudinal series of interviews of undergraduate students from the university of X. The formant values of the vowels of 13 speakers (3 males, 10 females) whose L1 is French were automatically extracted with a Praat [4] script (see [2] for a more detailed description of the protocol). We selected the male datapoints corresponding to the 9 monophthongs from our database, resulting in 6,354 tokens.

2.3. The k-nn algorithm

The k-nearest neighbour algorithm is a point-to-point classifier that assesses the Euclidian distance between each learner datapoint (for F1 and F2) and a variable number of neighbouring datapoints. For this paper, we have limited the experiment to a comparison between two varieties, but the principle can be extended to multiple classifications (and therefore as many reference pronunciation models as deemed sensible) as well as to other dimensions (F3, F4, vowel duration).

For each phonemic vowel type, F1 and F2 learner tokens were automatically compared to the k neighbouring datapoints established in the two reference studies (for Standard British English and for General American). For each datapoint (a vocalic realisation and its corresponding F1 and F2), the

distance between k nearest neighbours of each instance of the reference varieties (the training datasets, in our case, the result of British and American reference studies) was measured, and the system returned a score establishing whether the considered learner vowel was closer to a British or an American realisation. We have limited the investigation to male speakers (the only subjects studied in [8]) for better data comparability.

3. RESULTS

We used a tenth of the training set to check the consistency of the training data. Phoneme ellipses are notoriously messy and zones of overlaps exist (as evidenced in Figure 4 in [8]). In spite of these encroaching zones for the phoneme ellipsis within the two varieties, the F1/F2 values resulted in consistent results as to discrimination between varieties. With a 10-fold cross-validation, the confusion matrix for the training phase of the 1,663 data points reads as follows: only 7,39% of the tokens were misclassified. To minimise outliers, k is optimal with 12 neighbours.

==== Confusion Matrix ====

a b <-- classified as

174 99 | a = SBE

24 1,366 | b = GenAm

As to the test phase, male learner tokens were mostly classified as Gen Am (see percentages and row data in Table 1).

Table 1: Majority votes for the Gen Am reference variety per vowel per speaker

vowel	speaker1	speaker2	speaker3
had	77%(352)	64%(129)	78%(228)
head	78%(260)	50%(202)	78%(207)
hid	49.87%(794)	45%(655)	45%(655)
heed	49.17%(181)	31%(112)	26%(129)
heard	100 % (6)	61%(18)	0%(2)
hod	78%(360)	72%(201)	79%(295)
hood	37%(322)	25%(173)	17%(348)
who'd	33%(83)	24%(58)	15%(79)
hud	74%(241)	70%(156)	72%(196)

The two sets of problematic categorisations are *hid* vs. *heed* and *who'd* vs. *hood*. F3 values may refine the results for the latter. In two cases, the non-native datasets include CV vowel tokens, which are excluded from the hVd protocol; including CV tokens may have skewed the results, as the hVd reading lists in [8] and [11] preclude the comparison with CV realisational contexts.

We have submitted the recordings to six experts trained in phonetics to estimate the likely

pronunciation model of each speaker to validate the assumptions. The six experts confirmed the attribution of the label (British/American) for the three speakers (k=1).

4. DISCUSSION AND FUTURE RESEARCH

Two main caveats need to be taken into account. Because students may not know some of the hVd words, we did not ask them to read those words, and instead relied on a more limited set of words for read speech in isolation. To increase the data, we relied on unscripted speech, where variation is more likely. The full version of the paper discusses:

- evolution of scores over the three year period
- applicability to female subjects using Vocal Tract Length Normalization (VTLN) techniques.
- the bias-variance trade-off and the optimisation of the hyper-parameter k (number of neighbours considered in the analysis)
- extension to other pronunciation models in multiclass comparisons (using supplementary data from [8]). For L2 speakers under British influence, one could fine-tune the accent detection between the thirteen accents under scrutiny in [8]
- the optimisation of the formant measurements in hertz, log(herz) and barks to better approximate speaker perception, the need to rescale duration variation (in ms) to make it compatible with bark/herz variation
- the strengths and weaknesses of the algorithm for this classification task / conformity metric.

The k-nn algorithm is dependent on the training sets, but more pronunciation models can be learnt by rote, so that finer-grained recognition of other reference accents can be taught. The algorithm can function as multiclass classification, eg judging between British, Singapore or American realisations, acquiring more expertise as formant reference values are fed onto the system as training datasets.

5. ACKNOWLEDGEMENT

Thanks are due to Emmanuel Ferragne for granting access to their Ferragne & Pellegrino 2005 data and to Aurélie Fischer and Marie Candito for exchanges about the k-nn algorithm.

Part of this research was carried out during a research leave granted by the French national accreditation board (CNU), for which grateful thanks are acknowledged.

6. REFERENCES

- [1] Ballier, N. & Martin, Speech annotation of learner corpora. *Handbook of Learner Corpus Research*, CUP 107-134.
- [2] Méli, A. & Ballier, N. Assessing L2 phonemic acquisition : a normalization-independent method. Submitted, Glasgow.
- [3] Barreda, S. 2014. phonTools: Functions for phonetics in R. R package version 0.2-2.0. <http://cran.r-project.org/web/packages/phonTools/phonTools.pdf>
- [4] Boersma, Paul & Weenink, David 2012. Praat: doing phonetics by computer [Computer program]. Version 5.3.19, retrieved 24 June 2012 from <http://www.praat.org/>
- [5] Collins, B. S., & Mees, I. M. 2013. *Practical phonetics and phonology: a resource book for students*. Routledge.
- [6] Cruttenden, A. 2014. *Gimson's Pronunciation of English*, Routledge.
- [7] Dziubalska-Kolaczyk, K., Przedlacka, J. (eds.). 2008. *English pronunciation models: A changing scene* Bern: Peter Lang.
- [8] Ferragne, E., Pellegrino, F. 2010. Formant frequencies of vowels in 13 accents of the British Isles. *JIPA* 40(1), 1–34.
- [9] Fix, E., Hodges, J.L. 1951. Discriminatory analysis, nonparametric discrimination: Consistency properties. *Technical Report 4*, USAF School of Aviation Medicine, Randolph Field, Texas.
- [10] Graddol, D. (2006). *English next*. London: British Council.
- [11] Hillenbrand, J.M., Getty, L.A., Clark, M.J., and Wheeler, K. (1995). "Acoustic characteristics of American English vowels," *Journal of the Acoustical Society of America*, 97, 3099-3111.
- [12] Jenkins, J. 1998. Which pronunciation norms and models for English as an International Language?. *ELT journal*, 52(2), 119-126.
- [13] Wieling, M., Bloem, J. Mignella, K. Timmermeister, M., Nerbonne, J. 2014. Measuring foreign accent strength in English: Validating Levenshtein distance as a measure. *Language Dynamics and Change*, 4(2), 253-269.

Notes:

The influence of intelligibility, comprehensibility and degree of foreign accent in evaluating and categorizing non-native pronunciation errors

Mario Carranza

Universitat Autònoma de Barcelona
Departament de Filologia Espanyola
Campus UAB–Edifici B, 08193, Bellaterra (Barcelona) Spain
mario.carranza@uab.cat

Keywords: non-native accent, pronunciation errors, speech perception, error assessment.

1. OBJECTIVE

Characterizing a foreign realization as mispronounced depends on several factors, such as the degree to which the realization deviates from a standard pronunciation model, the phonological implications of the mispronounced sound(s) and finally, the individual tolerance level of non-nativeness by the interlocutor [1]. On the other hand, recent methodological approaches to pronunciation training do not aim at complete native-like proficiency, but focus instead on increasing the intelligibility and comprehensibility of learners [2]. The objective of our experiment was to test the severity of frequent pronunciation errors by Japanese learners of Spanish as a foreign language. A sample of 50 utterances that contained both words with frequent mispronunciations and words correctly pronounced was drawn from a spoken corpus of Japanese-accented Spanish L2 [3].

2. METHOD

Following the studies of [4], we asked 12 participants, native speakers of Castilian Spanish, to judge the intelligibility, comprehensibility and the degree of foreign accent of each utterance. The participants were not experts –phoneticians or language teachers– and reported no knowledge of Japanese. Stimuli were presented to listeners in two conditions: first, the words were presented in isolation and afterwards the same words were presented within the original utterances. Listeners were asked to rate on the three dimensions after each stimulus was heard.

Five types of pronunciation errors were selected according to their frequency of appearance in the corpus and to their different phonological status. Following the distinction proposed in [5], phonemic, phonetic and phonotactic errors were considered. Each group was formed by 5 instances of the

mispronunciation uttered by different speakers in different words. In each group, 5 instances of the same words but correctly pronounced, or words which presented the same target sound or sequence of sounds, were included as control items. All the stimuli were extracted from the sentences in which they were uttered and grouped into two categories: isolated words and words in context. In total, 100 stimuli were created. The experiment was built into a web page and conducted on-line using the tool described in [6]. Participants were asked to do the test with headphones, in a quiet environment and without taking any pause between the stimuli. A training phase with 6 stimuli was also included.

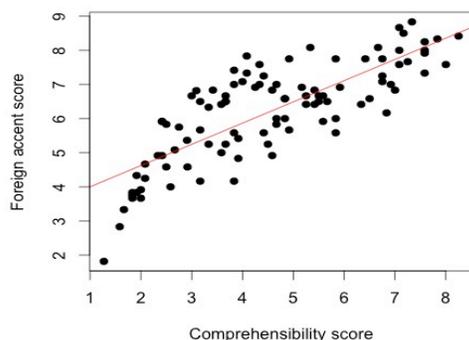
Intelligibility was assessed by asking the participants to write down into a textbox exactly what they perceived; comprehensibility and degree of foreign accent were evaluated using two Likert scales with 9 points (“1” being no accented and clear, and “9” highly accented and difficult to understand). After the first 50 stimuli (isolated words), it followed a pause of 2 minutes and then the second part (words in context) started; the procedure was exactly the same as in part one. Stimuli were randomly ordered and each stimulus could be heard up to three times. Before finishing the experiment, the participants had to fill in a questionnaire regarding the conditions of the experiment. Comprehensibility and foreign accent scores were directly obtained in the experiment. Intelligibility score was calculated afterwards by comparing each target word as transcribed by the participants with the word intended by the foreign speaker. If the target phone matched in both transcriptions, intelligibility was considered “true”; else, it was labelled as “false”, then the ratio of correctly identified words (true) was calculated.

3. RESULTS AND CONCLUSIONS

Comprehensibility and foreign accent scores yield a correlation of 0.65 ($t = 29.63$ $df = 1195$ $p < 0,001$), although comprehensibility ($\bar{X} = 6.16$ $SD = 2.31$) is globally evaluated as more problematic than foreign accent ($\bar{X} = 4.51$ $SD = 2.83$). These data can be

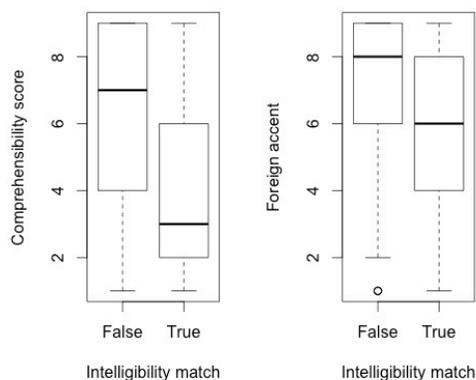
interpreted as follows: stimuli were evaluated with a significantly better score for foreign accent than for comprehensibility; in addition, as the degree of foreign accent increases, the comprehensibility decreases, as expected (Figure 1).

Figure 1: Correlation between foreign accent and comprehensibility



Intelligibility shows a significant relationship with comprehensibility ($F = 277$ $df = 1$ $p < 0.001$) and with foreign accent scores ($F = 88.4$ $df = 1$ $p < 0.001$), correctly identified words were globally evaluated as being less accented and more understandable (Figure 2), intelligibility ratio is higher for the control items (.84) than for the mispronounced items (.59); which suggests that the utterances tagged as mispronounced in the non-native speech corpus from which the stimuli for this experiment were obtained are more difficult to identify than the ones tagged as correctly pronounced.

Figure 2: Comprehensibility and foreign accent score according to intelligibility



As we expected, the presence of the context affected positively the comprehensibility score ($F = 65.08$ $df = 1$ $p < 0.001$) and the intelligibility ratio ($\chi^2 = 103.66$ $df = 1$ $p < 0.001$) but not the foreign accent

score ($F = 0.41$ $df = 1$ $p > 0.05$), which implies that for the same stimuli, the evaluation was consistent between isolated and in context conditions (Figure 3).

Figure 3: Comprehensibility and foreign accent score means according to the type of item

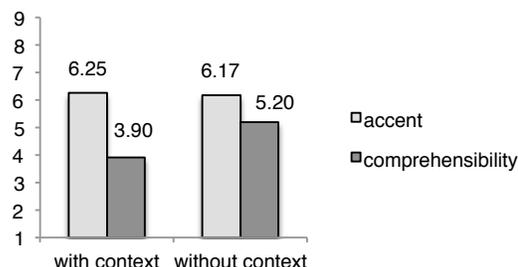


Table 1: Means for mispronounced stimuli (isolated and in-context) sorted by error type

Error type	Foreign accent	Comprehensibility	Intelligibility ratio
c. (syllabic)	7.66	5.81	.50
a. (phonemic)	6.88	5.66	.38
e. (phonetic)	6.44	4.78	.66
b. (phonemic)	6.30	4.50	.76
d. (phonetic)	6.00	4.14	.66
TOTAL	6.66	4.98	.59

We found a significant difference for the comprehensibility ($F = 8.726$ $df = 4$ $p < 0.001$), the foreign accent degree ($F = 12.83$ $df = 4$ $p < 0.001$) and the intelligibility ($\chi^2 = 32.72$ $df = 4$ $p < 0.001$) according to the type of error. Particularly, error type “c” (vowel epenthesis) produced the worst rates for comprehensibility and foreign accent, followed by the confusion in the contrast between [r] and [l] (type “a”, phonemic error). Phonetic errors seem to be the less problematic for native participants, since type “e” (substitution of [x]) and type “d” (substitution of [r]) errors obtained the better rates (Table 1). Vowel epenthesis originates when the learner must produce a combination of sounds that is not allowed by the phonotactic rules of the L1; therefore, the added vocalic element alters the word’s syllabic structure and, consequently, its phonological representation, reducing drastically the cues for identifying the target word. Classical contrastive analyses have not usually considered phonotactic rules, but have focused instead on establishing a mapping between L1 and L2 phonemic categories. In light of the results, we stress the need for reconsidering pronunciation error analysis, taking into account not only phonemic contrasts but also phonetic ones and, especially, dissimilarities at the phonotactic level between the target and the source languages.

4. REFERENCES

- [1] Flege, J. E., Munro, M. J., MacKay, I. R. A. (1995). Factors affecting strength of perceived foreign accent in a second language. *Journal of the Acoustical Society of America*, 97, 3125-3134.
- [2] Derwing, T. M., Munro, M. J. (2009). Putting accent in its place: Rethinking obstacles to communication. *Language Teaching*, 42(4), 476–490.
- [3] Carranza, M., Cucchiari, C., Llisterrí, J., Machuca, M. J., & Ríos, A. (2014). A corpus-based study of Spanish L2 mispronunciations by Japanese speakers. In *Edulearn 2014 Proceedings* (pp. 3696–3705). Barcelona, Spain.
- [4] Munro, M. J., Derwing, T. M. (1995). Foreign accent, comprehensibility, and intelligibility in the speech of second language learners. *Language Learning*, 45(1), 73–97.
- [5] Moulton, W. (1962). Towards a classification of pronunciation errors. *The Modern Language Journal*, 46(3), 101–109.
- [6] Ménétrey, P., Schwab, S. (2014). Labguistic: a web platform to design and run speech perception experiments. In Y. Congosto, M. L. Montero, A. Salvador (Eds.), *Fonética experimental, educación superior e investigación* (Vol. I, pp. 543-556). Madrid: Arco/Libros.

Notes:

BAS RESOURCES FOR PHONETIC LEARNER CORPORA

Christoph Draxler

Bavarian Archive for Speech Signals, Institute of Phonetics and Speech Processing,
Ludwig Maximilian University Munich, Germany
draxler@phonetik.uni-muenchen.de

Keywords: CLARIN, workflow, SpeechRecorder, WebMAUS, percy

1. CLARIN

CLARIN (Common Language Resource Infrastructure) is a European initiative to establish a stable research infrastructure for language and social sciences. In Germany, CLARIN-D consists of nine centres, three of which offer resources and services for speech: Institut für Deutsche Sprache (IDS) in Mannheim, Hamburger Zentrum für Sprachkorpora (HZSK), and Bayerisches Archiv für Sprachsignale (BAS) in Munich. The resources provided by these centres can be used free of charge by academics in Europe.

2. SPEECH DATABASES

A *speech database* is a well-structured collection of digital speech related data on three levels: primary data comprises audio, video, and sensor signal data; this data is immutable (except for format conversion). Secondary data are annotations and derived signal data; they always relates to some signal data, and they can be extended and modified, e.g. by adding annotation levels or correcting existing annotations. Finally, meta-data describe the database contents, ownership, license contracts, processing logs, validation reports, etc.

BAS [1] makes available its speech databases via its CLARIN repository, which may be searched using the virtual language observatory.

2.1. Phonetic learner speech databases

A phonetic learner speech database contains speech signals of non-L1 speakers with at least an orthographic transcription. It may also comprise derived signals such as f0 and formant data, which can be computed automatically, and a phonetic segmentation. This segmentation can be computed automatically, or be performed manually. The speech material in general consists of prompted items which were designed for learners of a language, e.g. minimal pairs, phrases with specific intonation patterns,

exercises, etc.

