Proceedings Fourth International Conference on Language Resources and Evaluation (LREC 2004), Lisboa, 1395–1398.

CrossTowns: Automatically Generated Phonetic Lexicons of Cross-Lingual Pronunciation Variants of European City Names

Stefan Schaden

Institut für Kommunikationsakustik (IKA)
Ruhr-Universität Bochum
D-44780 Bochum, Germany
stefan.schaden@rub.de

Abstract

The *CrossTowns* lexicons are part of a study that focuses on the phonetic variants that occur when speakers of different native languages (L1) with varying degrees of target language (L2) proficiency pronounce foreign city names. Based on a collection of speech data from this domain, it is one of the aims to identify the most common pronunciation errors in a particular L1/L2 pair (*language direction*) and to model them by phonological rewrite rules. Although derived from only a small corpus of names, the rule sets already generate plausible variants when applied to unseen material. Yet there is a need for improvement. To demonstrate the current state of affairs, sample lexicons of 1.000 place names for English, French, and German were compiled and converted into various *interlanguage* pronunciation lexicons using the accent rule sets. In the paper, the procedures involved in the data collection, an outline of the rule-based accent generation technique, and a discussion of the problems involved in modelling non-native pronunciations on the lexicon level will be presented.

1. Introduction

The pronunciation of place names by non-native speakers is a problematic issue in speech technology applications such as travel information or car navigation systems: In this application scenario, a broad spectrum of potential mispronunciations and pronunciation variants must be anticipated, ranging from minor phonetic shifts to strongly accented forms that hardly resemble the canonical forms provided in ASR and TTS phonetic dictionaries. In ASR, severe mismatches between expected pronunciation and actual pronunciation may lead to a significant decrease in recognition performance (cf. van Compernolle, 1999). In current speech synthesis research, on the other hand, naturalness, acceptability, and personalised voices are topical issues which might be addressed e.g. by modelling accents that are similar to the user's own speech style (cf. Dahlbäck et al., 2001).

The *CrossTowns* pronunciation lexicons are part of a research project that attempts not only to identify the most common non-native pronunciation errors occurring in various language directions (i.e. L1/L2 pairs), but also to model these variants on the lexicon level by applying phonological rule sets that systematically introduce selected pronunciation errors into canonical lexicons (cf. Schaden, 2003). The language-specific rule sets are designed to model varying degrees of partial L2 knowledge that speakers typically apply when pronouncing L2 material. Rule sets of this type have been compiled for various language directions and are constantly being updated and improved.

The present contribution is a model application for this rule-based approach applied to unseen vobabulary. The rules used for the *CrossTowns* lexicons are based on non-native speech data that was compiled for this specific purpose (Schaden, 2002). This database includes place names from a number of European languages (English, German, French, Italian, and Dutch) that were pronounced by speakers of different native languages, and thus represents potential pronunciation variants of numerous L1/L2 combinations.

It is the aim of this contribution not only to demonstrate the approach pursued in this study by introducing (freely downloadable) sample lexicons, but also to stimulate feedback – especially by native speakers of the languages investigated – in order to improve the underlying rule sets as well as the overall approach where necessary.

The topic addressed in this study is not entirely new. Similar research – though on a considerably larger scale – has been carried out within the framework of the European *Onomastica* project. In particular, the so-called *Onomastica Interlanguage Pronunciation Lexicon* (cf. Onomastica Consortium, 1995) is comparable to the *CrossTowns* project in many respects. However, the data compiled within this framework has unfortunately never been made available to the research community.

Despite these parallels, it should be stressed that it is not the ultimate aim of this study to just list potential pronunciation variants in a static lexicon. Rather, it is attempted to describe the linguistic regularities involved in the most common non-native pronunciation errors in such a way that this knowledge can be used to systematically reproduce these mispronunciations for new vocabulary. This study pursues the approach that non-native pronunciation errors are motivated by linguistically traceable principles, and that these principles can be described on a generic level to be employed in speech-based systems.

2. Languages and Base Lexicons

2.1 Language Directions

The languages included in this study so far are English (ENG), German (GER), French (FRA), Italian (ITA), Dutch (DT), and Spanish (SPA) in various L1/L2 combinations. Since language-specific influences of the speakers' L1s typically lead to different types of phonetic errors, each language direction is regarded as a separate unit

of analysis. Although in some cases speakers transfer L1-specific features equally to several L2s, most errors are specific of a particular L1/L2 pair. Therefore it is required to establish a separate *CrossTowns* lexicon for each of them. The current language directions are (in the notation L1 → L2):

CT 1: *English → German*	**CT 2**: *German → English*
CT 3: *French → German*	**CT 3**: *German → French*
	CT 4: *German → Italian*

Table 1: Existing *CrossTowns* lexicons (as of Apr 2004)

Lexicons for further language directions involving e.g. Italian and Spanish as L1 are scheduled for the near future. However, the general design of all lexicons will follow the exemplary design described in this paper.

