HAL
open science

Pronunciation Dictionaries for the Alsatian Dialects to Analyze Spelling and Phonetic Variation

Lucie Steiblé, Delphine Bernhard

To cite this version:

Lucie Steiblé, Delphine Bernhard. Pronunciation Dictionaries for the Alsatian Dialects to Analyze Spelling and Phonetic Variation. 11th edition of the Language Resources and Evaluation Conference, 7-12 May 2018, May 2018, Miyazaki, Japan. hal-01704814

HAL Id: hal-01704814

https://hal.science/hal-01704814v1

Submitted on 20 Jun 2018

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Pronunciation Dictionaries for the Alsatian Dialects to Analyze Spelling and Phonetic Variation

Lucie Steiblé, Delphine Bernhard
LiLPa, Université de Strasbourg, France
lucie.steible@unistra.fr, dbernhard@unistra.fr

Abstract

This article presents new pronunciation dictionaries for the under-resourced Alsatian dialects, spoken in north-eastern France. These dictionaries are compared with existing phonetic transcriptions of Alsatian, German and French in order to analyze the relationship between speech and writing. The Alsatian dialects do not have a standardized spelling system, despite a literary history that goes back to the beginning of the 19th century. As a consequence, writers often use their own spelling systems, more or less based on German and often with some specifically French characters. But none of these systems can be seen as fully canonical. In this paper, we present the findings of an analysis of the spelling systems used in four different Alsatian datasets, including three newly transcribed lexicons, and describe how they differ by taking the phonetic transcriptions into account. We also detail experiments with a grapheme-to-phoneme (G2P) system trained on manually transcribed data and show that the combination of both spelling and phonetic variation presents specific challenges.

Keywords: Alsatian, transcription, spelling variation

1. Introduction

The Alsatian dialects, which belong to the High German dialects, are still spoken by approximately 500,000 speakers in Alsace, a region in north-eastern France (INSEE et al., 1999). Being non-dominant varieties, in a French-speaking country, they are mostly considered as *oral languages*, and have no standardized spelling system. There are however some occasions on which these dialects are written, dating back to 1816 when the first theater play in Alsatian, *Der Pfingstmontag* by Jean-Georges-Daniel Arnold, was published. Despite this written production, now stretching back to two centuries, almost no computational tools exist for the Alsatian dialects, which can therefore be considered as low-resourced.

Given the absence of a standardized spelling, this article aims at investigating the relations between oral and written forms. This work was performed using lexicographic and lexical resources: for one resource, we used the transcriptions which were already provided and for the other resources we manually transcribed recordings of native speakers illustrating the lexical entries. Letter sequences and phonemes were then automatically aligned. Finally, we evaluated the challenges for a grapheme-to-phoneme (G2P) tool facing both spelling and phonetic variation. Our ultimate objective is to provide resources for the Alsatian dialects, gain a better understanding of the phonological and spelling systems and develop tools able to deal with spelling variation in text corpora.

The main contributions of this article are as follows:

- We present the first pronunciation dictionaries for the Alsatian dialects, obtained by the manual transcription of several datasets ;
- We investigate phonetic and spelling variation in these datasets, with a focus on the specific status of consonants ;
- We compare various spelling systems for the Alsatian dialects, with a focus on their orthographic depth ;
- We train and evaluate a G2P system based on the manual transcriptions.

2. Related Work

Spelling variation is an issue for many different applications which have to process texts lacking spelling consistency (informal texts on the Web 2.0, text messaging, historical texts, dialects, etc.). While speech is rarely taken into account in traditional (written) text processing pipelines, it can be useful when dealing with spelling variation. Phonetic indexing algorithms, like Double Metaphone (Philips, 2000), have been developed in the context of information retrieval to account for spelling differences in words or names. The goal is to encode the input string using simplified phonetic rules. More sophisticated methods like grapheme-to-phoneme can also be used for searching words in corpora and dictionaries.[1] The goal is either to be able to retrieve words without knowing their exact spelling, or to abstract from the spelling variations in non-standardized languages, by using a phonetic index (Divay and Vitale, 1997).