2.2. Workflow

To establish a phonetic learner speech database, the workflow consists of the specification of the database structure and contents, recording audio data, annotating these recordings, and, finally, exploiting the database for phonetic experiments and statistical analyses (fig. 1).

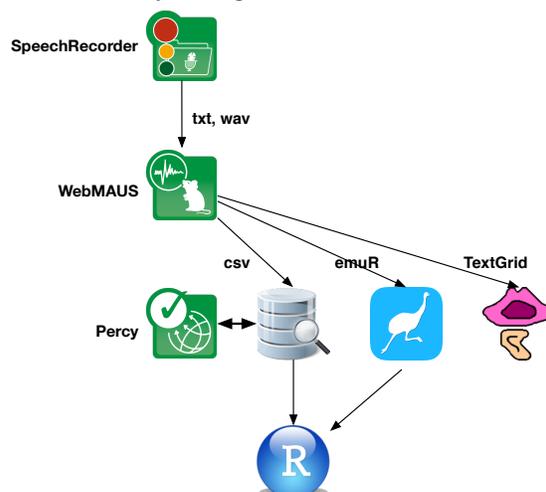


Figure 1: Schematic workflow with BAS and other tools

3. RECORDING: SPEECHRECORDER

SpeechRecorder is a multiplatform application for scripted audio recordings [4]. The contents and structure of a recording session are defined in an XML-formatted recording script. This script is divided into sequential sections, which in turn contain prompt items which may be randomized. Prompt items are either text, images or audio files – the latter two are useful for recording speech without text prompts, e.g. in dialectology, speak-after-me-exercises or with speakers who can't read, e.g. children.

SpeechRecorder supports multiple displays for

different views for speakers and recording supervisors. This facilitates performing recordings in non-standard environments because the speaker interaction can be separated from the controlling device, e.g. in a car, class room, or a classical recording studio.

Each recording session consists of a separate directory on disk. Each utterance is written to a separate file; this audio file may be accompanied with a parallel text file containing the prompt item. The contents of the text file are defined in the recording script to allow a normalized form suitable for further processing, e.g. automatic segmentation.

SpeechRecorder is being ported to mobile devices, and it is embedded into the WikiSpeech system for online recordings [5].

4. ANNOTATION: WEBMAUS

WebMAUS (Munich Automatic Segmentation) is a suite of three tools provided as web services to the community. A web service is accessed either via forms or drag&drop interaction with a browser, via direct procedure calls in a terminal shell, or embedded into other applications. For the user, web services are very attractive because they do not need any software installation, all computing is performed on the server side, and the newest version of the software is being used by default.

The distinguishing feature of WebMAUS is that it not only uses acoustical models for the segmentation, but also applies rules that model the coarticulation patterns of a given language [8]. For example, the standard pronunciation of 'haben' (*have*) denoted in a lexicon would be /h a b @ n/, but in general will be produced as /h a b n/, /h a b m/ or even /h a m/. WebMAUS will attempt to align the signal with these alternatives and return the best match. The overall performance of WebMAUS comes close to that of human labellers, at a fraction of time.

WebMAUS takes as input pairs of audio file and its associated orthographic transcript text file. It currently supports 14 languages plus language-independent SAM-PA, and new languages are added once suitable speech databases become available.

The three WebMAUS services differ in their interface: the basic service simply requires a pair of audio and text file, plus the selection of a language; it returns a Praat TextGrid [2] or an Emu database file [10]. The general service features a large variety of input options, whereas the multiple version offers a drag&drop interface for entire directories of audio and transcript files, and returns an archive file with a segmentation for every input pair of files.

WebMAUS may also called from within the

ELAN annotation editor [9, 6], and it is integrated into an experimental system to provided feedback to L2 learners in language learning applications.

The grapheme-to-phoneme converter G2P, on which WebMAUS is based, is now available as a web service of its own [7]. It takes as input an orthographic transcript and returns not only its canonic pronunciation, but also POS tags, syllabification, morphology and prosodic information.

5. EXPLOITATION: PERCY

In the context of phonetic learner databases, research questions often concern language interference, learning progress, pronunciation and intonation, and many other. These are often tested in perception experiments.

Percy is a framework for media-rich online experiments [3]. It is designed to run on any device supporting a modern browser, including mobile phones, tablets, computers, game consoles, and TV sets. In percy, the experimenter sets up an online experiment by specifying the experiment items, the associated audio files, and the allowed user input. The standard input is simply an empty text field, which is appropriate e.g. for items like 'Type the word you heard', or a series of buttons labelled with input options. Any other interaction element provided by HTML5 may be used, e.g. popup menus, sliders, etc., but this must be programmed by the percy administrator.

Once the experiment has been tested and approved, the experimenter sends a link to potential candidates, who then log in and perform the experiment. Large audiences can be reached very quickly, e.g. via mailing lists, social media, or other. Any user input is immediately saved to the server. This allows a real-time monitoring of progress and access to the data in the running experiment, e.g. for intermediate analyses, testing analysis procedures and scripts, etc. All data is held in a relational database which can be accessed directly from statistics applications such as R or SPSS.

6. SUMMARY

The three tools presented here are available free of charge. They are tailored to the needs of phoneticians and researchers working with audio data. To improve our tools we apply them in our daily work, integrate them into our curricula, and encourage users to provide feedback. In particular, web services will change the way researchers work, because they offer increased functionality but at the same time reduce or even eliminate the need for software installation and maintenance.

7. REFERENCES

- [1] BAS CLARIN-D online repository. <http://hdl.handle.net/11858/00-1779-0000-000C-DAAF-B>.
- [2] Boersma, P. 2001. Praat, a system for doing phonetics by computer. *Glott International* 5(9/10), 341–345.
- [3] Draxler, C. 2011. Percy – an HTML5 framework for media rich web experiments on mobile devices. *Proc. Interspeech* Florence, Italy. 3339–3340.
- [4] Draxler, C., Jansch, K. 2004. SpeechRecorder – a universal platform independent multi-channel audio recording software. *Proc. LREC* Lisbon, Portugal. 559–562.
- [5] Draxler, C., Jansch, K. 2008. Wikispeech – a content management system for speech databases. *Proc. Interspeech* Brisbane. 1646–1649.
- [6] Kisler, T., Schiel, F., Sloetjes, H. 2012. Signal processing via web services: the use case WebMAUS. *Proc. of Digital Humanities* Hamburg. pp. 30–34.
- [7] Reichel, U. D., Kisler, T. 2014. Language-independent grapheme-phoneme conversion and word stress assignment as a web service. In: Hoffmann, R., (ed), *Elektronische Sprachverarbeitung 2014* volume 71 of *Studientexte zur Sprachkommunikation*. Dresden, Germany: TUDpress 42–49.
- [8] Schiel, F. 1999. Automatic phonetic transcription of non-prompted speech. *Proc. ICPhS* San Francisco. 607–610.
- [9] Sloetjes, H., Russel, A., Klassmann, A. 2007. ELAN: a free and open-source multimedia annotation tool. *Proc. Interspeech* Antwerp. 4015–4016.
- [10] Winkelmann, R., Raess, G. 2014. Introducing a web application for labeling, visualizing speech and correcting derived speech signals. Chair), N. C. C., Choukri, K., Declerck, T., Loftsson, H., Maegaard, B., Mariani, J., Moreno, A., Odijk, J., Piperidis, S., (eds), *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC'14)* Reykjavik, Iceland. European Language Resources Association (ELRA).

Notes:

ONE CORPUS, ONE RESEARCH QUESTION, THREE METHODS - GERMAN VOWELS PRODUCED BY FRENCH SPEAKERS

Frank Zimmerer^a, Jürgen Trouvain^a, Anne Bonneau^b

^aComputational Linguistics & Phonetics, Saarland University, Saarbrücken, Germany ^bSpeech Group, LORIA Inria, Université de Lorraine, CNRS, Villers-lès-Nancy, F-54600, France
zimmererltrouvain@coli.uni-saarland.de; anne.bonneau@loria.fr

Keywords: German, French learners, vowels, perception, production.

1. INTRODUCTION

Learning a foreign language (L2) after puberty is not an easy task. Especially interference from the phonological and phonetic system of the native language (L1) is one of the main reasons for this difficulty.

One aspect that has received considerable amount of attention is the acquisition of vowels in L2 (e.g., among many others, [5, 9, 13, 14]). Computer assisted language learning is one way to train and improve perception and production of various L2. For successful implementation of software solutions, phonetic learner corpora are essential.

Such a phonetic learner corpus has been created in the course of the project IFCASL (“Individualized Feedback for Computer-Assisted Language Learning”, <http://www.ifcasl.org>) [4, 12] which investigates the problems that arise in the French-German language pair. This corpus, for instance, allows for the investigation of the interference French speakers have with German vowels. The data of the corpus can be analyzed with different methods, which focus on different aspects of this problem, and which ultimately contribute different results which create a more detailed analysis of the processes in question.

The German and French vowel systems are an interesting possibility to study L1-L2 interferences for several reasons. The German vowel system consists of 16 monophthongs with long and tense as well as short and lax vowels /i:, e:, ε:, a:, o:, u:, y:, ø:, ɪ, ε, a, ɔ, ʊ, ʏ, œ, ə/. On the other hand, the French vowel system exhibits similar properties, but also important differences in comparison to the German vowel system. These differences occur on the phonological and phonetic level [3]. The French vowel system uses 11 oral monophthongs: /i, e, ε, a, y, ø, œ, u, o, ɔ, ə/, see e.g. [10]. This means that the two systems are quite similar with respect to contrasts in vowel height and roundness, although there seem

to occur small acoustic differences between the two languages (e.g. [11]). However, crucially, French does not contrast long and short vowels, but German does [10, 3].

In specialized teaching materials the problem of vowel length/tenseness has been recognized (see e.g. [6, 7]). However, it is still unclear how frequent vowel errors occur among the learners and which pairs of long and short vowels create the most serious difficulties for learners and also native listeners. We shortly present three methods to analyze the interference processes occurring in the L2-productions of French learners of German. We also show the different results which are all based on the IFCASL learner corpus to illustrate that the interference has several aspects which can be investigated with different methodological approaches.

2. METHODS AND RESULTS

All analyses have been carried out using parts of the IFCASL learner corpus. The corpus consists of recordings of read speech from French learners of German and German learners of French in both their respective L1 as well as their respective L2.

2.1. Automatically Extracted Confusion Matrices

A first investigation of the vowel productions of French speakers is an automatically generated list of production alternatives, based on the annotations that are part of the corpus [4, 8, 12]. Results indicate the two vowels that were the most difficult were /œ/ which was correctly produced in 57% of the cases, and as [ø or ø:] in 24% of the cases, and /ʏ/ which was realized correctly 63% of the time ([y,y:] in 24% of the cases). The short /a/ was the vowel that was labelled as being produced correctly most often, that is, in 91% of the cases, its long counterpart /a:/ was labeled as being correctly produced in 83% [a] in 16% (see [8]). This study gives an overall impression which segments were produced correctly, and we can see on an overall level, which vowels seem

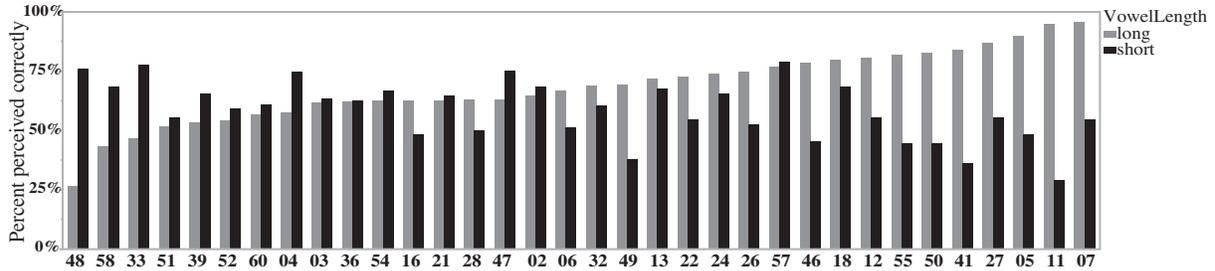


Figure 1: Percent of correctly perceived vowels depending on the speaker and vowel length for the BEG group.

to be harder to produce correctly for French learners of German, and which vowels seem to be easier.

2.2. Perception Experiment

The second analysis we present focuses on the perceptual side of the French learners' productions [14]. While for the first investigation, trained phonetics students provided the basis for the analysis, in this case, listeners decided what word they heard. In the German part of the corpus, 11 (near) minimal word pairs with the vowel as the crucial difference were recorded (e.g. /i:/ vs. /ɪ/: *Miete* 'rent', *Mitte* 'mid'). All of these words were excised from their sentence context. These words were played in isolation and 11 native German listeners indicated in a forced choice perception experiment, which member of the minimal pair they heard. For the experiment, 1157 items were played to the listeners. Results indicate that long vowels were identified correctly 76.6% of the time, whereas short vowels were perceived correctly in 63.9% of the cases, a difference that proved to be significant. We also found that advanced learners (ADV) were perceived more often correct compared to beginners (BEG). BEG were correctly perceived as long in 69.5% of the cases, and as short in 58.4% (see Figure 1 for the BEG), whereas ADV learners were perceived correctly 86.9% for long vowels, and 72.2% for the short ones. Generally, speakers seemed to have more problems with round vowels than with un-round vowels. Furthermore, individual differences were found, some speakers were better in producing long vowels compared to short vowels, but others showed the opposite tendency. Some were equally correct in both cases, others were equally incorrect for long and short vowels.

2.3. Acoustic Measurements

The third method that is presented here used the results of the perception experiment as point of depar-

ture. Based on the individual patterns (see Figure 1 for the BEG), we chose six speakers varying with respect to the ratings in the perception task (having short and long vowels perceived well, with a bias for either short or long vowels, or being perceived bad for both categories, e.g., speaker 51 in Figure 1. All segments of the words were hand-labelled and analysed with PRAAT [1]. Duration, as well as formant values at vowel mid-points were analyzed. Whereas the number of speakers is too small for statistical evaluations, the patterns that emerged were quite striking. Most of the speakers did not differentiate the vowel categories correctly, with a great overlap regarding their duration. Furthermore, they also showed some problems to produce acceptable formant values. Thus, the individual differences can be used for the creation of an individualized vowel training software that focuses on duration as well as formant values of vowel productions [2].

3. SUMMARY

In this article, we presented three methods that investigated the vowel production of French learners of German. All these methods used the same data base. Depending on the method the results shed light on slightly different aspects of the same process, the interference of the French phonetic and phonological system on the production of the German L2 vowels. Whereas the first method revealed that especially rounded vowels are problematic in the long/short distinction, we could show with the second method, that particularly [o: - ɔ] seem to be hard to produce for French learners. The third method corroborated this finding and added acoustic details on duration and formants. The results of the studies can be used to create individualized training and feedback for foreign language learners, aimed at reducing their accent in L2.

4. REFERENCES

- [1] Boersma, P., Weenink, D. 2013. *PRAAT - Doing phonetics by computer*.
- [2] Carroll, P., Trouvain, J., Zimmerer, F. 2015. A visual feedback tool for German vowel production. *Proc. 26. Konferenz Elektronische Sprachsignalverarbeitung 2015 (ESSV)*. Eichstätt, Germany.
- [3] Delattre, P. 1964. Comparing the vocalic features of English, German, Spanish and French. *IRAL - International Review of Applied Linguistics in Language Teaching* 2, 71–97.
- [4] Fauth, C., Bonneau, A., Zimmerer, F., Trouvain, J., Andreeva, B., Colotte, V., Fohr, D., Jouvét, 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. Conference on Language Resources and Evaluation (LREC) 2014* Reykjavik, Iceland. European Language Resources Association (ELRA) 1477–1482.
- [5] Flege, J. E., MacKay, I. R. A. 2004. Perceiving vowels in a second language. *Studies in Second Language Acquisition* 26, 1–34.
- [6] Hirschfeld, U. 2014. *Bewegte Phonetik. Ausspracheübungen mit französischen Schülern im Deutsch- und Englischunterricht (classes bilangues)*. Paris: Goethe Institut (<http://www.goethe.de/ins/fr/lp/prj/clb/unt/pho/deindex.htm>).
- [7] Hirschfeld, U., Reinke, K., Stock, E., (eds) 2007. *Phonothek intensiv. Aussprachetraining*. München: Langenscheidt.
- [8] Jouvét, D., Bonneau, J., Anneand Trouvain, Zimmerer, F., Laprie, Y., Möbius, B. 2015. Analysis of phone confusion matrices in a manually annotated French-German learner corpus. *Proceedings of SLaTE*.
- [9] Kingston, J. 2003. Learning foreign vowels. *Language and Speech* 46(2-3), 295–349.
- [10] Pustka, E. 2011. *Einführung in die Phonetik und Phonologie des Französischen*. Berlin: Erich Schmidt Verlag.
- [11] Strange, W., Weber, A., Levy, E. S., Shafiro, V., Hisagi, M., Nishi, K. 2007. Acoustic variability within and across German, French, and American English vowels: phonetic context effects. *The Journal of the Acoustical Society of America* 122, 1111–1129.
- [12] Trouvain, J., Laprie, Y., Möbius, B., Andreeva, B., Bonneau, A., Colotte, V., Fauth, C., Fohr, D., Jouvét, D., Mella, O., Jügler, J., Zimmerer, F. 2013. Designing a bilingual speech corpus for French and German language learners. *Proc. Corpus et Outils en Linguistique, Langues et Parole: Statuts, Usages et Méusages* Strasbourg. 32–34.
- [13] Weber, A., Betta, A. M. D., McQueen, J. M. 2014. Treack or trit: Adaptation to genuine and arbitrary foreign accents by monolingual and bilingual listeners. *Journal of Phonetics* 46, 34–51.
- [14] Zimmerer, F., Trouvain, J. 2015. Perception of French speakers' German vowels. *Proceedings of Interspeech 2015*.