2.2 Input Lexicons

2.2.1 Selection of Names

For each target language, a sample lexicon of 1.000 entries was compiled. The orthographic input lexicons are random selections from the *GEOnet* place names database. This data compilation, which is accessible on the Web[1], covers a huge number of place names[2] from more than 50 countries worldwide. Although the database clearly reflects its primarily geographical purposes, it can also be viewed as a valuable linguistic corpus for studies in the domain of geographical names.

At the present stage, the *CrossTowns* lexicons focus on rather small and unknown place names (examples see Table 2 below), whose pronunciations by non-natives is less predictable than in the case of well-known cities, where speakers may apply various sources of previous knowledge about their pronunciation. The *GEOnet* data supports this selection criterion in that some of the country databases contain specificiations of the size and geopolitical importance of towns. This information can be used by explicitly specifying this feature in the selection.

English	*Appleton, Bangor, Bridgwater, Kennington, Longborough, Maidstone, Warminster*
German	*Barsinghausen, Drakenburg, Eichstetten, Rosenheim, Schwalmtal, Sigmaringen*
French	*Abbeville, Beaubray, Bésignan, Cavaillon, Longchamp, Prévenchères, Rambouillet*

Table 2: Examples of names in the input lexicons

As a side-effect, a restriction to unknown names will also rule out cities for which there is an *exonym* in a particular L2 (e.g. *Londres* for *London*, *Munich* for *München*). In these particular cases – which mainly occur for large or well-known cities –, it is reasonable to include the L1-specific exonym in the list of potential pronunciation variants in addition to the automatically generated variants.

[1] http://www.nima.mil/gns/html

[2] Approx. 5.45 million as of Febr 2004

2.2.2 Phonetic Transcriptions

Within the domain of place names, devising a canonical phonetic transcription is not always a straightforward task. Even if a large subset of a country's toponyms roughly follows the regular pronunciation rules of the dominant language, there are still numerous exceptions. Place names may reflect older historical stages of the language and therefore preserve archaic grapheme-to-phoneme relations; they may have distinct linguistic origins and exhibit orthographic and phonetic traces of their source language (e.g. Welsh names in the UK), or their pronunciation may vary even within the borders of one language area, which makes it hard to stipulate a 'canonical pronunciation' for them at all. Thus a particular degree of idealisation seems indispensable in a study of this type. Therefore, the standard transcriptions that were adopted in the *CrossTowns* lexicons by and large follow the pronunciation rules of the relevant language, supplemented by some well-known particularities of toponymic pronunciation.

For the reference transcription of the lexicon entries, the standard SAMPA inventory for the corresponding languages is used (Wells, 2003) [3]. While the phonetic transcriptions for German could be extracted from an existing pronunciation dictionary, the English and French pronuncations had to be generated from scratch. To this aim, initial transcriptions of the names were generated by English and French GTP converters, followed by manual revision and correction by phonetically trained transcribers. These lexicons, containing an orthographical and phonetic transcription, were used as the input for the accent rule sets.

In order to characterise non-native pronunciation variants with sufficient phonetic detail in the word-level transcriptions, standard SAMPA is insufficient. In a cross-lingual situation as discussed here, speakers will typically produce a mixture of native, foreign, and intermediate speech sounds. It is therefore necessary to extend the basic language-specific inventories and to interpret the symbols in terms of language-independent phonetic values instead of language-specific phonemic values. Although principally designed as a language-independent phonetic alphabet, SAMPA symbols are regularly interpreted in terms of their language-specific phonemic values. For instance, the symbol /r/ is used for the 'r' sound in both English and Italian, but phonetically represents an approximant [ɹ] for English and a trill [r] for Italian. Therefore the X-SAMPA set proposed by Wells (2003) is applied where standard SAMPA would cause ambiguities.

3. Generating Variants by Rules

3.1 Deriving Rules from Speech Data

The majority of rules applied to generate the *CrossTowns* lexicons is derived from speech data that was compiled for this particular purpose. The data collection comprises at least 20 native speakers of ENG, GER, FRA, ITA, SPA with varying proficiency levels of the individual L2s. The speakers pronounced city names of five European countries in (i) a reading task and (ii) a perception/repetition

[3] In this paper, IPA transcriptions are used.

task (for details cf. Schaden, 2002). In these experiments, 45 names from each target language were used; the material is thus not identical to the lexicons presented here. Yet a restriction to a relatively small set of names was necessary to effectively conduct this study, as it includes a manual phonetic transcription of the recorded speech material.