In addition to improving the recall of queries, phonetic indexing and grapheme-to-phoneme techniques have been used to normalize texts in non-standard spellings. For instance, Cook and Stevenson (2009) take the phonemes of the standard word into account when normalizing graphemes in SMS text messages. Porta et al. (2013) integrate grapheme-to-phoneme transcription rules as well as rules expressing phonological change in a rule-based transducer from Old Spanish to Modern Spanish. These studies demonstrate that using knowledge about phonemes is relevant when dealing with non-standard spellings.

[1] See e.g. the *Trésor de la Langue Française* French dictionary (`http://atilf.atilf.fr/tlf.htm`) which uses pseudo-phonetic input or the Picartext textual database for the Picard language (`https://www.u-picardie.fr/LESCLaP/PICARTEXT/Public/`) which matches word forms based on phonetic correspondences.

In order to apply a grapheme-to-phoneme system to non-standard text a first obvious condition is to have such a system at disposal. In particular for low-resource languages, which also present high levels of spelling variation, the development of grapheme-to-phoneme systems is hindered by the lack of complete studies about the phonological system as well as the absence or the small volume of existing pronunciation dictionaries (word-pronunciation pairs). Moreover, when spelling is not normalized, designing grapheme-to-phoneme systems is known to be a complex task (Adda-Decker et al., 2011). The lack of resources can possibly be addressed by re-using data from other languages: Deri and Knight (2016) present an approach to train G2P models for low-resourced languages based on word-pronunciation pairs acquired from Wiktionary and on an adaptation of high-resource languages G2P models to closely related low-resource languages. This method nevertheless requires phoneme inventories to compute a phonetic distance metric between languages.

3. Datasets and Preprocessing

There are five main dialectal areas in Alsace, characterized by differences in their sound inventories: Rhine Franconian, South Franconian, High Alemannic, Low Alemannic from the north of the region, and Low Alemannic from the south of the region (Huck et al., 1999). The last two Low Alemannic variants are dominant on the Alsatian territory, and are the varieties studied here. They differ in several phonetic aspects, including the use of `[C]` in the south and `[X]` in the north after a front vowel, upholding of the `[g]` between vowels only in the south (in the north, the `[g]` became `[v]`) and more open vowels in final position in the south. All of them are German dialects, which creates a specific situation in this French region. According to a recent study, 43% of Alsatians can speak the dialect (OLCA / EDinstitut, 2012), and all of them also speak and write the national language, French.

The four datasets under study for the Alsatian dialects are the following:

- The DICTMULTI dataset is extracted from a printed multilingual dictionary (Adolf, 2006). It uses its own spelling system and provides the phonetic transcription for each word, made by the author of the dictionary. The whole dictionary is transcribed.

- The ELSASSICH dataset is taken from an online dictionary (*Elsässich Web diktionnair*) (Bitsch and Matzen, 2007). It uses its own spelling system, and the phonetic transcription was made manually by one of the authors of the present article, using the voice recordings provided for each lexical item. The whole online dictionary contains 3,333 entries, of which we transcribed 702.

- The OLCA datasets were produced by the local office for the preservation of Alsatian (Office pour la Langue et les Cultures d'Alsace et de Moselle, nd). It follows the ORTHAL spelling system (Zeidler and Crévenat-Werner, 2008), designed to be close to German spelling rules while at the same time preserving the variation (here between the northern part of the Alsace region –hereafter OLCA67– and the southern part –hereafter OLCA68). We used the same manual transcription process as for the ELSASSICH datasets, for a total of 2,859 entries, out of the 10,719 available.

The first dataset already had phonetic transcriptions, made by the dictionary author. Without audio files for the purpose of double checking, these transcriptions were used in their original form. The two last datasets, coming with audio, were phonetically transcribed by one of the authors, using the X-SAMPA transcription system (Wells, 1995). The resulting pronunciation dictionaries are available on the Zenodo platform (Steiblé, 2018) along with a documentation (Steiblé and Bernhard, 2018).

The written form was intentionally hidden during the transcription phase to avoid, as much as possible, the influence of the orthographic norms in use in the datasets. When there were doubts about some phonemes, specifically stops and their voice or voiceless quality, the waveforms were analyzed using PRAAT (Boersma and Weenink, nd).