Notes:

ERROR ANNOTATION IN SPOKEN LEARNER CORPORA

Anke Lüdeling, Simon Sauer, Malte Belz, Christine Mooshammer

Humboldt-Universität zu Berlin

corresponding author: anke.luedeling@rz.hu-berlin.de

Keywords: multi-layer annotation, error annotation, spoken learner corpus, BeMaTaC

1. INTRODUCTION

In order to find learner-specific linguistic properties, patterns in learner corpora are often analyzed quantitatively and compared to patterns in other learner corpora or native speaker corpora (see e.g. [11, 12, 23, 13]). Some patterns can be found on the surface of the learner text using word forms or properties of the sound signal but many research questions require the analysis of more abstract patterns involving, for example, phonemes, tones, lemmas, parts of speech, syntactic phrases, or error categories. This paper is largely methodological and focuses on the question of how error annotation can be done consistently and transparently in a spoken learner corpus. We will illustrate our points with data from the *Berlin Map Task Corpus* (BeMaTaC v. 2013-02.1, [22], see Section 2).

There are at present not many spoken learner corpora. Only some of the existing spoken learner corpora contain sound files, and only some of these are time-aligned and stored in a multi-layer corpus architecture where different annotation layers can be added freely (see [2] for a typology of spoken learner corpora and a discussion of these issues). Some spoken learner corpora are produced for phonetic questions and annotated and analyzed in a tool dedicated for phonetic phenomena. Other spoken learner corpora are collected for lexical, syntactic, or communicative purposes and these often do not contain the signal. Annotation is done using tools dedicated for token, or span annotation, syntactic annotation, or sometimes pointing relations. We will argue below that many properties of learner language and learner speech can only be understood through the combination of information on many layers. This implies a corpus architecture that allows annotation through different tools that are then merged into a common corpus.

In Section 3 we argue that error identification implies the implicit or explicit formulation of a target hypothesis, and that there can be different target hypotheses for the same text depending on the research

question and desired granularity. Since error annotation can pertain to phonetic phenomena in learner speech as well as to grammatical or even communicative properties of learner language - and all of these concurrently in the same corpus - we use a corpus architecture which allows for the alignment of the signal to a transcript, multiple tokenizations and as many annotation layers as necessary [15, 21].

2. BEMATAC

The Berlin Map Task Corpus (BeMaTaC; <https://u.hu-berlin.de/bematac>) is a freely available corpus of spoken German. It consists of an L1 subcorpus recorded with native speakers of German and an identically designed L2 subcorpus with advanced speakers of German as a foreign language (to date, all learners in the corpus are native speakers of English and have test scores equivalent to ECFR level C1 or above). BeMaTaC uses a map-task design, where one speaker (the instructor) instructs another speaker (the instructee) to reproduce a route on a map with landmarks [1]. The corpus is accessible via ANNIS [15], an open-source browser-based search and visualization tool.

3. ERROR ANNOTATION AND SPOKEN LEARNER CORPORA

3.1. Error identification

Error annotation is a difficult task (see [16] for a more thorough discussion). The main reason for this is conceptual: It is not always clear what constitutes an error. This has been discussed extensively in the literature on second language acquisition and foreign language teaching, and there are many suggestions for a definition of ‘error’, some involving purely grammatical criteria, others focusing more on the adequacy of an utterance in a given context, the comparison of what a learner does with what a native speaker would do in a given situation, etc. In essence, however, there can be no general definition of error, and the decision of what constitutes an error depends on the research goal (see, among many others, [6, 5, 9, 10, 7, 19]).

The first step in error annotation is error identification, i. e. a decision on the exponent of the error. Even if the research goal is clear and a precise error definition can be derived from it, it is often unclear how to interpret a learner utterance. Each error is a difference between the utterance and an explicit or implicit ‘correct’ utterance. This is sometimes called **target hypothesis**. Here we define ‘error’ as the difference between the learner utterance and a target hypothesis. There can be many target hypotheses for a given learner utterance. A target hypothesis does not constitute the ‘truth’ or the ‘only correct way of saying something’ but is an interpretation of the utterance for the purpose of a given research goal [17].

We want to illustrate this using a purely grammatical notion of error and two examples from a written learner corpus containing texts from advanced learners of German as a foreign language (the Falko corpus, [18]). (1) contains a number mismatch between an adjective and the noun it modifies. This can be ‘corrected’ in several ways: the number of the adjective can be changed, the number of the noun can be changed, or the noun phrase can be labeled as a whole. Each of the error marking strategies can be defended. In an error analysis the different strategies would lead to different error counts on adjectives, or nouns. The verb *erlernen* in (2) does not subcategorize for a reflexive and, while being possible, it is not the ideal verb here. One could ‘correct’ this sentence in several ways, and again the target hypothesis will influence the error analysis that follows: delete the reflexive (\rightarrow argument structure error), change the verb (for example to *aneignen* “to acquire” \rightarrow lexical/stylistic error), or do both (*neues Wissen zu erwerben* \rightarrow argument structure error and lexical error).

- (1) Um die richtige Strategien in diesen
in-order-to the right.SG strategies.PL in these
Bereichen wählen zu können
areas to-choose to be-able
‘In order to be able to choose the right strategie(s) in these areas’
- (2) bevor man überhaupt anfangen kann, sich neues
before one even start can, REFL new
Wissen zu erlernen
knowledge to learn
‘before one can even start to acquire new knowledge’

The consequence of these issues is that it is necessary to construct an **explicit** target hypothesis (or several) according to transparent criteria (see [20] for a description of several target hypothesis pertaining to different research questions in the Falko corpus). It is equally necessary to construct a target hypothesis following the same criteria for each corpus the

learner corpus is compared to. A shared baseline is essential, as native speakers do not always behave in a way grammar would predict. Constructing target hypotheses is difficult even for fairly advanced, written learner language. It becomes more difficult for varieties that are further away from a ‘standard’.

3.2. Schwa elision

We argued that the comparison of patterns in learner corpora and native speaker corpora across several annotation layers leads to interesting acquisition results. We want to briefly illustrate our point by looking at final schwa in German. In spontaneous German speech, schwa elision occurs quite frequently in word-final position (this is a reduced account; we are aware of the fact that schwa/non-schwa is not a binary decision and that many phonetic parameters have to be taken into account; for a thorough study see e.g. [14]). In BeMaTaC, we can find instances of final schwa elision through a comparison between the diplomatic (narrow transcription) and the normalized transcription, cf. Table 1.

Table 1: Example of the multi-layer architecture in BeMaTaC.

dipl	das	hab	ich	nicht	gesagt
norm	das	habe	ich	nicht	gesagt
pos	PDS	VAFIN	PPER	PTKNEG	VVPP
gloss	that	have	I	not	said

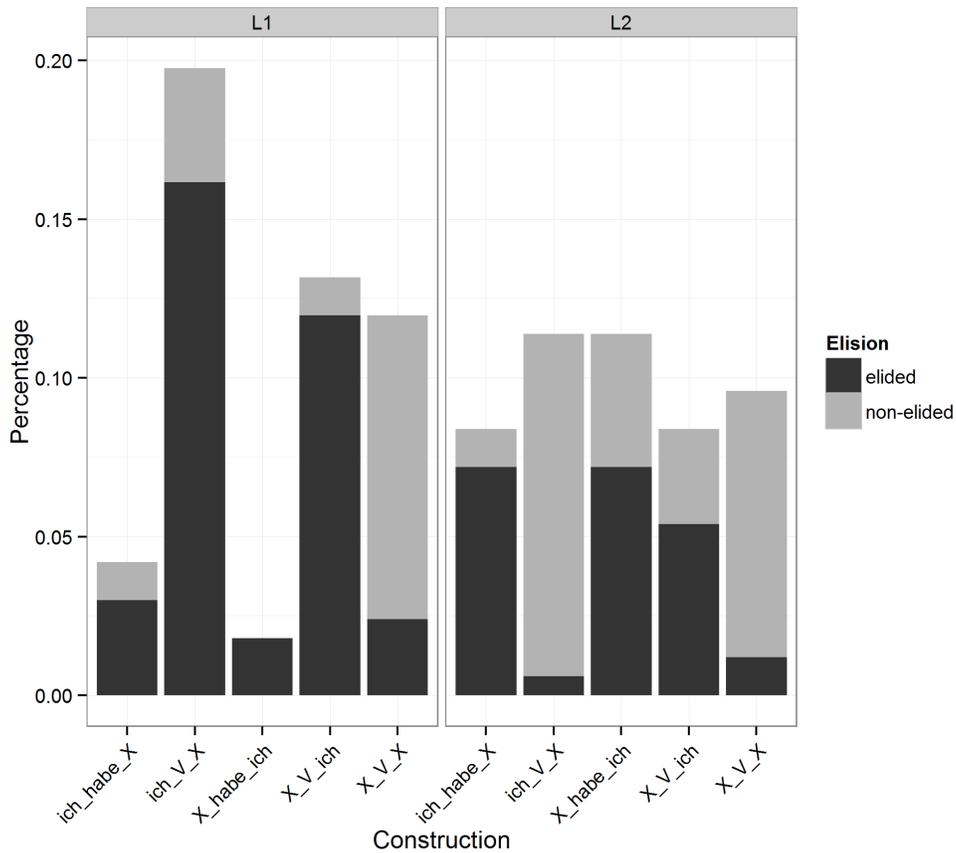
Here, the normalization can be used as a target hypothesis. We are able to integrate information from different linguistic annotations, such as part-of-speech tags or lemmatization. A first analysis reveals that both learners and native speakers do not elide schwas in nouns. Schwas in verbs, however, behave differently (imperatives are excluded from our analysis, as they are paradigmatically schwa-less).

Table 2: Frequencies of schwa elision in BeMaTaC.

	\emptyset -forms (dipl)	-e-forms (norm)	%
L1	67	108	62
L2	44	107	41

The interpretation of the patterns depends on the research agenda. When the normalized layer with unelided forms is seen as a target hypothesis, native speakers produce more ‘errors’ than learners (cf. Table 2). However, this target hypothesis reflects a conceptually written standard. When adhering to a setting of spontaneous speech, we

Figure 1: 3-gram constructions of finite verb forms ending with *-e* in BeMaTaC. V stands for all verbs other than *haben* ‘have’, X stands for any element except *ich* ‘I’



may conclude that learners have not yet achieved the level of schwa elision that is typical of native speakers. A more detailed comparison reveals interesting patterns of 3-gram constructions (cf. Figure 1). The most prominent difference is that L1 speakers use schwa elisions more productively, with a wider range of verbs (12 hapax legomena, e.g., *beschreib* ‘explain’ or *find* ‘find’). L2 speakers, on the other hand, predominantly elide schwas in the construction *ich_habe_X/X_habe_ich* ‘I_have_X/X_have_I’, which may serve as a teddybear construction [8] in the acquisition of verb-final schwa elision.

Spontaneous speech deviates from written language in many other ways. The comparison between a diplomatic layer and a normalized layer in usage data allows us to find these instances and compare overuse and underuse between native speakers and learners.

4. CONCLUSION

We have shown that the definition of ‘error’ depends on the underlying concept of a target hypothesis. The target hypothesis must be defined according to the research question. Therefore, multiple target hypotheses can be applied. Target hypotheses can include or even combine various linguistic domains, such as phonetics, morphology and syntax. This is only possible using a multi-layer corpus architecture.

5. REFERENCES

- [1] Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S., Weinert, R. 1991. The HCRC Map Task Corpus. *Language and Speech* 34, 351–366.
- [2] Ballier, N., Martin, P. 2015. Speech annotation of learner corpora. In: Granger, S., Gilquin, G., Meunier, F., (eds), *The Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press.
- [3] Belz, M. 2013. Disfluencies und Reparaturen bei Muttersprachlern und Lernern – eine kontrastive Analyse. Master’s thesis Humboldt-Universität zu Berlin.
- [4] Belz, M., Sauer, S., Lüdeling, A., Mooshammer, C. 2015. Repair behaviour of advanced German learners in the Berlin Map Task Corpus. Trouvain, J., Zimmerer, Frank, Gásy, Mária, , Bonneau, A., (eds), *IFCASL Workshop on Phonetic Learner Corpora*.
- [5] Bley-Vroman, R. 1983. The comparative fallacy in interlanguage studies: The case of systematicity. *Language Learning* 33(1), 1–17.
- [6] Corder, S. P. 1981. *Error Analysis and Interlanguage*. Oxford: Oxford University Press 4. impr. edition.
- [7] Díaz-Negrillo, A., Fernández-Domínguez, J. 2006. Error tagging systems for learner corpora. *Revista española de lingüística aplicada / RESLA* 19(19), 83–102.
- [8] Ellis, N. C. 2012. Formulaic language and second language acquisition: Zipf and the phrasal teddy bear. *Annual Review of Applied Linguistics* 32, 17–44.
- [9] Ellis, R. 1994. *The Study of Second Language Acquisition*. Oxford University Press, USA.
- [10] Ellis, R., Barkhuizen, G. 2009. *Analysing Learner Language*. Oxford applied linguistics. Oxford: Oxford Univ. Press [nachdr.] edition.
- [11] Granger, S. 1996. From CA to CIA and back: An integrated approach to computerized bilingual and learner corpora. In: Aijmer, K., (ed), *Languages in Contrast: Papers from a Symposium on Text-based Cross-linguistic Studies Lund 4 - 5 March 1994* volume 88 of *Lund studies in English*. Lund: Lund University Press [u.a.] 37–51.
- [12] Granger, S. 2002. A bird’s-eye view of learner corpus research. In: Granger, S., (ed), *Computer Learner Corpora, Second Language Acquisition, and Foreign Language Teaching* volume 6 of *Language learning and language teaching*. Amsterdam: Benjamins 3–33.
- [13] Granger, S. 2008. Learner corpora. In: Lüdeling, A., Kytö, M., (eds), *Corpus Linguistics. An International Handbook* volume 1. Berlin: Mouton de Gruyter 259–275.
- [14] Kohler, K. J., Rodgers, J. 2001. Schwa deletion in German read and spontaneous speech. *AIPUK* 35, 97–123.
- [15] Krause, T., Zeldes, A. 2014. Annis3: A new architecture for generic corpus query and visualization. *Digital Scholarship in the Humanities*.
- [16] Lüdeling, A., Hirschmann, H. 2015. Error annotation. In: Granger, S., Gilquin, G., Meunier, F., (eds), *The Cambridge Handbook of Learner Corpus Research*. Cambridge: Cambridge University Press.
- [17] Lüdeling, A. 2008. Mehrdeutigkeiten und Kategorisierung: Probleme bei der Annotation von Lernerkorpora. In: Walter, M., Grommes, P., (eds), *Fortgeschrittene Lernervarietäten*. Tübingen: Niemeyer 119–140.
- [18] Lüdeling, A., Doolittle, S., Hirschmann, H., Schmidt, K., Walter, M. 2008. Das Lernerkorpus Falko. *Deutsch als Fremdsprache* 45(2), 67–73.
- [19] Ragheb, M., Dickinson, M. 2011. Avoiding the comparative fallacy in the annotation of learner corpora. Granena, G., Koeth, J., Lee-Ellis, S., Lukyanchenko, A., Botana, G. P., Rhoades, E., (eds), *Selected Proceedings of the 2010 Second Language Research Forum: Reconsidering SLA Research, Dimensions, and Directions* Somerville, MA. Cascadilla Proceedings Project 114–124.
- [20] Reznicek, M., Lüdeling, A., Hirschmann, H. 2013. Competing target hypotheses in the falko corpus: A flexible multi-layer corpus architecture. In: Díaz-Negrillo, A., Ballier, N., Thompson, P., (eds), *Automatic Treatment and Analysis of Learner Corpus Data*. Amsterdam: John Benjamins 101–124.
- [21] Sauer, S., Lüdeling, A. to appear. Flexible multi-layer spoken dialogue corpora. *International Journal of Corpus Linguistics. Special Issue on Spoken Corpora*.
- [22] Sauer, S., Rasskazova, O. 2014. BeMaTaC – eine digitale multimodale Ressource für Sprach- und Dialogforschung. Workshop Grenzen überschreiten – Digitale Geisteswissenschaft heute und morgen. Berlin 2014. Poster. Digital Humanities Conference, Berlin.
- [23] Tono, Y. 2004. Multiple comparisons of IL, L1 and TL corpora: The case of L2 acquisition of verb sub-categorization patterns by Japanese learners of English. In: Aston, G., Bernardini, S., Stewart, D., (eds), *Corpora and Language Learners* volume 17 of *Studies in corpus linguistics*. Amsterdam and Philadelphia: John Benjamins 45–66.

L1 PHONETIC DRIFT IN DUTCH L2 SPEAKERS OF ENGLISH

Rosemary Orr ¹, David van Leeuwen², Jacky Zoë de Rode ¹, Georg Lohfink ³ and Hugo Quené ³

¹University College Utrecht, Utrecht University, The Netherlands

²CLS/CLSTRadboud University, Nijmegen, The Netherlands

³Utrecht Institute of Linguistics OTS, Utrecht University, The Netherlands

r.orr@uu.nl

Keywords: L1 attrition, phonetic drift, speech learning model (SLM), assimilation, dissimilation

1. INTRODUCTION

This longitudinal study examines the development of voice onset time (VOT) for the stops /t/ and /d/ and the central frequency of the /s/ sibilant in a group of fluent L2 speakers of English whose L1 is Dutch, over a period of 3 years.

The speakers are Dutch students who live and study at a residential college where the *lingua franca* is English. The student community is international, with ca. 60% being Dutch L1 speakers, and the remaining 40% comprising over 30 different L1s. A detailed description of the making of this corpus can be found in [5].