Rather, this study is based on the central assumption that characteristic pronunciation errors, even if derived from only a restricted sample vocabulary, can be extrapolated to unseen vocabulary, provided that they have occurred with a relatively high inter-speaker consistency. Among other things, it is this assumption that is to be tested by the application of the rules to the unseen 1.000 entries of the *CrossTowns* lexicons. Presently, the average number of rules per language direction is 80–100. The application of these rule sets to new vocabulary is likely to provide valuable indication of required improvements, additions, and modifications of the individual rules.

3.2 Variation and Accent Gradation

Non-native accents are similar to local or regional dialectal pronunciation variants in that they represent a *phonetic deviation from the standard variety.* Yet there is a crucial difference: While in dialectal speech, deviations from the standard variety are relatively consistent for large speaker groups, foreign-accented pronunciations will always vary considerably according to individual speaker characteristics such as L2 proficiency, age, education, and many other potential influences. Non-native pronunciations are to a much lesser extent shaped by a *regular correspondence* to the standard variety. Yet looking closely at a speaker group of the same L1 background pronouncing material of a particular L2, a number of inter-speaker regularities can be identified, especially for speakers at comparable proficiency levels. This situation can basically be described as a *graded continuum* of potential mispronunciations – in the sense of 'interlanguages' (Selinker, 1972) – ranging from slightly accented forms with only minor allophonic shifts up to strongly accented pronunciations with extreme deviations from the L2 standard. Therefore it is inadequate to model variants for a particular L1/L2 combination by adding just *one* single L1-specific variant to each L2 lexicon item. On the other hand, though, it is neither a practical aim to take *all* potential variants into account. In the present approach, it is therefore suggested to break up the continuum into discrete categories by defining a number of prototypical foreign-accented pronunciations per word, where each of these prototypes represents a particular *accent level*.

Accent levels range from near-native pronunciation to gross mispronunciations. Currently, this accent gradation model is based on four levels $0 < N < 4$, where 0 marks the canonical L2 pronunciation and higher integers indicate increasing deviations from the canonical form. The topmost level 4 is a strongly accented pronunciation that follows almost completely the grapheme-phoneme correspondences of the speaker's L1. Table 3 illustrates this conception for the English town name *Winchester* and its corresponding accent gradation for native speakers of German:

Item: ***Winchester***	**Accent gradation**
Level 0 (canonical)	[w'ɪntʃɪstə]
Level 1	[w'ɪntʃ**ɛ**st**ɐ**]
Level 2	[w'ɪn**ʃ**ɛstɐ]
Level 3	[**v**'ɪnʃɛstɐ]
Level 4	[v'ɪn**ç**ɛstɐ]

Table 3: Rule-generated accent gradation; English *Winchester* for L1 German

Accordingly, the rule system is built up in such a way that for each input word, multiple variants representing the accent level prototypes can be generated. The output is a modified dictionary containing *N* pronunciation variants per word, where *N* is the number of accent levels as defined above.

Based on this conception, multiple pronunications that reflect varying L2 proficiency levels of L1 speakers are derived for each lexicon entry. In the present stage, four prototypical accent levels plus the canonical form are distinguished; hence each *CrossTowns* lexicon contains a total of 5.000 pronunciations.

3.3 Rule Format and Generation of Lexicons

The basic rule format as well as a number of typical applications of the rules is outlined in Schaden (2003) and can only be roughly sketched in the present context. Generally, all rules operate as phonetic substitution rules using the notation adopted from generative phonology. Here, an L2 sound X_{L2} is substituted a sound Y if the immediate left and right contexts LC and RC are valid:

$$X_{L2} \rightarrow Y \quad / \quad LC \_\_ RC$$

While this general rule format is well established and widely used, there is one essential feature that distinguishes the rules applied in this study from traditional rewrite rules: Since many pronunciation errors in non-native speech are mediated or triggered by orthography rather than being purely phonetic interference phenomena, the rules include the graphemic representation of words and make use of this information to model a number of errors. This proved to be a convenient technique to model e.g. non-native mispronunciations caused by a transfer of L1 letter-to-sound rules onto the target language.

The input lexicons require only a minimum of linguistic annotation. Plain pronunciation dictionaries containing only an orthographic word and its canonical transcription (phonemic or as surface allophones) suffice as input for the rule system. By keeping the requirements for input lexicons at this minimum level, the rule system can be applied to existing phonetic dictionaries without the need for introducing specific annotation schemes.