Due to the high variability of the Alsatian dialects, no complete inventory of phonemes is available. We used, though, some area-specific inventories from sociolinguistic publications to help us (Zeidler and Crévenat-Werner, 2008; Huck et al., 1999; Bothorel-Witz et al., 1984).

Since the Alsatian dialects are in close contact with both French and Standard German, we also used datasets in these languages to draw a comparison between them and Alsatian. Our data for German come from two sources: MaryTTS (MARY) (Schröder and Trouvain, 2003), and Voxforge (VOX) (VoxForge, 2007). The French dataset is Lexique3 (LEX3) (New, 2006).

A quantitative summary of the datasets used in this study is provided in Table 1.

Language	Source	Entries	Phonemes
Alsatian	DICTMULTI	1,594	7,277
	ELSASSICH	702	4,372
	OLCA67	1,458	10,825
	OLCA68	1,401	10,472
German	MARY	26,233	218,669
	VOX	8,463	55,688
French	LEX3	125,733	834,011

Table 1: Summary of data resources

All word / phonetic transcription pairs were automatically aligned using the Phonetisaurus tool, a WFST-driven grapheme-to-phoneme framework (Novak et al., 2016). The alignment pairs link one or several letters (graphemes) with one phoneme. Figure 1 shows the alignment obtained automatically for the word *Spritzkànn* (watering can).

This alignment process allows us to observe the consistency between sounds and graphemes, in other words the possible differences between acoustical reality and the use of a matching letter or sequences of letters. All the analyses hereafter pertain to the 100 most frequent phoneme/grapheme pairs in each dataset (these pairs amount to 82 to 97% of all the pairs).

Figure 1: Example of a grapheme-phoneme alignment by Phonetisaurus. The phonetic transcription uses X-SAMPA.

4. Spelling and Phonetic Variations

4.1. Characters and Phonemes in Alsatian

Overall, there are 28 unique characters common to the spelling systems used in the four Alsatian datasets.[2] They correspond to 98.7% of the total character occurrences in all the datasets. Among the remaining 6 characters found,[3] *é*, *ï* and *x* are very rare and can be seen only in loanwords. *ë* and *ì* on the contrary are quite frequent. *ë*, pronounced `[E]`, is only used in DICTMULTI and *ì*, pronounced `[I]`, is used in DICTMULTI, OLCA67 and OLCA68.

There are 42 unique phonemes common to all the 4 Alsatian datasets.[4] These phonemes can be considered as the essential set required to describe the phonology of Alsatian dialects. They correspond to 95.6% of the total phoneme occurrences in all the datasets. The other phonemes are present in three, two or only one dataset.[5] These rarer phonemes are linked to the high variability seen in the Alsatian dialects, since some of them are allophones of the phonemes from the main list above: `B` (only in Low Alemannic from the south) and `b`, `C` (in Low Alemannic from the south) `X`, etc. Some of the variation is not related to the north/south geographical gap, but rather depends on individual variation: some speakers use `R`, and others `r`, for example. There is also a tendency found in some varieties from the south of the region to diphthong some vocalic sounds, which are very rare in other forms of Alsatian. Some phonemes are characteristic for French loanwords, such as the nasal vowels `[a_~]` and `[o_~]`.

4.2. Variation at the Sublexical Level

Based on the automatic alignments between graphemes and phonemes (see Section 3.), we first investigate the orthographic depth and consistency of each spelling system and then focus on the special case of consonants.

4.2.1. Orthographic Depth and Consistency

There are many ways to spell sounds. In alphabetical systems, a sort of ideal could be to have a straightforward relation between phonemes and characters. This optimal situation would create a direct relation between one phoneme and one character: always the same pronunciation for one character and always the same way to write one given phoneme. Such a system could be described as having a consistent, *shallow* orthography. On the opposite, when it is difficult to assess the pronunciation from the spelling, or when one phoneme can be written in various ways, the orthography is called *deep*.