The core hypothesis which motivated the collection of this corpus is that, over the three years of these students' undergraduate study, their accents of English will gradually converge to some kind of *UCU Accent of English*, which is not a native accent for any speaker. As part of this convergence, the influence of the Dutch accent of English is expected to play a large role, since the Dutch population forms the majority.

For this study, we look at possible phonetic drift in word-initial /d/ and /t/, and the sibilant /s/, which are realised differently in Dutch and English. In non-clustered word-initial position, typical VOT values for Dutch voiceless stop /t/ and the English voiced stop /d/ are quite similar. Dutch voiced stops have a shorter lag time and English voiceless stops have a much longer lag time, being generally aspirated. Dutch has only one sibilant /s/ whereas English has two, namely /s/ and /ʃ/ [3, 2]. The articulation of the Dutch /s/ is described as being somewhere between the two English sibilants, having a more retracted position of articulation, a flatter tongue, and more lip rounding [3].

Because these particular phonemes exhibit *phonetic*, rather than *phonemic* differences in Dutch and English, it is interesting to explore them in the context of Flege's speech learning model (SLM), which suggests that the ability to perceive within-phoneme differences between an L1 and an L2 may drive the

formation of a new phonetic category within a single phoneme [7]. Conversely, if a speaker does not perceive the difference, this new category may not be formed at all, but both L1 and L2 values will assimilate towards each other. We hypothesize that this will be the case for the Dutch L1 speakers' VOT values for /t/ and /d/, and for the production of /s/.

2. METHODS

The first and last recordings for 50 Dutch L1 speakers in the first two cohorts of the UCU Accent Project speech corpus are examined.

2.1. Speakers

The speakers were 50 undergraduate students. Their age range for the first recording was 17 to 19, and for the second recording, three years later, 19 to 21. 10 were male and 30 female. All had started learning English before the age of 8.

2.2. Recording equipment

Recordings were made in a quiet furnished office, using a close-talking microphone (Sennheiser Headset HSP 2ew), via a Saffire Pro 40 multichannel AD converter and preamplifier, using *Audacity*, open source software for recording and editing sounds (see <http://audacity.sourceforge.net/>).

2.3. Materials

Speech material was taken from recordings 1 and 5, so at the beginning of the first semester and at the end of the last, 3 years later. We extracted 2-minute informal monologues in Dutch and English and located the word-initial instances of /d/ and /t/, as well as instances of /s/ are located and used in analysis.

2.4. Parameters

VOT for /d/ and /t/ were measured manually, using Praat [1]. VOT was measured from the release of the stop burst to the onset of voicing.

/s/ was located in a semi-automatic fashion. Candidate /s/ segments were generated with the *Kaldi*

speech recognition system [6], trained on *wsj0* and *wsj1* using the *s5* recipe and an *nnet2-online* neural net configuration. Values ranging from English /s/ to /ʃ/ were generated obtain candidates for both Dutch and English /s/. Candidate segments were listened to and accepted or rejected, one by one. The COG was calculated for all of the accepted candidates.

3. ANALYSIS AND RESULTS

3.1. Voice Onset Time (VOT), for /d/ and /t/

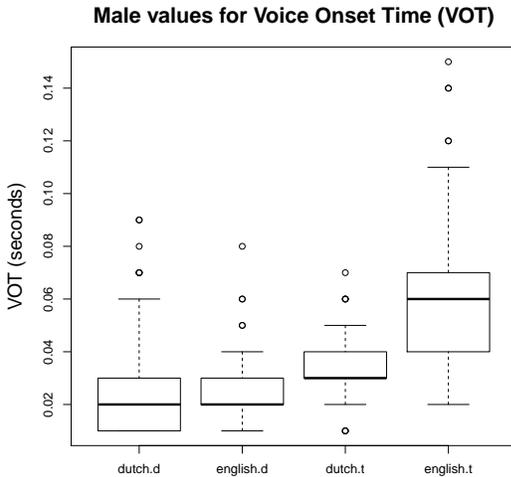


Figure 1: Values for VOT for /d/ and /t/ for English and Dutch for male speakers

Figure 1 shows the VOT measurements for Dutch and English for the male speakers for both /d/ and /t/ segments. Similar results can be found for female speakers. For mean VOT values in both Dutch (L1) and English (L2), no notable change was measured between recordings 1 and 5. For /d/, there was no significant language-dependent difference in duration of prevoicing though there were fewer instances of pre-voicing in English. The mean VOT value for /t/ in English was significantly longer by about 28.5 ms, or than in Dutch.

3.2. Centre of Gravity (COG for /s/)

Figure 2 shows the COG for the /s/ segments in the spontaneous monologues for the male speakers, for both English and Dutch, in rounds 1 and 5 of the recordings. Again, similar results can be shown for female speakers.

Using the *lme4* package in R, a linear mixed effects model was constructed to model the relationship between COG, recording and language. Fixed effects were *sex*, *recording* and *language*, with *speaker* as a random effect, and by-speaker random slopes for the effect of *language*. Significance was

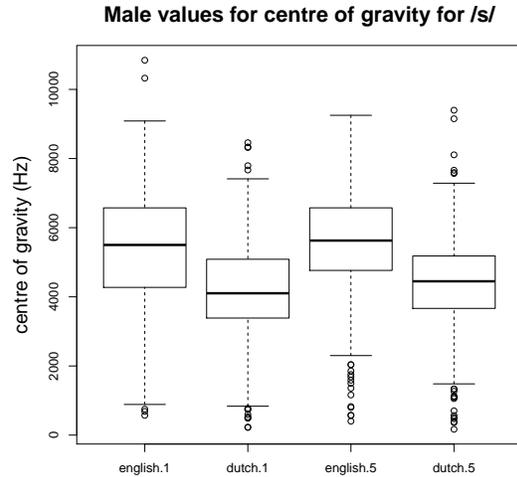


Figure 2: Values for COG for /s/, for English and Dutch in rounds 1 and 5 of the recordings, for male speakers

estimated by likelihood ratio tests on the full model with the effects of *language* and *recording round* against the model without these effects. COG for Dutch was significantly lower than that of English in both recording rounds and for both languages, higher for round 5.

4. DISCUSSION AND CONCLUSION

While the difference in VOT values for /t/ in English was significant, the amount of increase was still small, less than 30ms. This may represent some phonetic drift of /t/ towards more typical English values, but hardly a new phonetic category. COG values for /s/ started and remained lower than those for English. In this case, it seems that a new phonetic category was formed before the first recording. This makes sense, since all students are expected to be competent in English on entry. The significant though slight rise in COG for English and Dutch /s/ across recordings indicates some further phonetic drift.

A community of speakers in which L1 speakers are in the minority and there is a dominant L2 is unusual. This may explain the relative stability of these segments over time, despite the English-speaking environment. Since the majority (60%) of the speakers has Dutch L1, it might be expected that, in line with accommodation theory [4], phonetic features of the Dutch are adopted by speakers of other L1s. With English L1 speakers in the minority, we might expect that their influence is less strong.

We might also expect that other L1 speakers, including English L1, may experience accommodation to Dutch values for these segments, and that will be the next step in the investigation of phonetic drift in this community.

5. REFERENCES

- [1] Boersma, P. 2001. Praat, a system for doing phonetics by computer. *Glott International* 5(9/10), 341–345.
- [2] Boersma, P., Hamann, S. 2008. The evolution of auditory dispersion in bidirectional constraint grammars. *Phonology* 25(2), 217–270.
- [3] Collins, B., Mees, I. June 2003. *The Phonetics of English and Dutch*. Brill Academic Pub. 5 edition.
- [4] Giles, H., Coupland, N., Coupland, J. 1991. Accommodation theory: communication, context and consequence. Giles, J. C. H., Coupland, N., (eds), *Contexts of Accommodation: Development in Applied Sociolinguistics*. Cambridge University Press 1–68.
- [5] Orr, R., Quené, H., van Beek, R., Diefenbach, T., van Leeuwen, D., Huijbregts, M. 2011. An international english speech corpus for longitudinal study of accent development. *Proceedings of the 12th Annual Conference of the International Speech Communication Association* 1889–1892.
- [6] Povey, D., Ghoshal, A., Boulianne, G., Burget, L., Glembek, O., Goel, N., Hannemann, M., Motlicek, P., Qian, Y., Schwarz, P., Silovsky, J., Stemmer, G., Vesely, K. 2011. The kaldi speech recognition toolkit. *IEEE 2011 Workshop on Automatic Speech Recognition and Understanding* Hilton Waikoloa Village, Big Island, Hawaii, US. IEEE. IEEE Catalog No.: CFP11SRW-USB.
- [7] Yeni-Komishian, G., Flege, J. 2000. Pronunciation proficiency in the first and second languages of korean-english bilinguals. *Bilingualism: Language and Cognition* 3(2), 131–149.

Notes:

Phonetic and prosodic features in automated spoken language assessment

♪ Calbert Graham, ♪ Andrew Caines, ♪ Paula Buttery

♪Phonetics Laboratory, ♪ALTA Institute: DTAL, University of Cambridge

1. INTRODUCTION

The traditional approach of using trained human assessors to evaluate spoken language proficiency has proven to be both expensive and time-consuming. This has led to renewed effort to develop tools for the automated assessment of non-native speech. Previous work on automated assessment has generally not focused on acoustic features, although it has long been established that listeners use such features in judging the naturalness of L2 pronunciation. Including such features in assessment will go a long way in making the implicit knowledge examiners use in assessing pronunciation more explicit, and can be very useful, for instance, in a computer-assisted language learning (CALL) pronunciation training programme. In this paper, we describe one strand of research in our ALTA¹ project and discuss the role of phonetic features in automated assessment of non-native speech. We will briefly discuss some of the challenges we face in automatically measuring phonetic/prosodic features, and how we go about striking a compromise between what may be a linguistically meaningful feature and what is actually measurable given the constraints of our system.

Keywords: automated assessment, CALL, prosody, non-native, speech

2. CORPUS DESCRIPTION

The dataset discussed in this paper comes from a Cambridge English BULATS test of business English comprising elicited spontaneous speech (in the form of a short bio and a monologue testing the business knowledge of the candidate). Currently, our pilot dataset consists of over 1000 candidates and 20,000 recordings. Candidates are speakers of various source languages: Gujarati, Hindi, Urdu, Thai, Spanish, Portuguese, and others. As preparation for our phonetic analyses, the data was orthographically transcribed using multiple crowd-sourcers and a speech recogniser according to the procedure described in [1]. The transcribed data was then automatically segmented and aligned using a Hidden Markov Model Toolkit (HTK) MLP-based algorithm to determine word and phone boundaries.

¹ A partnership between Cambridge Assessment and the University of Cambridge.

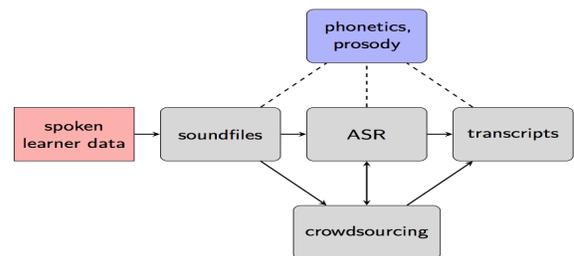
2.1. Features extracted from audio

We have extracted the following features (1) Audio-based: f0, Energy (mean, minimum, maximum, mean absolute deviation); (2) Fluency-based: long silence, silence duration, disfluencies, words, phones (mean and mean-weighted features similar to audio-based features). See [2] for a complete list. From our current feature set the system performs remarkably well with a Pearson's correlation of 0.8 with the human graders.

2.2. Phonetic features

One of the central goals of our project is to identify linguistically meaningful, criterial features in spoken language assessment, and to build a CALL system that allows candidates to improve their pronunciation based on these criteria. To that end, we seek to examine phonetic and prosodic features that can also be integrated into a pronunciation training system, as shown in Figure 1 below.

Figure 1: Phonetic and prosodic features in automated assessment systems



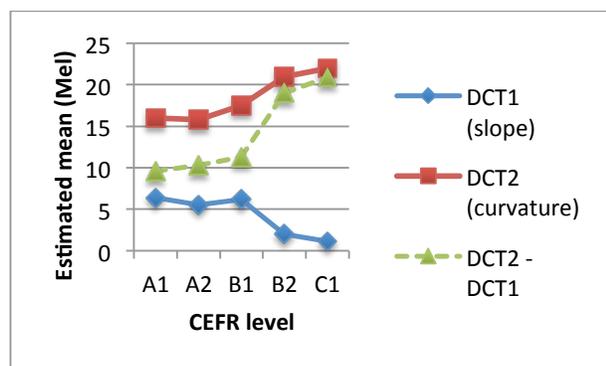
We aim to capture both the segmental and the suprasegmental (hereafter prosodic) level in speech production. Examples of features we are currently investigating include: vowel metrics, phonetic vowel reduction and sentence boundary detection. In preparation for phonetic analysis the Master Label File (MLF) output of the HTK of a pilot dataset was converted to Praat Textgrids using a script. The candidates were mostly L1 Gujarati speakers, who were multiply graded and placed into different CEFR levels. In the next section we summarise some of our on-going research.

2.2.1. Vowel quality metric

Previous research has established the relationship between vowel production accuracy and intelligibility in the L2 ([3], [4]). If confirmed as a reliable metric in

automated assessment, vowel articulation could be an excellent candidate on which to provide feedback in a CALL context. Research has shown that formant frequencies contain the primary information for the distinction of vowels [5]. Our previous pilot work [6] suggests a correlation between the vowel space of speakers (measured from formant tracking data) and their oral proficiency in the L2. However, we also observed that formant tracking may not be entirely practical with the mixed quality of the data in our corpus, which is generally unavoidable given the varying conditions in which tests are recorded in testing centres around the world. As an alternative to formant tracking we explored the Mel-scaled Discrete Cosine Transformation (DCT) method (see [5] for a detailed description). We then calculated the log Euclidean distance ratios between target vowels (e.g. the distance of [ɪ] and [i:] to a fixed anchor point (e.g. [æ]) to test the specific hypothesis of whether Gujarati speakers were making a distinction between tense and lax vowels in their L2 English, given that their L1 does not have this distinction. The log Euclidean distance ratios for the Gujarati speakers suggest a strong correlation between their vowel space and their Common European Framework of Reference (CEFR) level in English. More specifically, there is a correlation between the extent to which they realised a distinction between tense and lax vowels and their CERF levels (as shown in Figure 2).

Figure 2: Estimated DCT coefficient mean by CEFR level



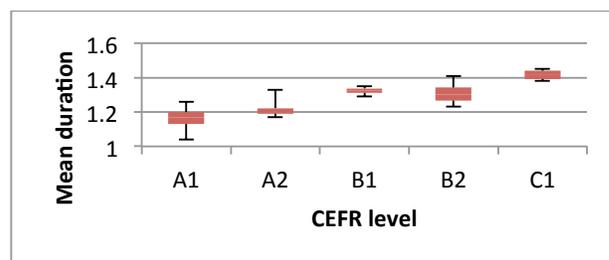
There is inherent difficulty in automatically creating accurate phone alignments. However, we find that this becomes much less problematic once we define our parameters consistently across conditions and extend our training data to include learners of different proficiencies, L1s, speaking conditions, and so on.

2.2.2. Stress and vowel reduction

In English there is a well-documented tendency for vowels in lexically unstressed position to be reduced (see e.g. [7]) when they do not carry lexical stress. This is characterised by the reducing of the acoustic correlates (f₀, energy, duration) relative to what would

be typical when the vowel is stressed. According to Bolinger [7], stress reduction can be seen as a possible measure of the rhythm of a language. Previous research suggests a strong link between the level of vowel reduction and the proficiency of the speakers (e.g. [8]). In our study, we calculated the ratio of all stressed to unstressed vowels produced by each speaker (the use of ratios means no further normalisation was necessary), and examined how effective this measure is if applied automatically (no manual corrections applied). The results suggest a very high correlation with proficiency.

Figure 3: Ratio of stress to unstressed vowel duration by CEFR level



2.2.3. Sentence boundary detection (SBD)

We also use prosodic information in the task of SBD in continuous speech, taking pitch and duration measurements to model spoken behaviour. This model is combined with a language model and sentence-length model in a log-linear fashion to decide on an optimal set of sentence breaks for a given recording, à la [9]. Where our work differs from [9] is that they worked with native speaker data whereas we apply the same methods to non-native learner data [10].

3. DISCUSSION AND CONCLUSION

Our starting point in this project is that automated assessment systems need to reflect the intuitions and implicit criteria used by human assessors to judge pronunciation. Given the reported link between phonetic and prosodic competence and the intelligibility of L2 speech, we explored ways to implement these features in the assessment of non-native English. Vowel quality and stress reduction are both easy to implement in automatic system as well as to teach in a CALL medium. SBD may also be useful in assessing the fluency of non-native speakers. Overall, the results that begin to emerge appear to support this view. In the future, we plan on extending our work by looking at higher-level intonation and other prosodic features, which we hope will ultimately lead to the building of a phonetically grounded CALL pronunciation training system. We also intend that the tools and methodologies we develop can be used in similar research on other source languages apart from English, and welcome the opportunity to further collaborate.