However, in order to take advantage of the above mentioned graphemic information, an *alignment* of the grapheme and phoneme sequences in the input lexicon is a necessary precondition. This procedure maps each phone in the input string to the grapheme or grapheme sequence

that represents it, as illustrated in the following example of the French name *Questembert*:

qu	e	s	t	em	b	e	r	t
|	|	|	|	|	|	|	|	|
k	ɛ	s	t	ã	b	ɛ	ʁ	–

Fig. 1: GP alignment for the French name *Questembert*

The GP alignment must be applied prior to the accent rules, since it is required for the graphemically constrained rules to operate properly. The alignment module is based on a set of language-specific rules containing all potential graphemic representations of each phoneme of the language. The overall rule system is designed to operate *postlexically*. This means that virtually any existing canonical phonetic lexicon can be converted into an adapted dictionary for specific non-native speaker groups without interfering with the original input lexicon.

4. Preliminary Evaluation

In an evaluation of a rule system that is designed to model pronunciation *errors*, it is not a straightfoward task to distinguish 'correct' vs. 'erroneous' output of the system. While rules designed to generate canonical transcriptions can be checked against a correctly transcribed reference form in order to evaluate their performance, it is not particularly clear what the appropriate target forms are in the case of non-native speech. In any case, the notion of a 'correct' transcription is probably not adequate as a reference for the evaluation of the system. Yet the automatically generated variants can be assessed in terms of (a) *plausibility* and (b) *coverage* when measured against the actual pronunciations of a reference speaker group.

In order to establish the basic rule sets that were used to generate the *CrossTowns* lexicons, manual phonetic transcriptions (approx. 20.000) of non-native pronunciations were created. Although this manually transcribed vocabulary is not identical to the entries of the *CrossTowns* lexicons, it may well be used for an exemplary evaluation of the overall approach. Such an evaluation procedure is presently being elaborated and has already yielded some encouraging preliminary results. It is the basic idea to compare the automatically generated variants of a particular input lexicon to actual speaker variants for the very same vocabulary in order to determine the degree of *phonetic approximation* achieved by the rule-based variants. This approximation, calculated by a phonetically weighted distance metric, should be above that of the canonical transcription of the lexicon item. In a first informal evaluation for a limited vocabulary, at least one of the automatically generated variants achieved a better approximation than the canonical form in about 50% of all cases.

5. Conclusions

The approach presented in this paper is a sample application of a method of handling a particular type of pronunciation variation. Although a high degree of variability must be expected in all non-native pronunciations, it is hoped that a number of characteristic errors can be realistcally formulated by rule sets. Among others, name pronunciation is just one potential application of the rules. Future work will include not only a constant update and improvement of the underlying rules, but also an extension to new language directions as well as an application of the rules to standard vocabulary.

The *CrossTowns* sample lexicons are accessible online for all interested parties. Any kind of feedback is welcome as a valuable contribution to the improvement of the underlying rule sets that generated the lexicons.

Note: Since the web interface of the *CrossTowns* lexicons is not yet set up at the time of the paper submission, please contact the author for details.

Acknowledgements

This study is funded by the *Deutsche Forschungsgemeinschaft* (DFG). The author would like to thank Ms Katja Kesselmeier for her invaluable support in the phonetic transcription and Mr Anders Krosch for carrying out countless speech recordings.

References

Dahlbäck, N.; Swamy, S.; Nass, C.; Arvidsson, F.; Skågeby, J. (2001): "Spoken Interaction with Computers in a Native or Non-native Language – Same or Different?" Proceedings *INTERACT* 2001.

Onomastica Consortium (1995): "The Onomastica Interlanguage Pronunciation Lexicon". *Proceedings Eurospeech 1995*, Madrid, Spain, 829–832.

Schaden, S. (2002): "A Database for the Analysis of Cross-Lingual Pronunciation Variants of European City Names". *Proceedings Third International Conference on Language Resources and Evaluation (LREC 2002)*, Las Palmas de Gran Canaria, Spain, Vol. 4, 1277-1283.

Schaden, S. (2003): "Generating Non-Native Pronunciation Lexicons by Phonological Rules". *Proceedings 15th International Conference of Phonetic Sciences (ICPhS 2003)*, Barcelona, Spain.

Selinker, L. (1972): "Interlanguage". *International Review of Applied Linguistics* (IRAL) 10, 209-231.

van Compernolle, D. (1999): "Speech Recognition by Goats, Wolves, Sheep and Non-Natives." *Proceedings Workshop on Interoperability in Speech Technology.* Leusden, The Netherlands.

Wells, J. C. (2003): *SAMPA. Computer Readable Phonetic Alphabet*. http://www.phon.ucl.ac.uk/home/sampa/home.htm