Orthographic depth can be evaluated by counting the average number of pronunciations for one grapheme, and, in contrast, the average number of graphemes that can be used to spell one phoneme. These figures provide an accurate picture of the consistency between sounds and graphemes. We obtained those ratios by counting the average number of pronunciations for one given grapheme, and the number of ways to spell one given phoneme, for the 100 most frequent alignment pairs from our datasets. Of course, even when the phonetical transcriptions were made by one and the same person for the OLCA and ELSASSICH datasets, the writing system is not the same, leading to differences in the average numbers. Figure 2 displays the orthographic consistency for all our datasets, including German and French for the sake of comparison. The Alsatian dialects and German have quite similar ratios. Both can be described as having an almost "shallow orthography", or "transparent orthography". Compared to French, a language known for its complex spelling system, Alsatian is quite straightforward, despite the lack of a unified orthographic norm.

The most transparent spelling system in Alsatian is ELSASSICH. One possible explanation could be that the authors of this specific spelling system are linguists, which could account for a more stable relation between phonemes and graphemes. While these ratios are certainly a method to evaluate orthographic depth, other experiments should be performed to assess the degree of readability and usability of each system.

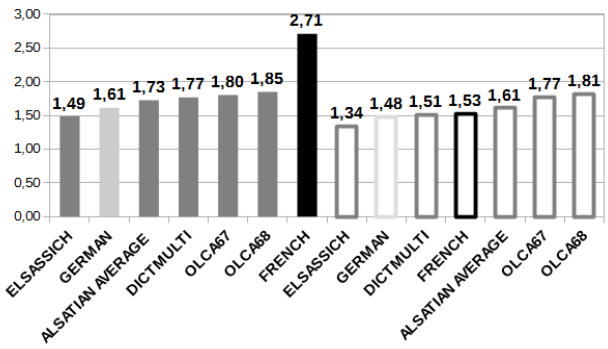

Figure 2: Average number of graphemes by phoneme, per language/spelling system (plain bars), and average number of pronunciations by grapheme (hollow bars).

4.2.2. Consonants and "Voicing"

The status of the consonants in Alsatian is quite unclear. In dictionaries, pronunciation tables are sometimes provided, but they are contradictory, and the consonants set differs from one to another. Fricatives are usually considered as always voiceless (Adolf, 2006) which leads to the lack of distinction between `[f]` and `[v]`, for example. The categorization of plosives is less clear.

In Alsatian, all plosives are voiceless (Steiblé, 2014). Despite this, there are two categories of plosives, opposed by the feature [*fortis*] (Jessen, 1998; Kohler, 1979). In writing, minimal pairs can be represented, using the same characters

[2] a à ä b c d e f g h i j k l m n o ö p q r s t u ü v w z

[3] ë é ï ì x y

[4] 9: 2 @ a a: ai b d e E e: E: f g h i I i: I: j k l m n N o O o: O: p r s S t u U u: v w X y y:.

[5] 2: 9 a_~ B Ei EI o_~ aI ao aU C R ts y9 Y z 3: ? @: 2I 2y 9i ei i@ ia Ia ie iE IE Oi U: Y: y9: ya

as in French or German, such as: *Gass* (street) versus *Kass* (crate). The difference between the two sets of plosives is linked to speech events, involving two distinct temporal patterns for the production of the consonants, as shown in Figure 3. The distinction between the series is based on various cues (Steiblé, 2014). The main cues are linked to the temporal orchestration of stops production. In both absolute and relative durations, the opposition is statistically significant. The Voice Termination Time (Agnello, 1975) is shorter in the *fortis* stops (13% of the total stop, versus 30% for the *lenis* stops), and the burst and Voice Onset Time (Klatt, 1975) phase is longer for the *fortis* stops (25% of the total stop, versus 17% for the *lenis*). The silent stop gap duration is longer in the *fortis* stops (63% of the total stop, versus 53% for the *lenis*).

During the transcription by one of the authors of the ELSASSICH, OLCA67 and OLCA68 datasets, if the consonant status was unclear, the recordings of the entries were observed on PRAAT (Boersma and Weenink, nd). If this visual check (searching for the patterns showed above) was not certain, the intra-segmental relative durations were compared to the values found in (Steiblé, 2014) for *fortis* and *lenis*, respectively. Despite the acoustical and written evidence, it has been said that the Alsatian dialects do not show any distinction between the consonants (Hug, nd), probably because they differ from both German and French sounds. In fact, our dataset tends to prove that there is enough distinction between the two series to allow for a strong phonological awareness, leading to the use of the two series of letters.