4. REFERENCES

- [1] Van Dalen, R., Knill, K., Psiakoulis, P., & Gales, M. (2015). Improving Multiple-Crowd-Sourced Transcriptions Using a Speech Recogniser. ICASSP.
- [2] Van Dalen, R., Knill, K., Psiakoulis, P., & Gales, M. (2015). Automatically Grading Learners' English Using a Gaussian Process. SLaTE Workshop.
- [3] Bohn, O and Flege, J. (1990). Interlingual identification and the role of foreign language experience in L2 vowel perception. *Applied Psycholinguistics*, 11, 303-328.
- [4] Bohn, O and Flege, J. (1992). The production of new and similar vowels by adult German learners of English. *Studies in Second Language Acquisition*, 14, 131-158.
- [5] Harrington, J. (2010). *Phonetic Analysis of Speech Corpora*. Wiley-Blackwell, Oxford.
- [6] Graham, C., Nolan, F., Caines, A., Buttery, P. (2015). Using vowel formant characteristics in automated assessment. PaPE, University of Cambridge (accepted).
- [7] Bolinger, D. (1986). Two kinds of vowels, two kinds of rhythm. Ms., distributed by Indiana University Linguistics Club, Bloomington, IN.
- [8] Graham, C. (2011). A case study of Japanese acquisition of American English stress. *Journal of the Phonetic Society of Japan*, 15, (3), 87.
- [9] Lee, A. & Glass, J. (2012). Sentence detection using multiple annotations. Proceedings of INTERSPEECH 2012.
- [10] Caines, A., Moore, R., Graham, C., Buttery P. Sentence boundary detection (in prep).

Notes:

ENGLISH READ BY JAPANESE PHONETIC CORPUS

Takehiko Makino

Chuo University, Tokyo, Japan
mackinaw@tamacc.chuo-u.ac.jp

Keywords: pronunciation, corpus, English Read by Japanese, Intonation Variation Transcription System

1. INTRODUCTION

English Read by Japanese (ERJ) Phonetic Corpus is a phonetically transcribed electronic corpus of the pronunciation of English by learners whose L1 is Japanese. It is intended to be a source of all the phonetic characteristics of Japanese speakers' English speech, which has tended to be described rather informally [3]. So far the transcription of individual phones is completed and prosodic transcription is underway.

2. ENGLISH READ BY JAPANESE SPEECH DATABASE

The corpus uses as its basis the 800 out of more than 70,000 recordings in English Read by Japanese (ERJ) speech database [5], which was originally collected mainly in order to help CALL system development but has not been much utilized for linguistic research.

Some of the characteristics of ERJ speech database is given below:

- 807 different sentences and 1,009 different words/word sets were used for recording. A majority of the sentences were from those used for TIMIT speech database.
- They were read aloud by 200 (100 male and 100 female) university students in 20 different recording sites all over Japan.
- The speakers did not record all of the materials; each sentence was read by about 24 speakers and each word (set) by 40.
- Phonemic transcriptions in ARPAbet were given for practice; in the recording session, the speakers only looked at orthographic sentences.

3. BUILDING PROCEDURE OF THE CORPUS

The 800 recorded sentences selected for the corpus building were the same set used in [6]. In that study, the sentences were played back on the phone to Americans who were then asked to repeat what they thought they heard, and the correlations between "errors" and the speakers' general pronunciation

scores (which were evaluated independently of this study) were calculated. The orthographic transcription of what the Americans repeated could be utilized for future study about the correspondence between pronunciation deviations and intelligibility.

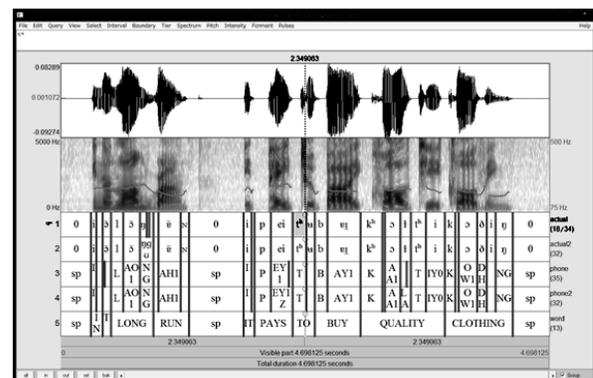
3.1. Segmental annotations

In order to facilitate the transcribing process, the recording was pre-processed by the Penn Phonetics Lab Forced Aligner (p2fa) [9], which produced forced aligned transcriptions of English words and phonemes for each file in the Praat [1] TextGrid format.

The output from p2fa was full of errors, which is not surprising at all as it was developed for native-speaker English. Thus most of the corpus-building work involved manual transcription of "actual phones," which represent what was actually pronounced in narrow phonetic IPA, and manual alignment of both "target phones," a phonemic transcription in ARPAbet which the speaker should have aimed at, and "words," to actual phones, all using Praat.

The TextGrid has two more tiers, namely "actual2" and "phone2" (Figure 1). They are there to collapse into one two or more consecutive actual phones or target phones corresponding to a single target/actual phones. They are necessary because ELAN [7], the software which is used for searching, cannot currently handle one-to-many correspondences.

Figure 1: Segmental annotations.



3.2. Prosodic annotations

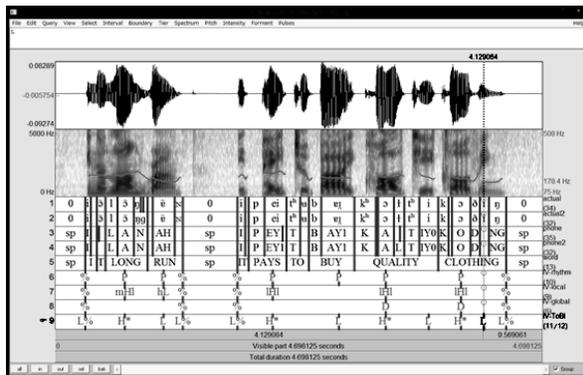
Because the prosodic system in question is part of an interlanguage [2], it cannot be transcribed with the notational system for Japanese or English. Thus it was necessary to devise a new notational system for this corpus.

Actually, this should also have been a problem for segmental transcription. But as far as individual phones are concerned, there is a framework of narrow phonetic transcription, and it was possible to use this to transcribe actual phones. Its actual implementation was not at all straightforward, though, since the truly narrow phonetic transcription independent of any language is an ideal which could not be reached. But still, it is doable.

Not so with prosody. There is no ready-to-use framework for its narrow phonetic transcription. The only framework which has been found for possible use for L2 (interlanguage) prosody is Intonation Variation Transcription System [8], abbreviated to IVTS, which was originally devised to transcribe dialect differences.

There are four tiers in IVTS: (1) rhythmic beat, (2) local phonetic pitch, (3) global pitch change such as downstep, and (4) tentative phonological pitch targets (Figure 2).

Figure 2: Segmental and IVTS annotations



The specific adaptation of IVTS to accommodate Japanese speakers' prosody of English is inevitably tentative.

Since the interlanguage system is fluid (i.e., proportions of influences from L1 and L2 should be different in one speaker/situation from another), it is not possible to devise a watertight notational system. Adjustments will be made as we transcribe more sentences, until we finish with all the 800 files. Even then the system is not complete, since there may be other utterances which could require further adjustments.

4. SOME PRELIMINARY FINDINGS

Some of the preliminary findings from the segment-only corpus survey [4] are given below:

- Voiceless plosive targets are frequently realized as fricatives, especially for /p/ (Tables 1, 2 and 3). This occurs not only in “weakening” positions (e.g., between vowels) but also in “strong” positions. Why /p/ is spirantized much more often than /t, k/ remains to be explored.

Table 1: Actual phones for target /p/ (n=549).

p	p ^h	p̄	φ	pφ	pi	others
259	84	14	123	20	15	34
				Spirantized (26.0%)		

Table 2: Actual phones for target /k/ (n=770).

k	k ^h	k̄	k ^h	k̄	k'	x	kx	ki	others
330	204	53	21	16	14	44	12	23	53
							Spirantized (7.2%)		

Table 3: Actual phones for target /t/ (n=1,280)

t	t ^h	t̄	t'	ts	tʃ	tɕ	θ	s	ti	others
579	325	58	13	28	18	16	15	10	64	154
							Spirantized (6.8%)			

- Target /r/ and /l/ are correctly pronounced in half of their tokens (Tables 4 and 5). This should mean that the distinction between them is far from absent, as is often pointed out.

Table 4: Actual phones for target /r/ (n=879).

ɹ	ʁ	r	ʃ	ə	ɐ	l	ɥ	others
389	40	162	27	69	36	57	29	70
(48.8%)	Japanese /r/ (21.5%)		Vocalized					

Table 5: Actual phones for target /l/ (n=1,037).

l	ɫ	ɾ	ʃ	ɹ	li	ri	others
385	203	139	20	102	37	36	115
(56.7%)	Japanese /r/ (15.3%)		Hyper-Corrected?		Vowel insertion		

4. REFERENCES

- [1] Boersma, P., Weenink, D. 2015. Praat: doing phonetics by computer (version 5.4.08) [Computer software]. Available: <http://www.praat.org/>
- [2] Ioup, G., Weinberger, S. H. 1987. *Interlanguage Phonology*. Cambridge, MA: Newbury House.
- [3] Makino, T. 2013. Pronunciation characteristics of Japanese speakers' English: a preliminary corpus-based study. *Pronunciation in Second Language Learning and Teaching: 5th Annual Proc.*, Ames, IA, 120-136.
- [4] Makino, T. 2014. Vowel and consonant patterns of Japanese speakers' English: a study based on English Read by Japanese Phonetic Corpus. Presentation at the 3rd Conference of the International Society for the Linguistics of English, Zurich.
- [5] Minematsu, N., et al. 2002. English speech database read by Japanese learners for CALL system development. *Proc. LREC 2002*, Las Palmas, Spain, 896-903.
- [6] Minematsu, N., et al. 2011. Measurement of objective intelligibility of Japanese accented English using ERJ (English Read by Japanese) database. *Proc. Interspeech 2011*, Florence, Italy, 1481-84.
- [7] Wittenburg, P., et al. 2006. ELAN: A professional framework for multimodality Research. *Proc. LREC 2006*, Genoa, 1556-59.
- [8] Post, B., Delais-Roussarie, E. 2006. Transcribing intonational variation at different levels of analysis. *Proc. Speech Prosody 2006*, Dresden, paper 190.
- [9] Yuan, J., Liberman, M. 2008. Speaker identification on the SCOTUS Corpus. *Proc. Acoustics '08*, Paris.

Notes:

DEVELOPMENT AND USE OF THE ESTONIAN L2 CORPUS

Einar Meister, Lya Meister

Institute of Cybernetics at Tallinn University of Technology, Estonia

einar@ioc.ee, lya@phon.ioc.ee

Keywords: Estonian, L2 speech, corpus design, quantity opposition

1. INTRODUCTION

Corpus-based approach in L2 speech research is an increasing trend facilitated by technological progress enabling cheap storage of large amounts of speech data and by availability of free software tools for manual or (semi)automatic annotation, and analysis of annotated corpora. An ideal corpus of L2 speech should comprise a large variety of speakers with different native language backgrounds, different speaking styles, extensive annotations in terms of both segmental and suprasegmental phonology, and more than one target language [2].

The paper introduces the Estonian Foreign Accent Corpus (EFAC) [8] which aims to provide high-quality L2 speech data for studies of L2 phonology and for language technology developments. The corpus collection was initiated at the Institute of Cybernetics, Tallinn University of Technology in 2006. Currently, it includes speech recordings of 180 L2 speakers of Estonian representing 18 different native languages and the reference group of 20 native Estonian speakers. The corpus is annotated on word and segmental levels.

The research on Estonian L2 speech based on this corpus has been so far focused on the acquisition of Estonian vowel categories, and on L2 prosody, specifically on the acquisition of short/long categories and quantity contrasts.

2. CORPUS DESIGN

The corpus is designed for studies of L2 acquisition of the main categories of the Estonian phonological system – vowels, consonants, diphthongs, consonant clusters, and quantity oppositions.

The text corpus involves 130 neutral sentences including all Estonian vowels and frequent diphthongs as well as all consonants and frequent consonant clusters in two-syllable target words representing the Estonian quantity oppositions. The target words are embedded in short meaningful sentences of similar structure, e.g.:

- *Kaotasin sada krooni raha.* Q1: *sada* /sata/ 'hundred', nom.sg.
- *Palun saada talle artikli koopia.* Q2: *saada* /saata/ 'to send', sg.imperat.
- *Soovin saada kolme kuu aruannet.* Q3: *saada* /saa:ta/ 'to get'
- *Kaunis lugu kõlas raadiost.* Q1: *lugu* /luku/ 'story', nom.sg.
- *Uue lukku paigaldus maksab sada krooni.* Q2: *luku* /lukku/ 'lock', gen.sg.
- *Vana lukku pole mõtet paradada.* Q3: *lukku* /luk:ku/ 'lock', part.sg.

The constructed sentences are rather simple in order to be easily comprehensible and readable for learners with intermediate knowledge of Estonian. In addition, the text corpus involves eight questions, two passages, and some prompts to elicit spontaneous speech (self-introduction, description of three pictures).

3. SPEAKER RECRUITMENT

The main criteria for L2 speaker selection were:

- proficiency level of Estonian (at least "intermediate" or higher, foreign accent must be perceived by native listeners),
- native language,
- age of learning Estonian (adult learners were preferred),
- no hearing and speaking disorder.

Ideally, it would be good to achieve balance by sex and age, but in reality it is rather impossible. A group of native speakers has also been recruited – 10 male and 10 female subjects from monolingual Estonian-speaking families living in the capital area.

Different recruitment schemes have been used – invitations have been distributed in local universities and newspapers, teachers of Estonian giving language courses for adults and teachers of Estonian working at different foreign universities have been approached, etc. All recruited speakers have filled a questionnaire containing questions about their age, native language, age of learning Estonian, where and how they have learnt the language and how often they use it, as well as self-assessment of their knowledge of Estonian ("elementary", "intermediate", "ad-

vanced" or "proficient"). All participants were paid a small amount of money.

The number of speakers in the L2 subject groups is: Russian – 50, Finnish – 30, Latvian – 20, German – 15, Lithuanian – 13, French – 13, Japanese – 6, Swedish – 6, Spanish – 5, English – 5, and Italian – 5; other language (Dutch, Slovak, Polish, Portuguese, Hindi, Azeri, and Irish) are represented by 1-2 subjects only. Attempts will be made to recruit more subjects so that all L2 subject groups will be represented by at least 10 speakers.

Among the L2 subjects ca. 60% are females and 40% males. Age of the L2 subjects ranges from 16 to 67 years (mean 29.4, median 26.1). Most L2 subjects started to learn Estonian at the age 18-30 years, some subjects at the age 7, and some at the age over 40. All subjects have studied several foreign language, mostly English, German, or French.

4. DATA COLLECTION

Majority of the speech recordings have been done in our own recording studio; several L2 speakers have been recorded at their home universities in Finland (Oulu, Helsinki and Turku), France (Paris), Austria (Vienna), Latvia (Riga), Lithuania (Vilnius). By now we have recorded 180 non-native speakers of Estonian representing 18 different language backgrounds and the reference group of 20 native Estonian speakers. Currently, the total duration of speech data is approximately 80 hours.

Technically, the recordings are mostly of high quality (sampling frequency 44.1 kHz, 16 bit, wav-format); most speakers were recorded with two condenser microphones (a close-talking mic and a desktop mic). During the recording prompts were displayed on a screen one by one, correctness of reading was monitored by a recording operator. On average, a recording session lasted 25-30 minutes.

5. ANNOTATION

The corpus is annotated in different ways: the whole corpus is segmented and labelled automatically on word and phone levels (by using an Estonian text-to-speech aligner), and the target words representing quantity oppositions are segmented manually (currently only for native Estonian speakers and for L2 speakers with Russian, Finnish, Latvian and Japanese backgrounds) using Praat software [1]. We proceed with the manual annotation since the automatic text-to-speech aligner trained on Estonian native speech does not provide satisfactory results in the case of L2 speech.

6. CORPUS-BASED STUDIES

The corpus constitutes a valuable resource for the studies of L2 acquisition of Estonian phonological categories. Estonian is a quantity language exploiting the duration cue for manifesting phonological quantity oppositions. The quantity oppositions can occur in vowels and diphthongs, and also in consonants and consonant clusters. The corpus allows us to study the acquisition of quantity contrasts by L2 subjects with different phonological relevance of duration in their L1, e.g:

- Russian and Spanish: none,
- English and German: only as a secondary cue for vowels,
- Swedish: a primary cue for most vowels and consonants,
- French: only some vowels and consonants,
- Italian: only consonants,
- Latvian and Lithuanian: only vowels,
- Finnish and Japanese: consonants and vowels.

So far, we have reported the results on the acquisition of short/long categories by Russian subjects [6] and quantity contrasts by native speakers of Russian, Finnish, and Latvian [9], [4], [3], [5], as compared to L1 speakers. The results reveal differences in L1 and L2 groups due to the different role of duration in subjects' L1.

In addition, we have studied the acquisition of Estonian vowel categories by Russian and Japanese subjects [7], [10].

In our future research we plan to study:

- how does native language type (quantity vs. non-quantity language) affect the production and perception of quantity contrasts,
- to what extent do L2 subjects use spectral and other prosodic cues (such as F0 and intensity) as secondary cues in the production and perception of Estonian quantity contrasts,
- the relationships between L2 production and perception,
- the role of orthography in the production of Estonian quantity oppositions,
- L2 speech rhythm.

As an ultimate goal, we hope to propose a new (or at least to extend an existing) L2 theoretical model that is adequately able to address prosodic features, especially duration.

7. AVAILABILITY

The corpus will be available via the Center of Estonian Language Resources (<http://keeleressursid.ee/eng/>) and via EU CLARIN infrastructure (<http://www.clarin.eu/>).

8. REFERENCES

- [1] Boersma, P., Weenink, D. Praat: doing phonetics by computer [computer program]. Version 5.4.08, retrieved 25 March 2015 from <http://www.praat.org/>.
- [2] Gut, U. 2009. *Non-native Speech. A Corpus-based Analysis of Phonological and Phonetic Properties of L2 English and German*. Peter Lang.
- [3] Meister, E., Meister, L. 2013. Native and non-native production of Estonian quantity degrees: comparison of Estonian, Finnish and Russian subjects. Asu, E. L., Lippus, P., (eds), *Nordic Prosody: Proceedings of the XIth conference, Tartu 2012* Frankfurt am Main. Peter Lang Verlag 235–243.
- [4] Meister, E., Meister, L. 2013. Production of Estonian quantity contrasts by native speakers of Finnish. *Interspeech 2013: 14th Annual Conference of the International Speech Communication Association* Lyon, France. 330–334.
- [5] Meister, E., Meister, L. 2014. Estonian quantity degrees produced by Latvian subjects. *Linguistica Lettica* 22, 85–106.
- [6] Meister, L., Meister, E. 2011. Perception of the short vs. long phonological category in Estonian by native and non-native listeners. *Journal of Phonetics* 39(2), 212–224.
- [7] Meister, L., Meister, E. August 2011. Production and perception of Estonian vowels by native and non-native speakers. *Interspeech 2011* Florence, Italy. 1145–1148.
- [8] Meister, L., Meister, E. 2012. Aktsendikorpuse ja vöörkeele aktsendi uurimine. *Keel ja Kirjandus* 55(8–9), 696–714.
- [9] Meister, L., Meister, E. 2012. The production and perception of Estonian quantity degrees by native and non-native speakers. *Interspeech 2012: 13th Annual Conference of the International Speech Communication Association* Portland, Oregon. 886–889.
- [10] Nemoto, R., Meister, E., Meister, L. 2015. Production of Estonian vowels by Japanese subjects. *18th International Congress of Phonetic Sciences* Glasgow, UK. xx–xx.

Notes:

ASSESSING PHONEMIC ACQUISITION: A NORMALIZATION-DEPENDENT PROCEDURE?

Adrien Méli

Univ Paris Diderot – Sorbonne Paris Cité
CLILLAC-ARP – EA 3967
adrienmeli@gmail.com

Keywords: automatic formant extraction, phonemic acquisition, normalization, interphonology

1. INTRODUCTION

Spontaneous speech learner corpora are likely to feature skewed word and phoneme frequencies that violate the requirements of commonly used methods to analyze phonetic data: vowel-extrinsic normalization methods such as Nearey [7], Lobanov [4] or Watt & Fabricius [12] require that acoustic measurements for all the vowels of a speaker’s system be collected in roughly the same amount.

The question therefore arises of the influence that methods of normalization may have upon assessing the evolution of learners’ phonemic acquisition over time.

2. DATA

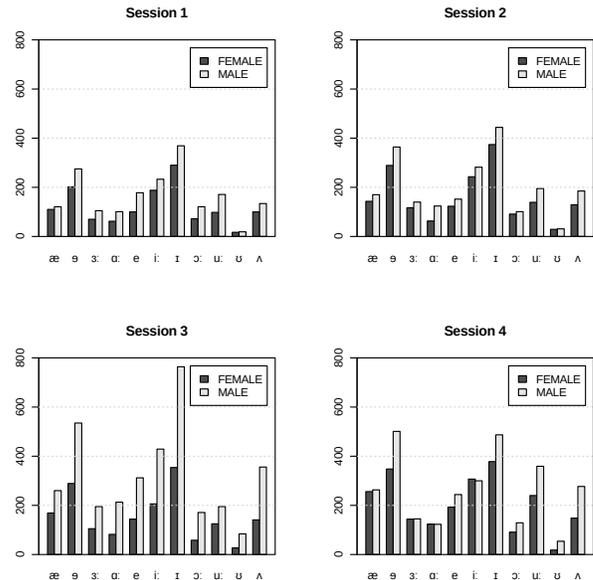
2.1. Corpus

The data used for this investigation come from a project by University Paris Diderot, in which 5 French native speakers, 3 male and 2 female students, were recorded in English four times over six-month intervals. 20 recordings were thus obtained. The interviews were conducted by a native English speaker, and recorded in an individual stereo 16-bit resolution sound file at a sampling rate of 44100 Hz captured in an uncompressed, pulse code modulation format using an Apex435 large diaphragm studio condenser microphone with cardioid polar pattern.

2.2. Procedure of acoustic extraction

The recordings were analyzed in the following fashion: the transcriptions of short, consistent sentences were aligned on a PRAAT TextGrid (Boersma [3]), which were then extracted and automatically aligned at the segmental level with SPPAS (Bigi [2]) using an American transcription of the CMU dictionary. For each vowel, a PRAAT script then collected,

Figure 1: Per-sex monophthong distribution across the four sessions.



among other things, the mid-temporal F1, F2 & F3 values. 17,407 monophthongs were thus collected and their distribution can be found in Figure 1.

3. ANALYSIS

3.1. Normalization

The data was then normalized using the four following methods:

1. Traunmüller’s [11] Bark method;
2. The Bark Difference Metric [10];
3. Nearey’s [7] Extrinsic method (Nearey 2);
4. Lobanov’s [4] z-score.

The computations were made using the statistical software R [9], and the PhonTools [1] package for the last two. As shown in Table 1, this set of methods features all possible combinations of Vowel/Formant ex-/in-trinsic computations. No speaker extrinsic method was adopted due to the differences in vowel

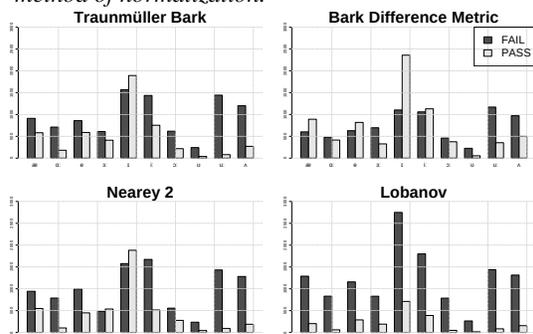
Table 1: Reminder of the specificities of the normalization methods.

Method	Vowel	Formant
Bark	Intrinsic	Intrinsic
BDM	Intrinsic	Extrinsic
Lobanov	Extrinsic	Intrinsic
Nearey 2	Extrinsic	Extrinsic

count from one speaker and one session to another. Because the evolution of pronunciation is a key research interest, the dataset was split into session-based subsets prior to normalization.

3.2. Phone-gating

Figure 2: Cross-session, cross-speaker count of phonemes phone-gated as “FAIL” or “PASS” by method of normalization.

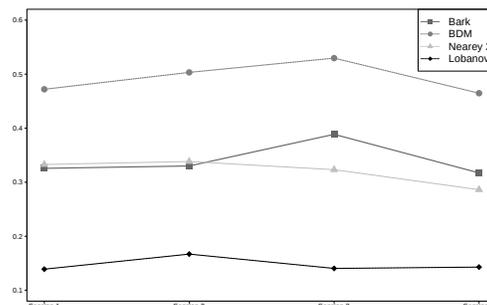


A native dataset was then normalized. For this study, data from Peterson & Barney [8]), which comes with the PhonTools [1]) package, was used for convenience. Corresponding sex-dependent minima and maxima of the F1 and F2 formant values for each vowel and each method of normalization were then stored and extracted. This procedure returned a 10-row dataframe, one row for each vowel (/ə/ was excluded), with 8 columns: the male and female F1/F2 minima and maxima. The normalized F1 and F2 values from the learners’ datasets were then checked against these extreme values, and flagged as correct (“PASS”) if both their F1 and F2 values were comprised within the natives’ range, “FAIL” otherwise.

4. DISCUSSION

Figure 2 shows the count of vowels flagged as “FAIL” or “PASS” for each session and each method of normalization. The total proportion of vowels flagged as accurate is displayed in Figure 3. Because the phone-gated values cannot be assumed to be really independent from one method of nor-

Figure 3: Proportion of phonemes tagged as accurate by method of normalization.



malization to another, pair-wise McNemar’s Chi-squared tests of independence with one degree of freedom were performed from the contingency tables in order to examine the relation between the method of normalization and the phone-gating results. The results of those pair-wise comparisons are displayed in Table 2. The extremely low p -

Table 2: Pair-wise comparisons of the phone-gating results by normalization methods using McNemar’s chi-squared tests.

Method	χ^2	p value
Bark vs. BDM	454.96	$< 2^{16}$
Bark vs. Nearey 2	1639.21	$< 2^{16}$
Bark vs. Lobanov	3173.79	$< 2^{16}$
BDM vs. Nearey 2	439.77	$< 2^{16}$
BDM vs. Lobanov	1395.56	$< 2^{16}$
Nearey 2 vs. Lobanov	3577.21	$< 2^{16}$

values mean that the null hypothesis that no differences exist between the phone-gated results from one method to another can safely be rejected.

Although these results might suggest a certain inadequacy between normalization constraints and the unbalanced characteristics of natural language (as opposed, for instance, to controlled lists of words), they may also reveal the limitations of automatic formant extractions. The extrinsic methods used in this paper rely on F3 measurements, whose proximity with F2 in vowels such as /ɪ/ may account for the wide gap in phone-gating results between the BDM and Lobanov methods for this vowel. One way to work around the relative unreliability of automatic formant extraction in high-dispersion corpora such as learners’ corpora would be to study vowel-inherent spectral change (*cf.* Nearey (1986) [6] or Morrison (2007) [5]), rather than mid-temporal values.

5. REFERENCES

- [1] Barreda, S. 2014. *phonTools: Functions for phonetics in R*. R package version 0.2-2.0.
- [2] Bigi, B. 2012. Sppas: A tool for the phonetic segmentations of speech. LREC, , (ed), *Proc. of LREC 2012* 1748–1755.
- [3] Boersma, P. 2001. Praat, a system for doing phonetics by computer. *Glott International* 5(9/10), 341–345.
- [4] Lobanov, B. 1971. Classification of Russian vowels spoken by different speakers. *J. Acoust. Soc. Am.* 49, 606–608.
- [5] Morrison, G. S., Nearey, T. M. jul 2007. Testing theories of vowel inherent spectral changea). *J. Acoust. Soc. Am.* 122(1), EL15–EL22.
- [6] Nearey, T., Assmann, P. 1986. Modeling the role of vowel inherent spectral change in vowel identification. *Journal of the Acoustical Society of America* 125, 2387-97.
- [7] Nearey, T. M. 1978. *Phonetic Feature Systems for Vowels*. PhD thesis Indiana University Linguistics Club.
- [8] Peterson, G. E., Barney, H. L. 1952. Control methods used in a study of the vowels. *The Journal of the Acoustical Society of America* 24(2), 175–184.
- [9] R Core Team, 2015. *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing Vienna, Austria.
- [10] Syrdal, A. K., Gopal, H. S. 1986. A perceptual model of vowel recognition based on the auditory representation of american english vowels. *The Journal of the Acoustical Society of America* 79(4), 1086–1100.
- [11] Traunmüller, H. 1990. Analytical expressions for the tonotopic sensory scale. *The Journal of the Acoustical Society of America* 88(1), 97–100.
- [12] Wand, D., Fabricius, A. 2002. Evaluation of a technique for improving the mapping of multiple speakers' vowel spaces in the F1 ~ F2 plane. *Leeds Working Papers in Linguistics and Phonetics* 9, 159-173.

Notes:

USING LANGUAGE TESTING CORPORA TO INVESTIGATE L2 PRONUNCIATION

Riikka Ullakonoja, Hannele Dufva, Mikko Kuronen, Maria Kautonen & Elina Tergujeff
Department of Languages, University of Jyväskylä, Finland
riikka.ullakonoja | hannele.t.m.dufva | mikko.j.kuronen | maria.e.e.kautonen | elina.tergujeff@jyu.fi

Keywords: L2 learning, prosody, Swedish

1. BACKGROUND

First, we will shortly present our project *Focus on learning pronunciation: Swedish as L1/L2* (www.jyu.fi/fokus). Second, we focus on the possibilities of using language testing corpora to investigate L2 pronunciation. The project focuses 1) on studying the phonetic aspects of L2 Swedish with speakers of Finnish, Russian and English as L1, and 2) studying aspects of Swedish L1 speakers' production of these languages as their L2. The variant we focus on is Finland Swedish, but also Swedish as spoken in Sweden will be considered (for Swedish phonology, see, e.g. [1], [3]; for differences between Finland Swedish and Swedish spoken in Sweden, see [14], [16]).

The cross-language design of the project was chosen for studying 1) the effect of different L1-L2 interrelationships (e.g. genetic relationship; writing system/orthography; phonological distance) on learning pronunciation and transfer phenomena between each L1-L2 pair and 2) the influence of various sociophonetic factors (e.g. language attitudes attached to different foreign accents) (see [13]). Hence, by using methods of acoustic and auditory analysis we will analyse features of learner language with Swedish either as L1 or L2 with the purpose of highlighting the potential language-specific difficulties, but also potential language-independent factors such as inter-individual differences (for second language pronunciation learning, see, e.g. [6], [18], [20]). While the particular focus of analysis is on prosody (see, e.g. [23]), we are also interested in segmental features typical for each variant of L2 speech (e.g. [15]), and further, in learners' own views on learning pronunciation (e.g. [12]).

In addition to its theoretical aims of gaining new information on the L2 oral performance, the project aims at a contribution to language teaching practices (see, e.g. [5], [7], [8], [19]) and design of teaching materials (see, e.g. [4], [15]). Finnish and Swedish, as two national languages of Finland, are mandatory subjects in formal education. English, while being a very widely used language of media, youth culture and business, is also the most frequent choice for the first foreign language at Finnish schools. Finally,

while Russian speakers are the largest linguistic group in Finland after the national language communities, the status of Russian at school is one of the lesser studied languages. Research-based knowledge is now particularly needed for developing teaching of oral skills because of the new oral proficiency exam to be launched (for the current practices, see, e.g. [22], [21]).

2. RESEARCH QUESTIONS

The main research questions of the project team are:

1. What phonetic/prosodic features are typical of Swedish as spoken by Finnish, Russian and English speakers? What are the most central problems for each L1 group? How can these problems be addressed in teaching pronunciation?
2. What phonetic/prosodic features are typical of English, Russian and Finnish as spoken by Swedish speakers? What are the most central problems in each L2? How can these problems be addressed in teaching pronunciation?
3. What features seem to be associated with L1 effects? What aspects may be language-independent? How do findings help to develop teaching pronunciation?
4. What is the added value of imitation experiments for understanding how pronunciation is learned?
5. How are different foreign accents evaluated in the society? How can this knowledge be used in teaching pronunciation?
6. What kind of pronunciation teaching methods seem to be effective (e.g. transcription, listening, visual feedback; teacher feedback)? How should this knowledge be used in designing teaching materials?

3. DATA

New data sets will be collected, but there are also existing corpora that will be used for analysis.

3.1 Corpora to be collected

The data sets to be collected ("FOKUS data") include: 1) Speech samples consisting of reading aloud (sentences, a short text) and spontaneous speaking (a short oral narrative task) to be analysed

acoustically and by perception tests. Data includes a) L2 Swedish as produced by Finnish, Russian and English speakers and b) L2 Finnish, English and Russian as produced by Swedish speakers. 2) Imitation tasks based on pilot studies by the authors [24, 25]. Task involves imitation of utterances by learners and non-learners (i.e. persons who have not studied the language) to be analysed acoustically and by perception tests. 3) Data on Finnish university students learning Swedish. The data includes both the students' oral performance, its analysis and the students' interviews. Data (both read-aloud and spontaneous speech) are being collected on Swedish pronunciation courses given at University of Jyväskylä.

3.2 Existing language testing corpora

The existing language testing corpora include: 1) The Finnish National Board of Education Learning Performances and 2) The National Certificates of Language Proficiency. The first corpus consists of the learning performances of Finnish ninth graders in Swedish, Russian and English in a test organized by the Finnish National Board of Education [9, 10, 11]. The oral skills sub-test has been video-recorded and evaluated on the Finnish National Curriculum Scale (a more fine-tuned version of the CEFR-scale). It consisted of four different tasks in each language, one of which was a monologue and three were dialogues. In addition, extensive background questionnaires have been collected from both the pupils and their teacher. All the participants are of the same age, come from a similar language learning background (Finnish schools), but have different L1s: Finnish, Finland-Swedish, and some Russian. They have taken the test in Swedish, Russian or English and we have taken a sub-sample of their performances to be investigated more thoroughly.

The National Certificates of Language Proficiency data consist of test performances given in a national language proficiency test evaluated across CEFR (Common European Framework for Languages) proficiency levels. The test takers come from a variety of language learning backgrounds (no specific course or study unit is required to take the test) and socioeconomic groups. In the test, each speaker is recorded typically doing two monologue tasks and some "interactional" tasks, where s/he is asked to have a conversation with the person s/he hears on the tape.

4. METHODS

Methods will include using tools of experimental phonetics such as acoustic analysis and perception tests. Programs such as Praat [2] and KayPENTAX'

Computerized Speech Lab (<http://www.kaypentax.com>) will be used. Results will be analysed using statistical means.

5. DISCUSSION

Using language testing corpora as data in L2 pronunciation research has both advantages and disadvantages. By using existing data, the steps of recruiting participants and making recordings can be avoided, and the speakers' oral proficiency level is reliably assessed. However, finding suitable samples from the data bank may be laborious due to missing video or audio files and consent forms. Also, as the conditions of the recordings vary, their quality is not always optimal. Further, due to ethical principles of the language testing system, the instructions of the speaking tasks themselves are not available and the use of background information of the subjects is also limited. Similarly, as the original video file is not available to researchers, it may be difficult to distinguish the speakers in dialogue tasks.

In comparison, we have also met challenges in finding suitable designs and tasks for the purposes of the present project, especially as concerns spontaneous elicitation. For example, when piloting tasks, we found a picture telling task unsatisfactory for investigating tone accent in Swedish as it resulted in pitch patterns that are typical of reading a list. Hence, while monologue tasks might be more suitable for such purposes, they present other types of challenges, such as their potential unauthenticity.