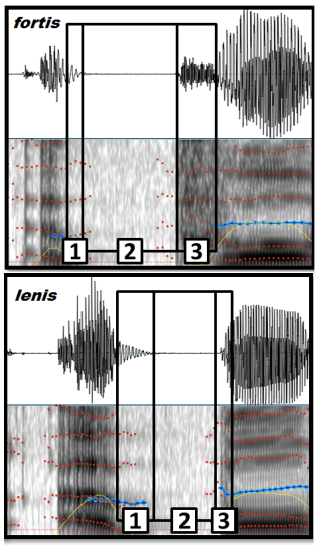

Figure 3: Typical temporal patterns of the Alsatian plosives: top = fortis plosive, bottom = lenis plosive. The rectangles frame three intra-segmental phases : 1, the Voice Termination Time, 2, the silent stop gap, and 3, the burst and Voice Onset Time.

In German, spelling rules would warrant that a voiceless consonant be written using the matching graphemes from the voiceless list (p,t,k,c,f,s,ss,ß,ch,sch,etc). In our German datasets, however, exceptions can be found and quantified, providing us with a consistency ratio. In Alsatian, admittedly, all consonants are voiceless, but the *fortis* ones are written using the same letters as the German voiceless set, and the *lenis* ones using the complementary voiced set (b,d,g,v,z,j,etc). The consistency ratio between a consonant and the characters used to spell it amounts to 95% in French (which accounts for the clear voiced/voiceless opposition in this language). In German, the ratio reaches 90%, and Alsatian reaches 89% (over all the four datasets), despite the lack of a unified orthographic norm. It is likely that the consistency ratio found in Alsatian is not related to the variation in spelling but to it being a German dialect.

The consistency ratio of Alsatian, when measured this way, is equivalent to that of standard German, despite the lack of a unified orthographic norm in Alsatian, as shown in Figure 4.

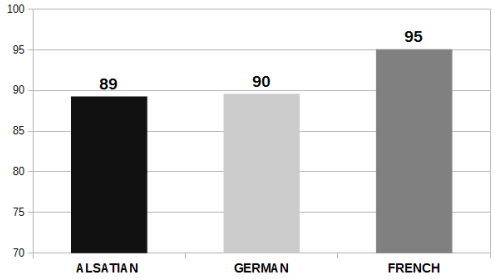

Figure 4: Percentage of consonant graphemes matching the phonemes, in terms of phonological oppositions

5. G2P for the Alsatian Dialects

In this section, we use the pronunciation dictionaries to train a G2P system. This presents a double challenge: (i) the small size of the pronunciation dictionaries and (ii) the presence of both phonetic and spelling variation. Given these challenges, our goal here is not to produce a full-fledged G2P system, but to assess the performance levels which can be reached for now, both on the four Alsatian datasets and on the task of retrieving spelling variants extracted from a corpus of texts in Alsatian.

For G2P, we chose to adopt a data-driven approach and trained the Phonetisaurus (Novak et al., 2016) system. The results are presented in Table 2 and are evaluated in terms of Word Error Rate (WER) – the percentage of lexicon entries with at least one error in their 1-best transcription– and Phoneme Error Rate (PER) – the Levenshtein distance between the predicted and the reference transcription, divided by the number of phonemes in the reference transcription (Hixon et al., 2011). We trained the G2P model on each dataset, then applied it to the other datasets. We also performed a closed test for each dataset by using the training data as test data. As could be expected, the results are low, given the small size of the training data. Training on the largest datasets (DICTMULTI and OLCA67) yielded the best results. The results obtained by training on OLCA67 and then applying the model to OLCA68 (and the other way round) could indicate that the use of the same spelling system plays a positive role. However, it should be mentioned that OLCA67 and OLCA68 also roughly correspond to the same vocabulary set, only for two different dialectal areas,

	DICTMULTI		ELSASSICH		OLCA67		OLCA68	
Training dataset	**WER**	**PER**	**WER**	**PER**	**WER**	**PER**	**WER**	**PER**
DICTMULTI	*11.70%*	*4.49%*	82.75%	29.83%	88.98%	35.49%	90.70%	38.02%
ELSASSICH	88.76 %	37.59%	*47.37%*	*19.04%*	96.12%	45.32%	97.14%	48.59%
OLCA67	71.59%	27.45%	85.67%	31.90%	*23.87%*	*7.18%*	66.52%	20.22%
OLCA68	72.33%	29.14%	85.67%	33.19%	67.02%	20.84%	*29.30%*	*9.52%*

Table 2: G2P results. The best results for each dataset have a grey background. The results of the closed tests are in italics.

and this could account for the results obtained. We also tried to train models using combinations of the datasets but this only leads to small improvements or no improvements at all, despite the increase in the amount of training data. This might indicate that the system is not able to handle what might often correspond to contradicting cues coming from heterogeneous datasets.

Finally, we applied the four Alsatian G2P models to a list of 110 words extracted from a corpus of texts in Alsatian written by several authors and corresponding to several dialectal areas and spelling systems. These words are grouped in 28 variant clusters, e.g., [*Frejndschaft ; Freundschaft ; Friindschàft ; Frindschàft ; Frìndschàft*] (friendship). Our goal here is to assess the ability of the G2P models to provide identical transcriptions for spelling variants. The model which performs best at this task is OLCA67, with an F-measure of 0.11. As could be expected, the precision is high (1.0), but the recall is very low (0.06), especially when compared with a rule-based Double Metaphone approach (Bernhard, 2014), which is less precise (0.47) but has a much better recall (0.90). Here however the results can be improved by pooling all the datasets for training the G2P model, leading to a better F-measure (0.15) due to an increase in recall (0.08), with only a very small decrease in precision (0.97). When the task does not require phonetic realism, but rather the ability to generalize over spelling variants, a non-homogeneous training dataset seems to be a better option. One research direction could be to add known spelling variants in the pronunciation dictionaries before training.

6. Conclusion

In this article, we have described four word-pronunciation datasets for the Alsatian dialects, covering different spelling systems. Our analysis of the relationship between sound and spelling has shown that we observe spelling and phonetic variation between the datasets, but that the spelling systems are rather self-consistent, having an almost shallow orthography. We have also identified a set of essential phonemes required to describe Alsatian phonology. In future work, we plan to increase the size of the pronunciation dictionary using a semi-automated approach relying on the G2P models described in Section 5.. The pronunciation dictionaries could be used in the future to create writing aids (spellchecking), or to assist dictionary lookup for spelling variants.

7. Acknowledgements

This work was supported by the French "Agence Nationale de la Recherche" (ANR) (project no.: ANR-14-CE24-0003). We would like to thank Paul Adolf for kindly providing an electronic copy of his dictionary.

8. Bibliographical References

Adda-Decker, M., Lamel, L., Adda, G., and Lavergne, T. (2011). A First LVCSR System for Luxembourgish, a Low-Resourced European Language. In *Language and Technology Conference*, pages 479–490. Springer.

Adolf, P. (2006). *Dictionnaire comparatif multilingue: français-allemand-alsacien-anglais*. Midgard, Strasbourg, France.

Agnello, J. (1975). Voice onset and voice termination features of stutterers. In L.M Webster et al., editors, *Vocal tract dynamics and dysfluency: the proceedings of the first annual Hayes Martin Conference on Vocal Tract Dynamics.*, New-York. Speech and Hearing Institute.

Bernhard, D. (2014). Adding Dialectal Lexicalisations to Linked Open Data Resources: the Example of Alsatian. In *Proceedings of the Workshop on "Collaboration and Computing for Under-Resourced Languages in the Linked Open Data Era" at LREC 2014*, pages 23–29, Reykjavík, Islande.

Bitsch, R. and Matzen, R. (2007). de Elsässich Web diktionnair. Online: `http://www.ami-hebdo.com/elsadico/index.php`.

Boersma, P. and Weenink, D. (nd). Praat, a system for doing phonetics by computer [Computer program]. Downloaded from: `http://www.praat.org`.

Bothorel-Witz, A., Philipp, M., and Spindler, S. (1984). *Atlas linguistique et ethnographique de l'Alsace*. CNRS.

Cook, P. and Stevenson, S. (2009). An Unsupervised Model for Text Message Normalization. In *Proceedings of the Workshop on Computational Approaches to Linguistic Creativity*, CALC '09, pages 71–78.

Deri, A. and Knight, K. (2016). Grapheme-to-phoneme models for (almost) any language. In *Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics*, volume 1, pages 399–408.

Divay, M. and Vitale, A. J. (1997). Algorithms for grapheme-phoneme translation for English and French: Applications for database searches and speech synthesis. *Computational linguistics*, 23(4):495–523.

Hixon, B., Schneider, E., and Epstein, S. L. (2011). Phonemic Similarity Metrics to Compare Pronunciation Methods. In *INTERSPEECH*, pages 825–828.

Huck, D., Laugel, A., and Laugner, M. (1999). *L'élève dialectophone en Alsace et ses langues*. Oberlin, Strasbourg, France.

Hug, M. (nd). Orthographe alsacienne. Online: `http://elsasser.free.fr`.

INSEE, Barre, C., and Vanderschelden, M. (1999). Insee - Population - L'enquête "Etude de l'histoire familiale" de 1999.

Jessen, M. (1998). *Phonetics and Phonology of Tense and Lax Obstruents in German*. Studies in Functional and Structural Linguistics. John Benjamins Publishing Company, Amsterdam, jan.

Klatt, D. (1975). Voice onset time, frication, and aspiration in word-initial consonant clusters. *Journal of speech and hearing research*, 18(4):686–706.

Kohler, K. (1979). Phonetic Explanation in Phonology: The Feature Fortis/Lenis. *Phonetica*, 36:332–343.

New, B. (2006). Lexique 3: Une nouvelle base de données lexicales. In *Actes de la Conférence Traitement Automatique des Langues Naturelles (TALN 2006)*.

Novak, J. R., Minematsu, N., and Hirose, K. (2016). Phonetisaurus: Exploring grapheme-to-phoneme conversion with joint n-gram models in the WFST framework. *Natural Language Engineering*, 22(06):907–938, nov.

Office pour la Langue et les Cultures d'Alsace et de Moselle. (nd). Olca, lexiques français-alsacien. Online: `https://www.olcalsace.org/fr/lexiques`.

OLCA / EDinstitut. (2012). Etude sur le dialecte alsacien. Online: `https://www.olcalsace.org/sites/default/files/documents/etude_linguistique_olca_edinstitut.pdf`.

Philips, L. (2000). The Double Metaphone Search Algorithm. *C/C++ Users Journal*, 18(6):38–43, June.

Porta, J., Sancho, J.-L., and Gómez, J. (2013). Edit Transducers for Spelling Variation in Old Spanish. In *Proceedings of the Workshop on Computational Historical Linguistics at NoDaLiDa 2013*, volume 87 of *Linköping Electronic Conference Proceedings*, pages 70–79.

Schröder, M. and Trouvain, J. (2003). The German Text-to-Speech Synthesis System MARY: A Tool for Research, Development and Teaching. *International Journal of Speech Technology*, 6:365–377.

Steiblé, L. (2014). *Le contrôle temporel des consonnes occlusives de l'alsacien et du français parlé en Alsace*. Phd thesis, Université de Strasbourg.

Steiblé, L. and Bernhard, D., (2018). *Phonetic Transcription for the Alsatian Dialects*. DOI: 10.5281/zenodo.1174219.

VoxForge. (2007). Free Speech... Recognition (Linux, Windows and Mac) - voxforge.org.

Wells, J. (1995). Computer-coding the IPA: a proposed extension of SAMPA. Online: `https://www.phon.ucl.ac.uk/home/sampa/ipasam-x.pdf`.

Zeidler, E. and Crévenat-Werner, D. (2008). *Orthographe alsacienne: bien écrire l'alsacien de Wissembourg à Ferrette*. J. Do Bentzinger, Colmar, France.

9. Language Resource References

Steiblé, L. (2018). Pronunciation Dictionaries for the Alsatian Dialects. DOI: 10.5281/zenodo.1174214.