6. ACKNOWLEDGEMENTS

The project *Focus on learning pronunciation: Swedish as L1/L2* is funded by The Society of Swedish Literature in Finland (SLS).

7. REFERENCES

- [1] Bannert, R. 2004. *På väg mot svenskt uttal*. Lund: Studentlitteratur.
- [2] Boersma, P., Weenink, D. 2014 PRAAT: Doing phonetics by computer, Version 5.4, [Computer program] <http://www.praat.org/> (University of Amsterdam, Amsterdam)
- [3] Bruce, G. 2012. *Allmän och svensk prosodi*. Lund: Studentlitteratur.
- [4] de Silva, V., Mäkilä, K., Kärkkäinen, H., Ullakonoja, R. 2010. *Uusi venäjän ääntämisopas*. Helsinki: Finn Lectura.
- [5] Derwing, T., Munro, M. 2005. Second Language Accent and Pronunciation Teaching: A Research-Based Approach. *TESOL Quarterly*, 39:3, 379–397.
- [6] Flege, J. E. 1995. Second language speech learning: Theory, findings, and problems. In: W. Strange (ed). *Speech Perception and Linguistic Experience. Issues in Cross-Language Research* Timonium: York Press, 233–277.
- [7] Gilbert, J. 2008. *Teaching Pronunciation. Using the Prosody Pyramid*. Cambridge: Cambridge University Press.
- [8] Henderson, A., Frost, D., Tergujeff, E., Kautzsch, A., Murphy, D., Kirkova-Naskova, A., Waniek-Klimczak, E., Levey, D., Cunningham, U. & Curnick, L. 2012. The English Pronunciation Teaching in Europe Survey: Selected results. *Research in Language* 10.1, 5–27. <http://www.diva-portal.org/smash/get/diva2:564137/FULLTEXT02>
- [9] Hildén, R., Rautopuro, J. 2014a. Ruotsin kielen A-oppimäärän oppimistulokset perusopetuksen päättövaiheessa 2013. Helsinki: Finnish National Board of Education. Available (with summary in English) <http://oph.fi/download/160073-ruotsin-kielen-a-oppimaaran-oppimistulokset-perusopetuksen-paattovaiheessa-2.pdf>.
- [10] Hildén, R., Rautopuro, J. 2014b. Venäjän kielen A- ja B-oppimäärän oppimistulokset perusopetuksen päättövaiheessa 2013. Helsinki: Finnish National Board of Education. Available (with summary in English) <http://oph.fi/julkaisut/2014/venajan-kielen-a-ja-b-oppimaaran-oppimistulokset-perusopetuksen-paattovaiheessa-2013>.
- [11] Härmälä, M., Huhtanen, M., Puukko, M. 2014. *Englannin kielen A-oppimäärän oppimistulokset perusopetuksen päättövaiheessa 2013*. Helsinki: Finnish National Board of Education. Available (with summary in English) <http://oph.fi/download/160066-englannin-kielen-a-oppimaaran-oppimistulokset-perusopetuksen-paattovaiheessa.pdf>.
- [12] Kautonen, M. 2014. *Språkliga identiteter hos blivande språklärare*. Master's thesis. Department of Languages, University of Jyväskylä. <http://urn.fi/URN:NBN:fi:jyu-201410243088>
- [13] Kautonen, M., M. Kuronen, R. Ullakonoja, E. Tergujeff, Dufva, H. (2015). På väg mot bättre språkundervisning – FOKUS på uttal, *Language, Education and Society*, March 2015. <http://www.kieliverkosto.fi/article/pa-vag-mot-battre-sprakundervisning-fokus-pa-uttal/>
- [14] Kuronen, M., Leinonen, K. 2000. Fonetiska skillnader mellan finlandssvenska och rikssvenska. In: L. Jönsson et al. (eds) *Svenskans beskrivning 24. Linköpings universitet*. <http://www.ep.liu.se/ecp/006/011/ecp00611.pdf>
- [15] Kuronen, M., Leinonen, K. 2010. Svenskt uttal för finskspråkiga – teori och övningar i finlandssvenskt och rikssvenskt uttal för finskspråkiga. Tampere: Tampereen yliopistopaino.
- [16] Kuronen, M. 2000. Vokalluttalets akustik i sverigesvenska, finlandssvenska och finska. *Studia Philologica Jyväskyläensia* 49. Jyväskylä: University of Jyväskylä. <https://jyx.jyu.fi/dspace/handle/123456789/25579>
- [17] Kuronen, M. 2015. Bryter finlandssvenskarna när de talar finska? *Folkmålsstudier* 52.
- [18] Major, R. C. 2008. Transfer in second language phonology: A review. In: J. G. H. Edwards, M. L. Zampini (eds) *Phonology and Second Language Acquisition. Studies in Bilingualism* 36 Amsterdam: John Benjamins Publishing Company, 63–94.
- [19] Norris, J.M., Ortega, L. 2000. Effectiveness of L2 Instruction: A Research Synthesis and Quantitative Meta-analysis. *Language Learning*, 50:3, 417–528.
- [20] Piske, T., MacKay, I., Flege, J. 2001. Factors affecting degree of foreign accent in an L2: a review. *Journal of Phonetics*, 29, 191–215.
- [21] Tergujeff, E. 2013. *The English Pronunciation Teaching in Finland*. Jyväskylä Studies in Humanities 207. Jyväskylä: University of Jyväskylä. <http://urn.fi/URN:ISBN:978-951-39-5322-5>
- [22] Tergujeff, E., Ullakonoja, R., Dufva, H. 2011. Phonetics and Foreign Language Teaching in Finland. In: S. Werner, T. Kinnunen (eds) *XXVI Fonetikan päivät 2010*. Joensuu: UEF. 63–68. <http://urn.fi/URN:ISBN:978-952-61-0391-4>
- [23] Ullakonoja, R. 2011. *Da. Eto vopros!: prosodic development of Finnish students' read aloud Russian during study in Russia*. Jyväskylä: Jyväskylän yliopisto. Jyväskylä Studies in Humanities 151. <http://urn.fi/URN:ISBN:978-951-39-4209-0>.
- [24] Ullakonoja, R., H. Dufva, M. Kuronen, P. Hurme. 2014. How to imitate an unknown language? Russians imitating Finnish. In K. Jähi, L. Taimi (eds.) *XXVIII Fonetikan päivät*, Turku: University of Turku. 10–18. <http://www.utu.fi/fi/sivustot/fp2013/Sivut/Konferenssijulkaisu.aspx>
- [25] Ullakonoja, R., Kuronen, M., Hurme, P., Dufva, H. 2014. Segment duration in Finnish as imitated by Russians. In: N. Campbell, D. Gibbon, D. Hirst (eds.) *Social and Linguistic Speech Prosody. Proceedings of the 7th Speech Prosody Conference*. 507–511. Dublin: Science Foundation Ireland; International Speech Communication Association. <http://fastnet.netsoc.ie/sp7/sp7book.pdf>

Notes:

The extended COREIL corpus: first outcomes and methodological issues

Elisabeth Delais-Roussarie¹, Fabián Santiago^{1,2} & Hi-Yon Yoo¹

¹UMR 7110-LLF, Université Paris Diderot, ²UMR 7018-LPP, Université Sorbonne Nouvelle

elisabeth.roussarie@wanadoo.fr, fabian.santiago.ling@gmail.com, yoo@linguist.jussieu.fr

Keywords: L2 prosody, recording protocol, subject selection, intonation

1. INTRODUCTION

Since several years, numerous researches on second language acquisition have been based on corpus studies (cf. [1] and [2] among others), allowing a better evaluation of possible correlations between learner's L1, grammatical competence and proficiency level in L2. Among the studies focusing on the acquisition of L2 phonology, phonetics and prosody, a distinction can be made according to the way of collecting and analysing the data. Acoustic data, for instance, may be gathered by means of experimental procedures or may be extracted from larger corpora [3]. Experimental approaches have the advantage of allowing the control of various elements that may come into play in the production process and to focus on the targeted structures, but the obtained data may not always be a good sample of the learner's proficiency and may be limited. Corpus-based approaches, by contrast, allow gathering a large data set that may provide better insights on the L2 phonological /phonetic acquisition process (cf. [4], [5] and [6]). Even if the structure and content of the corpus used may vary greatly, corpus-based studies display several advantages in L2 research:

- Analysing a large amount of data allows investigating the interaction of several factors at the same time (for instance, the relation between proficiency level and distribution of tonal patterns, the relation between syntactic complexity and prosodic phrasing, etc.).
- It is possible to put in perspective several explanatory factors affecting the L2 acquisition process and competence (be they linguistic or not), e.g.: L1 transfer, speech style (reading vs. spontaneous speech), age when learners start acquiring the L2, etc.
- When the corpus includes data from a large array of languages and from various speaking styles, it is more representative and allows making generalisations on the acquisition of a foreign language. In addition, cross-comparisons between language pairs are

possible and allow identifying the factors motivating L2 prosodic/ phonological 'errors'.

By taking all these issues, the COREIL corpus, a large learner speech database designed for studying the L2 prosody, was developed [6]. After presenting the principles at play in designing the data collection and annotation protocols, we will (i) describe the data that are currently compiled in the extended version of the COREIL corpus, (ii) briefly present some results obtained, and (iii) discuss some remaining methodological issues that must to be improved in order to successfully share these resources with the linguistic community.

2. THE EXTENDED COREIL CORPUS

2.1 Basic principles and facts

The protocols used to gather and annotate the data from the COREIL corpus were thought in such a way as to (i) avoid making strong presuppositions (such as the idea that L1 transfer is crucial); (ii) allow making contrastive analysis between learners' and natives' oral productions with comparable data sets; (iii) allow evaluating the learner oral competence and the L2 proficiency level while taking into account a large array of tasks/ skills (reading speech, monologal and interactive speech); and (iv) recording speakers with different L1.

2.2 Recording protocol and tasks

In order to obtain different speaking styles, the speakers were recorded while performing five distinct tasks that are classified into three groups. The first group includes two interactive oral production tasks (IOP). In one of them, the speakers were interviewed (they were asked to talk about their projects, their experience in French courses, etc.), while, in the second, they had to perform a role-play, in which they asked questions to complete an enrolment form. The second group consisted of two monologue oral production tasks (MOP), including first the description of a painting and second the narration/narration of a picture representing a group of people involved in an activity. The third group consisted in a reading task (RT), in which the speakers had to read short dialogues and several texts adapted from the

EUROM 1 corpus (cf., for more details, [5]). All participants were asked to read the texts and dialogues several times before the recording session.

The recordings took place in a quiet room and were done with an Edirol R09 digital recorder. The questions used in the current study were extracted from two types of tasks: IOP and RT.

2.3 Participants

Two groups of participants are distinguished in our corpus: the native speakers (or control groups) and the learners (experimental groups). In control groups speakers were recorded in their L1, whereas learners were recorded in the L2 target language. The COREIL corpus was designed to collect data in L2 French and English produced by speakers with English and Mandarin as L1. But in its current state, data from the following language background have been recorded, transcribed and are thus available: native spoken data from French, Mexican Spanish, German, Korean and Greek speakers, French L2 data from Mexican Spanish, German, Korean and Greek learners and French learners of Korean. As for the age of learners, no constraint was imposed, but the L2 was always acquired when the speakers had reach at least 10 years of age. Table 1 summarizes all these information. The proficiency level of learners was encoded according to the Common European Framework of Reference for Languages (CEFR).

Table 1. *Speakers' profile*

Group	L1	Participants	Level
L2 French	German	8	A2&B1
	Greek	4	A1, A2 & B1
	Korean	3	A1 & A2
	Spanish	15	A2&B1
L1 French	French	10	Native
L1 German	German	8	Native
L1 Greek	Greek	2	Native
L1 Korean	Korean	2	Native
L1 Spanish	Spanish	15	Native

3. USE OF THE COREIL CORPUS

Up to know, the COREIL corpus has been used to develop or evaluate annotation tools and systems, but also to study specific intonational patterns.

Since most prosodic annotation systems have been developed at the phonological levels, it was important to evaluate how they could be used to encode learner prosody in oral data [7]. In addition, some of this data has been used to develop PROSOTRAN, a transcription tools that relies on the prosodic parameters at the syllabic level to provide a symbolic transcription [8].

Studies on the acquisition of intonation in an L2 have also been achieved with data extracted from the COREIL corpus. In studying and comparing the

tonal configurations realized at the end of neutral yes/no questions and wh-questions in native French, Spanish and L2 French, [9] showed that some L2 intonational patterns, in particular high rises at the end of questions, cannot be attributed to an L1 transfer but rather to the effects of iconic universal tonal representations. In the same vein, [10] showed that German and Spanish learners of L2 French realize high rises at the end of non-final IP at early stages of the acquisition process, even if such form doesn't occur in their L1 nor in native French. This tonal pattern may be interpreted as a sign of linguistic insecurity.

4. CONCLUSION AND PERSPECTIVES FOR IMPROVEMENT

The COREIL Corpus had several features that allow getting interesting insights on the acquisition of intonation in an L2. As it includes data from different styles recorded by speakers with various proficiency levels, the results obtained are probably less disputable. The protocol used allows comparing productions from speakers with different L1, which offer promising results. Yet, the results of the studies achieved are still limited, several methodological issues requesting improvements:

- the protocol employed for eliciting questions does not allow gathering a large set of different question types (coordinated questions, echo-questions, etc.)
- the evaluation of the L2 proficiency levels is sometime problematic, as the CEFR does not really take into account suprasegmental features in the descriptors.

In order to improve these issues, we wish (i) to add to the protocol more interactive oral tasks for eliciting questions, and (ii) to develop tests for better identifying the proficiency level of the learners regarding prosody.

In order to reinforce some of the results obtained by comparing the productions of learners with different L1, it would be crucial to share the protocol for gathering a larger amount of data that can be compared.

ACKNOWLEDGEMENTS

This work was supported by the French Investissements d'Avenir - Labex EFL program (ANR-10-LABX-0083). It also benefited from a doctoral grant from CONACyT (Mexico).

REFERENCES

- [1] Granger S. (2003). The International Corpus of Learner English: A New Resource for Foreign Language Learning and Teaching and Second

- Language Acquisition Research. *TESOL* 37/3:538-546.
- [2] Hawkins J.A. & P. Buttery. (2009). Using learner language from corpora to profile levels of proficiency. *Studies in Language Testing*. Cambridge University Press.
- [3] Delais-Roussarie E. & H. Yoo (2014) Corpus and research in phonetics and phonology (I): Methodological and formal considerations. In Durand J., Gut U. & G. Kristoffersen (eds), *Handbook of Corpus Phonology*. Oxford University Press
- [4] Gut U. (2009). *Non-native Speech. A corpus-based analysis of the phonetic and phonological properties of L2 English and L2 German*. Peter Lang.
- [5] Herment S., A. Tortel, B. Bigi, D. Hirst & A. Loukina (to appear) AixOx, a multi-layered learners' corpus: automatic annotation. In Díaz Pérez, J. & A. Díaz Negrillo (eds.), *Specialisation and variation in language corpora*. Peter Lang.
- [6] Delais-Roussarie, E. & H. Yoo. (2011). Learner corpora and prosody: from the COREIL corpus to principles on data collection and corpus design. *Poznań Studies in Contemporary PSCIL* 471: 28-39.
- [7] Delais-Roussarie, E & H. Yoo (2011). Transcrire la prosodie : un préalable à l'échange et à l'analyse de données. *Journal of French Language Studies* 21/1 : 17-27.
- [8] Bartkova K., Delais-Roussarie E. & F. Santiago-Vargas (2012). PROSOTRAN : a tool to annotate prosodically non-standard data. *Proceedings of Speech Prosody 2012*.
- [9] Santiago, F. & E. Delais-Roussarie. (2015). The acquisition of question intonation by Mexican Spanish Learners of French. In Delais-Roussarie, E., M. Avanzi & S. Herment (eds.), *Prosody and languages in contact: L2 acquisition, attrition, languages in multilingual situations*. Springer Verlag.
- [10] Santiago, F. & E. Delais-Roussarie. (2015). What motivates extra-rising patterns in L2 French: Acquisition Factors or L1 Transfer?, *Proceedings of the 18th International Congress of Phonetic Sciences (ICPhS)*.

Notes:

THE DEVELOPMENT OF A HUNGARIAN–ENGLISH LEARNER SPEECH DATABASE AND A RELATED ANALYSIS OF FILLED PAUSES

Mária Gósy – Dorottya Gyarmathy – András Beke

Research Institute for Linguistics, Hungarian Academy of Sciences, Budapest
gosy.maria@nytud.mta.hu – gyarmathy.dorottya@gmail.com – beke.andras@gmail.com

Keywords: English–Hungarian database, three proficiency levels, five speech styles, annotations.

1. INTRODUCTION

The third revolution in the history of speech science has been identified as the development of large databases. By now, the importance and usefulness of such databases are unquestionable. They provide a unique possibility to do research on various speech phenomena from fine phonetic events to diverse characteristics of long narratives. Beside native speakers' databases language learner corpora have also started being developed. The size, contents, annotation level, and other characteristics of existing spoken corpora are very different. They focus on English as L2 and mainly contain written materials. The largest corpus of informal interviews of learners of English (from higher intermediate to advanced students) has been coordinated by the University of Louvain [1]. This corpus contains 554 interviews produced by learners with eleven first languages (Hungarian is not represented among them).

The present study has two principal purposes. (i) It intends to present the design and development of our learner database (HunEng-D). (ii) We present our research concerning filled pause patterns of Hungarian and English speech based on the database.

The two languages differ in a number of linguistic facts. Hungarian is an agglutinative language belonging to the Finno-Ugric language family. It has an extremely rich morphology with extensive affixation and, as a consequence, syntactic and semantic functions of noun phrases are primarily expressed via suffixes and postpositions. Case endings are used extensively with nouns, but pronouns, adjectives, and numerals also take case and number endings. Verbs also have a considerable number of prefixes and suffixes. It has an inventory of 39 phonemes, including also phonologically short and long ones (but there are no diphthongs, no neutral vowels, and no aspirated consonants).

2. THE LEARNER DATABASE (HUNENG-D)

This is the first spoken language learner database that (i) contains speech materials of Hungarian-speaking participants whose second language is English, (ii) is recorded under the same conditions

and following the same protocol, and (iii) contains annotated speech materials. The aim of the project is to develop a large database of 150 monolingual speakers whose native language is Hungarian and whose command of English represents various L2 proficiency levels.

2.1. Participants

At the time of writing, the total recorded material involves 60 speakers (half of them females) and amounts to about 30 hours. Hungarian is the mother tongue of all participants; they learn English as L2 at school. 20 of them are basic learners (BL) with two to three years of L2 instruction at school (14–15-year-olds), another 20 of them (18–19-year-olds) are intermediate learners (IL) and the third group consists of 20 advanced learners (AL) (ages between 22 and 28 years). The L2 language proficiency of the youngest group is B1, of the second group is B2 while that of the third group is C1 according to the categories of the Common European Framework.

2.2. Recording protocol

The database contains various types of spontaneous speech materials including also a word list to be read. The protocol consists of 5 modules. 1. Narratives about the participant's life, family, job, and hobbies. 2. Narrative-like opinions about a topic of current interest, provided by the interviewer. The topics are adjusted to the participants' age. 3. Précis (summary of content), that is, directed spontaneous speech. The participant reads a story and then s/he has to summarize its content in his/her own words. 4. Two types of methods (a map task and a speech game) to elicit quasi-natural conversations between the participants. 5. A word list consisting of forty-eight Hungarian words and another forty-eight English words (with voiced and unvoiced stop consonants as the initial segments of the words).

2.3. Recording conditions

Recordings are invariably made in the same sound-proof booth (at the Research Institute for Linguistics), under identical technical conditions, digitally, direct to the computer using the same professional microphones. In all recordings (both in English and in Hungarian) the interviewer was the second author.

2.4. Annotation

The speech material has been manually annotated by two trained transcribers, phoneticians with high-level proficiency in English) while two authors have continuously double-checked the annotations. The transcription was done in Praat at two levels (phrase level and word level) according to the criteria and rules that had been developed.

3. FILLED PAUSE ANALYSIS

The term ‘filled pause’ refers in this paper to the phenomenon of diverse vocalic, nasal, mixed or other sound events inserted into the spontaneous utterances. Filled pauses do occur in spontaneous narratives in any language. They may be used for several reasons and in various functions, and occur in diverse verbal forms of the languages (e.g., [2]). English speakers use *uh* and *um*, Portuguese ones use *uum*, [v], and [ə], Japanese speakers use *ano*, *e*, *eto*, *ma*, Basque speakers use *e*, *m*, *zera*, while Hungarians prefer neutral vowels [4, 5, 6], etc. Data of analysis of filled pauses in L1 and L2 speech show that they are more frequent and longer in L2 than in L1, and occur characteristically within the clause in L2 narratives (e.g., [3]).

The main theoretical interest that guided our research was to find out whether filled pauses show different occurrences and phonetic patterns in the two languages, whether they are flanked by lexical items or by silent periods, and whether they show different patterns suggesting different proficiency levels. Three hypotheses were formulated. (i) Filled pauses occur more frequently and with longer durations in L2 than in L1 narratives. (ii) Phonetic patterns and flanking character of the filled pauses reflect the influence of L1 on pausing in L2. (iii) Filled pauses occur more frequently at clause boundaries in L1 than in L2 utterances.

Thirty recordings were randomly selected for analysis from HunEng-D, ten from each proficiency level. Occurrences of the filled pauses in terms of their position in clauses, their phonetic forms, durations, and the first two formant frequencies were defined. To test statistical significance, cross-tabulation and GLMM tests were used (SPSS 19.0).

Results show that, as expected, the great majority of the filled pauses (72%) were neutral vowels in both L1 and L2 (out of 4612 filled pauses). Analysis was focused on these pauses. Language proficiency had a significant effect on the occurrence of filled pauses (Chi-square = 89.76; $p < 0.05$). More filled pauses were found in L2 (2411 instances, 2.5/min) than in L1 (928 instances, 1.2/min) irrespective of level of proficiency. In L1, 29.4%, in L2 21% of all filled

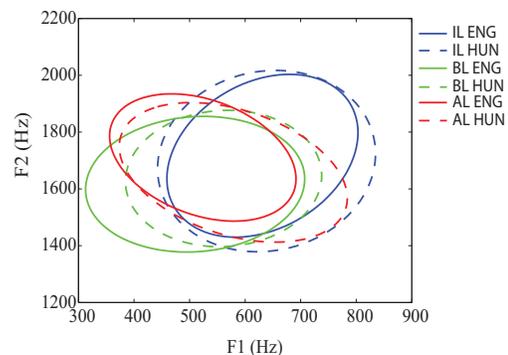
pauses were flanked by two silent pauses. They occurred most frequently with IL learners.

Durations of the filled pauses were analysed in terms of four fixed factors: L1 vs. L2, level of proficiency, position of the filled pause in the clause, and whether it was flanked by silent pauses. Statistical analysis revealed significant differences across positions ($F(1, 3113) = 228,626$; $p = 0.001$) and languages ($F(1, 3113) = 41,312$; $p = 0.001$). However, no significant differences were found in filled pause durations depending on language proficiency with the exception of IL vs. AL. The durational differences among speakers depending on level of proficiency were 10 to 26 ms, on average. The mean duration of the filled pauses was longer in L1 than in L2 (329 ms and 366 ms, respectively). The shortest durations occurred with AL and BL, while the longest ones were found with IL both in L1 and L2.

Pairwise contrasts confirmed significant durational differences depending on the positions of the filled pauses. Filled pauses were shorter within clauses than at clause boundaries both in L1 and in L2, particularly in the case of the BL and IL groups. In addition, filled pauses flanked by two silent pauses were significantly longer (397 ms, on average) than those surrounded lexically (304 ms, on average).

Formant frequencies of the filled pauses showed significant differences depending on level of proficiency (for the first formant: $F(2, 3113) = 104.330$; $p = 0.001$ and for the second formant: $F(2, 3113) = 53.161$; $p = 0.001$). F2s significantly differed also depending on the language ($F(1, 3113) = 4.048$; $p = 0.044$), see Fig. 1. The formant differences suggest that speakers produce the same neutral vowels as filled pauses but with different vowel qualities.

Figure 1: Formant frequencies of filled pauses across languages and groups.



Conclusions will be drawn on the usage of filled pauses in terms of L1 vs. L2, language proficiency and phonetic patterns of the filled pause articulation.

4. REFERENCES

- [1] Clark, H. H., Fox Tree, J. E. 2002. Using *uh* and *um* in spontaneous speaking. *Cognition* 84, 73–111.
- [2] Gilquin, G., De Cock, S. & Granger, S. (eds.) 2010. *LINDSEI Louvain International Database of Spoken English Interlanguage*. Presses Universitaires de Louvain.
- [3] Trouvain, J., Fauth, C. 2014. Zu phonetischen Details von Pausen – Untersuchungen von Lesesprache in L1 vs. L2. *Phonetik und Phonologie* 10. Konstanz.
- [4] Urizar, X., Samuel, A. 2014. A corpus-based study of fillers among native Basque Speakers and the role of *zera*. *Language and Speech* 57, 338–366.
- [5] Veiga, A., Candeias, S., Lopes, C., Perdigão, F. 2011. Characterization of hesitations using acoustic models. *Proc. of ICPhS*. Hong Kong, 2054–2057.
- [6] Watanabe, M., Hirose, K., Den, Y., Minematsu, N. 2008. Filled pauses as cues to the complexity of upcoming phrases for native and non-native listeners. *Speech Communication* 50, 81–94.

Notes:

REPAIR BEHAVIOUR OF ADVANCED GERMAN LEARNERS IN THE BERLIN MAP TASK CORPUS

Malte Belz, Simon Sauer, Anke Lüdeling, Christine Mooshammer

Humboldt-Universität zu Berlin

Corresponding authors: {malte.belz | simon.sauer}@hu-berlin.de

Keywords: disfluencies, repairs, learner speech, BeMaTaC, multi-layer annotation

1. DISFLUENCIES IN NON-NATIVE SPEECH

How do learners (L2) differ from native speakers (L1) when speaking disfluently? We will discuss this question for the use of self-repairs [10], henceforth called repairs, based on spoken corpus data.

Disfluencies such as repairs, filled pauses, truncations, prolongations etc. occur in about 6% of every 100 words uttered in spontaneous speech [3]. While disfluencies are widely accepted in L1 speech, they are sometimes regarded as problematic in L2 speech. Comparing the disfluencies of L1 and L2 speakers may instead lead to a better understanding of the adaptation processes of learners in different stages of interlanguage, when some features of the target language have not yet been acquired. For example, a non-native-like articulation of filled pauses may contribute to the so-called *foreign-soundingness* of L2 speech [7]. Reflections of the underlying L1 articulation system can often be observed, e.g. in the articulation of the filled pause *äh* [ɛ:] in German as *eu*h [œ:] in L2 German with a background of L1 French [6]. However, a foreign-sounding articulation constitutes only a part of what defines non-native speech. In addition to grammatical and lexical differences, a non-native-like proficiency is often detected by the use of deviating speech patterns and frequency distributions. It has been stipulated that learners produce disfluencies more frequently than native speakers. In a corpus study of 50 interviews with advanced German learners of English, 84% use significantly more filled pauses than native speakers [4]. However, it remains unclear whether this is merely caused by higher cognitive processing costs. It may as well be that “speakers have a strong inclination to transfer their pause profile from L1 to L2 performance” [8, p. 270]. With this in mind, we predict that repairs in L2 German show the some kind of transfer. We will examine this hypothesis using the Berlin Map Task Corpus.

2. THE BERLIN MAP TASK CORPUS

The Berlin Map Task Corpus (BeMaTaC; <https://u.hu-berlin.de/bematac>) is a freely available corpus of spoken German. It consists of an L1 subcorpus recorded with native speakers of German and an identically designed L2 subcorpus with advanced speakers of German as a foreign language (to date, all learners in the corpus are native speakers of English and have test scores equivalent to ECFR level C1 or above).

BeMaTaC uses a map-task design, where one speaker (the instructor) instructs another speaker (the instructee) to reproduce a route on a map with landmarks [1]. The speakers cannot see each other and are thus unable to communicate non-verbally. The dialogues are recorded with two separately placed microphones and a video showing the drawing hand of the instructee. Transcriptions are consistently tokenized and time-aligned. Instead of inline annotations, BeMaTaC is using a multilayer stand-off architecture [9]. This allows for a wide and easily extendable range of different annotations layers. These include layers for general disfluencies, repairs, repair sub-classifications and silent pauses. Extensive and anonymized metadata are provided with every dialogue. In learner dialogues, this also includes proficiency levels, surveyed by means of a C-Test [5], as well as the details of L2 acquisition.

In its pilot release (all data in this abstract are based on the 2013-02.1 release), BeMaTaC has 12 L1 dialogues with a total of 66 minutes and 13891 tokens and 5 L2 dialogues with a total of 77 minutes and 5155 tokens. Token counts are based on the diplomatic transcription which includes all verbal utterances, including filled pauses and truncated words. Contractions such as *haste* (*have-you*) are counted as one token only. Non-verbal utterances such as laughter and silent pauses are excluded. Disfluencies make up for 8.8% in native speakers and 14.7% in learners (we exclude unfilled pauses in this abstract). The corpus is accessible via ANNIS [13], an open-source browser-based search and visualization tool for deeply annotated corpora (<http://annis-tools.org>; for access-

Figure 1: Reparanda (*rd*), interregna (*ir*) and reparantia (*rs*) on *instructor_repair* and substitutions (*s*), repetitions (*r*) and insertions (*i*) on *instructor_subrep*, Engl. *turn-you uhm tu/ turn you*.

tok	649	650	651	652	653	654	655
<i>instructor_dipl</i>	🔊 schlägste		🔊 ähm		🔊 schä/	🔊 schlägst	🔊 du
<i>instructor_utt</i>	🔊 utt						
<i>instructor_df</i>	🔊 pr		🔊 f1				
<i>instructor_repair</i>	🔊 rd		🔊 ir		🔊 rs		
<i>instructor_subrep</i>					🔊 s1	🔊 s2	🔊 i1
<i>instructor_repair2</i>					🔊 rd	🔊 rs	
<i>instructor_subrep2</i>						🔊 s1	
<i>break</i>				🔊 1.5			
<i>len</i>	🔊 0.490	🔊 0.296	🔊 0.714	🔊 1.500	🔊 0.341	🔊 0.317	🔊 0.107

ing corpora that are publicly available see <https://korpling.german.hu-berlin.de/annis3>).

3. REPAIRS

Repairs consist of an original utterance (*reparandum*) and a replacement (*reparans*) [11]. Anything additional in between, e.g. filled or silent pauses, is called *interregnum*. Reparantia often subsume other disfluencies such as repetitions or truncations, as shown in Fig. 1 (following <https://u.hu-berlin.de/schlaegste> recreates the corpus query). The interregnum of the instructor is filled with *ähm* followed by a silent pause of 1.5 s. As L2 speakers usually need more processing capacity and working memory [12], we can predict that they should produce more filled pauses, as seen in [4]. Thus, learners should also use more repairs with interregna.

In a study about the structure of repairs of L1 and L2 German speakers in BeMaTaC, advanced learners with a competence level of C1 and above do not differ from native speakers in the frequencies of their repairs [2], cf. Table 1. The fact that they do not use more repairs with interregna than native speakers may indicate that either their planning capacity or their disfluency patterns are well adapted to the target language.

If there is no evidence for differences between L1 and L2 speakers in their repair frequencies, what is it that drives the appearance of interregna? A mixed-effects model in the study above suggests that repetitions and substitutions in the reparandum reduce the probability of interregna ($\beta = -.74$, $SE = .15$, $z = -4.78$, $p < .001$), as opposed to insertions. It

Table 1: Frequencies of repairs in BeMaTaC.

	N	per token
L1 repairs	245	0.0265
L2 repairs	261	0.0265

seems that the information structural processing of what can be seen as ‘new’ information is as challenging for non-natives as it is for native speakers. On the other hand, repeating or substituting what could be considered salient material tends to produce a higher amount of repairs without interregna, both for L1 and L2 speakers.

Thus, the same repair behaviour of the two groups might in fact be driven by constraints that apply language-internally and reflects the ease of repeating or modifying linguistic material as opposed to adding something new.

4. SUMMARY

We investigated the repair behaviour of L1 and advanced L2 speakers in a corpus of spontaneous speech. The results of this study imply that a native-like use of disfluency patterns can be achieved, at least for the number of repairs and within-repair behaviour. The resource employed for this study can easily be used and extended for future research in learner speech.

5. REFERENCES

- [1] Anderson, A. H., Bader, M., Bard, E. G., Boyle, E., Doherty, G., Garrod, S., Isard, S., Kowtko, J., McAllister, J., Miller, J., Sotillo, C., Thompson, H. S., Weinert, R. 1991. The HCRC Map Task Corpus. *Language and Speech* 34, 351–366.
- [2] Belz, M. 2013. Disfluencies and repairs in native and learner speech – a contrastive analysis. *Master's thesis*.
- [3] Bortfeld, H., Leon, S. D., Bloom, J. E., Schober, M. F., Brennan, S. E. 2001. Disfluency Rates in Conversation: Effects of Age, Relationship, Topic, Role, and Gender. *Language and Speech* 44, 123–147.
- [4] Brand, C., Götz, S. 2013. Fluency versus accuracy in advanced spoken learner language. In: Gilquin, G., Cock, S. d., (eds), *Errors and disfluencies in spoken corpora* volume 52 of *Benjamins Current Topics*. John Benjamins 117–137.
- [5] Eckes, T. 2010. Der Online-Einstufungstest Deutsch als Fremdsprache (onDaF): Theoretische Grundlagen, Konstruktion und Validierung. In: Grotjahn, R., (ed), *Der C-Test: Beiträge aus der aktuellen Forschung* volume 18 of *Language testing and evaluation*. Lang 125–192.
- [6] Gick, B., Wilson, I., Koch, K., Cook, C. 2004. Language-Specific Articulatory Settings: Evidence from Inter-Utterance Rest Position. *Phonetica* 61, 220–233.
- [7] Götz, S. 2013. *Fluency in Native and Nonnative English Speech* volume 53. John Benjamins Publishing Company.
- [8] Raupach, M. 1980. Temporal variables in first and second language speech production. In: Dechert, H. W., Raupach, M., (eds), *Temporal variables in speech*. Mouton 263–270.
- [9] Sauer, S., Lüdeling, A. to appear. Flexible multi-layer spoken dialogue corpora. *International Journal of Corpus Linguistics*.
- [10] Schegloff, E. A., Jefferson, G., Sacks, H. 1977. The preference for self-correction in the organization of repair in conversation. *Language* 53, 361–382.
- [11] Shriberg, E. E. 1994. Preliminaries to a Theory of Speech Disfluencies. *Unpublished dissertation*.
- [12] Temple, L. 2000. Second language learner speech production. *Studia linguistica* 54, 288–297.
- [13] Zeldes, A., Ritz, J., Lüdeling, A., Chiarcos, C. 2009. ANNIS: A Search Tool for Multi-Layer Annotated Corpora. *Proceedings of Corpus Linguistics*.

Notes: