Generating Non-Native Pronunciation Lexicons by Phonological Rules

Stefan Schaden
Institute of Communication Acoustics, Ruhr-University Bochum
D-44780 Bochum, Germany
E-mail: schaden@ika.rub.de

ABSTRACT

This paper presents a new approach to model prototypical foreign-accented pronunciation variants on the phonetic transcription level using rewrite rules. For each native language (L1) and target language (L2) pair, a set of post-lexical rules is designed to transform canonical phonetic dictionaries of L2 into adapted dictionaries for native L1 speakers. Potential applications are speech recognition and speech synthesis systems. A brief outline of the general framework is given, followed by an overview of the different rule types included in the system.

1 INTRODUCTION

Machine-readable pronunciation lexicons are a central component of speech synthesis and speech recognition systems, as they form the link between the acoustic and symbolic level of automatic speech processing. Typically, each entry in a lexicon is assigned a phonetic transcription of the standard pronunciation in the language the system is designed for – the 'canonical form'. Canonical lexicons, however, have the general drawback that every marked deviation from the standard will not be represented. In recent years, a number of approaches have been proposed to compensate for this mismatch between canonical transcription and actual pronunciations by various *lexical adaptation* techniques (for an overview see [7]). These techniques are usually applied to model common intra-lingual stylistic variations such as within-word or cross-word assimilations and elisions in informal speech.

It is the aim of this study to extend the lexicon adaptation approach from intra-lingual variation to the domain of foreign-accented pronunciation. Non-native speakers frequently produce variants that deviate markedly from the canonical form. They are characterized by phenomena such as changes in allophonic realizations, phoneme shifts, word stress shifts, or alternations in syllable structure caused by epenthesis or deletion of speech sounds. The idea is to model these errors by lexicon adaptation, based on the assumption that for each language direction – i.e. a pair of a native language (L1) and a target language (L2) – a number of characteristic pronunciation errors can be identified. Although there is always a considerable range of inter-individual variation even for speakers with the same native language background, it is assumed that some common mispronunciations can be formulated as rewrite rules to generate prototypical 'interlanguage' transcriptions.

Currently, the languages investigated are German (GER), English (ENG), and French (FR) in different L1/L2 combinations. The general approach, however, is not restricted to these languages, since it can be transferred to other L1s or L2s. A prototype of a task-specific rule interpreter was implemented, and phonological rule sets for the language directions ENG $\rightarrow$ GER, GER $\rightarrow$ FR, GER $\rightarrow$ ENG, and FR $\rightarrow$ GER were developed. The rules are based on pronunciation variants observed in a non-native speech database [3]. They are currently limited to the domain of foreign city names; yet it is expected that many findings can be generalized to other lexical domains. While the general design of the rule system has already been outlined in [4], this paper gives a more detailed account of the rule types required to model some characteristics of foreign-accented pronunciation variants.

2 OUTLINE OF THE RULE SYSTEM

As a general prerequisite for modelling pronunciation variation of any kind – be it speaker-specific, dialectal, or foreign-accented –, knowledge about the target forms to be modelled is necessary. In the case of non-native speech, however, it is not obvious how to define these target forms: While e.g. in dialectal speech, phoneme shifts and other deviations from the standard 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, or even dialectal origin. A common native language background does by no means constitute a homogenuous non-native speaker group. Neither are the pronunciation variants of non-natives predictable, e.g. by comparing and contrasting the L1 and L2 phoneme systems, as was claimed by the 'strong contrastive hypothesis' of the 1960s. Rather, there is a continuum of potential mispronunciations – in the sense of 'interlanguages' defined by Selinker [6] – ranging from slightly accented forms with only minor allophonic shifts up to heavily accented pronunciations with extreme deviations from the L2 standard.

It is therefore not adequate to model variants for a particular L1/L2 combination by adding one single prototypical L1-specific variant for each L2 lexicon item. On the other hand, it is neither a practical aim to take *all* potential variants into account. Instead, in order to achieve an adequate coverage of potential variants, it is 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*. The general idea of using accent levels as a descriptive device is based on the concept of 'interference gradation' (*Interferenzstufung*) sketched in [8]. Accent levels range from near-native pronunciation to gross mispronunciations. 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. Currently, the model is based on four accent 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.

The overall rule system is designed to operate *postlexically*. This means that 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. For each lexicon entry, multiple modified transcriptions can be generated. The output of the rule system is a modified dictionary containing *N* pronunciation variants per word, where *N* is the number of accent levels as defined above.

3 RULE TYPES

3.1 CONTEXT-FREE PHONE MAPPING

It is one of the most salient characteristics of foreign-accented pronunciation that non-native speakers tend to substitute L2 speech sounds by similar, yet not identical L1 equivalents or by intermediate forms that can neither be assigned to the L1 or the L2 sound inventory.

The first idea that suggests itself for modelling these substitutions are phone mapping tables[1] that replace particular L2 sounds by similar speech sounds from the L1 inventory. The following example shows a fragment of a mapping table for the language direction German (L1) → English (L2):

English phone	German substitution phone
[ɹ]	[ʁ]
[θ]	[s]
[æ]	[ɛ]
[əʊ]	[oː]

Table 1: Mapping of English phones to German phones

But although the general idea of mapping L2 phones onto L1 phones was adopted for the presented rule system, simple context-free mapping is problematic in at least two respects: First, for many L2 sounds it is not clear what the 'best' L1 equivalent is. Acoustic or articulatory proximity of an L1/L2 phone pair is not a reliable predictor of the sound shifts that speakers actually produce. Secondly, our data clearly indicates that in many cases, the choice of the substitution phone is related to the phonetic or – in the case of read speech – graphemic surroundings of the substituted phone. Therefore, in order to restrict their application to appropriate contexts, most rewrite rules require context conditions or constraints on the phone level and on the orthographic level. The types of constraints are described in the following sections.

3.2 CONTEXT-DEPENDENT MAPPING

Rule contexts that do not require information from linguistic levels other than the phone level can be formulated using the established rule notation adopted from generative phonology:

$$X_{L2} \rightarrow Y_{L1} \quad / \quad LC \;\_\_\; RC$$

Here, a phone X_{L2} (element of the L2 inventory) is substituted by Y_{L1} (element of the L1 inventory) if the immediate left and right contexts LC and RC are valid. In the rule system presented here, X and Y are usually phoneme or allophone segments, although it is also possible to apply rules to phoneme or allophone clusters.

In addition to context-dependent mapping rules, there is a special type of rule that is based on language-specific phonological properties and their transfer to L2: It has been noted by phonologists that some characteristic nuances of foreign accents are caused by a projection of systematically occurring phonological rules of L1 onto L2. Kenstowicz [2] draws a distinction between 'distinctive' and 'redundant' phonological properties of a language and argues that speakers tend to transfer the redundant features of their L1 unconsciously to L2. Well-documented examples are the transfer of word-final obstruent devoicing onto various L2s by native speakers of German, or the transfer of full vowel reduction onto L2 by native English speakers. Phonological rules of this category typically apply to entire phoneme classes that undergo similar alternations in particular phonetic/phonemic contexts.

More often than not, however, the decision whether a rule is a systematic phonological rule of L1 depends on the theoretical framework that is adopted. Yet the concept of systematic phonological rules and their transfer to L2 proved to be useful to formalize and implement some characteristic accent features. For instance, some native speakers of French tend to apply vowel nasalization to German words, e.g. in *Blankenstein* *[blãkənʃtaɪn] or *Frankfurt* *[fʁãkfʊʁt] [2]. This can be interpreted as a transfer of a two-stage phonological rule of French, formalized by Schane [5] as follows: (1) A vowel becomes nasalized if it precedes a nasal consonant plus a word boundary or an additional consonant; (2) the nasal consonant that triggered rule (1) is deleted:

[1] As a terminological convenience, I will use the terms 'phone' or 'sound' although some 'phones' may in fact have phonemic status in L1 and/or L2.

[2] In this paper, all phonetic transcriptions of potential non-native pronunciation variants are marked with an asterisk * .

(1) V [-nasal] → V [+nasal] / __ C [+nasal] { ## , C }

(2) C [+nasal] → Ø / V [+nasal] __

Technically, the transfer of native phonological rules onto L2 is just a special case of context-dependent phone mapping. Rules of this type were implemented in the rule sets for all three languages. In some cases, it is possible to apply L1-specific phonological rules immediately to the L2 surface representations (e.g. French vowel nazalization applied to German), in other cases, additional linguistic information is required. This information can partly be drawn from the orthographic representation, as shown in the following sections.

3.3 GRAPHEMIC CONSTRAINTS

In the particular case of read speech, as well as in cases where the speaker's acquisition of L2 pronunciation rules was mediated through orthography, mispronunciations by non-natives are often triggered by a partial or full transfer of L1 grapheme-to-phoneme (GTP) correspondences to L2 (spelling pronunciation errors).

For automatic speech recognition in the proper names domain, it has been suggested to model this error type by an application of L1 grapheme-to-phoneme converters to L2 orthographic input [1]. But although this approach proved to be beneficial in this recognition scenario, it does not model speaker behavior adequately. Non-native pronunciation variants are rarely based on unmodified L1 GTP rules applied to L2, since many speakers have an awareness of at least some pronunciation rules of L2 (e.g. the pronunciation of German <sch> as [ʃ]). Moreover, for some L2 orthographic sequences, a straight transfer of L1 GTP rules would yield artificial phoneme clusters, since the relevant GTP rules are not included in the L1 rule sets.

As an alternative to letter-to-sound conversion by L1 rules, it is therefore suggested to apply *graphemically constrained phone substitutions*. In this rule type, phone substitutions are tied to particular graphemic representations. For example, native English speakers frequently mispronounce German [v] as [w]. This substitution, however, will only occur if [v] is orthographically represented by <w>, while [v] represented by orthographic <v> fails to undergo this rule. Such a restriction can be formalized as follows:

Phone layer :	[v]	→	[w]
	|		
Grapheme layer :	<w>		

The association line between the grapheme and phone layer indicates that the substituted phoneme [v] must be represented by an orthographical <w>, otherwise the rule will not be applicable.

Graphemically constrained rules require that the phoneme string is correctly aligned with the grapheme string in order to map each phoneme to the grapheme segment or cluster representing it. Since phonetic lexicons usually do not include such an alignment, a rule-based grapheme-phoneme (GP) alignment module for English, German, and French is incorporated in the rule system. Here is an example of GP alignment for the French city name *Questembert*:

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

The GP alignment is based on a set of language-specific rules containing all potential graphemic representations of each phoneme of the language. Note that the alignment module inserts 'zero phoneme' symbols '–' in cases where a graphemic slot has no phonemic counterpart ('silent phonemes'), as in the final <t> of *Questembert*. This property can be exploited to model particular pronunciation errors that could not be handled by rules operating only on the phone layer (see below). The following sections provide some more examples of how an association of the grapheme and phone layer can be utilized to model a number of characteristic pronunciation errors.

3.3.1 RECONSTRUCTING MUTE PHONEMES

In the above example, word-final silent <t> in the graphemic representation of *Questembert* is aligned to a 'zero phoneme', i.e. it is not entirely deleted from the phone layer, but represented by the placeholder symbol '–'. This is a useful feature for modelling mispronunciations where a graphemically present silent phoneme is pronounced instead of being dropped. For instance, a rule for deriving [kɛstãbɛʁt] from [kɛstãbɛʁ] – a common error type among native German speakers with a low proficiency of French – can be written as follows:

Phone layer :	[–]	→	[t]	/	__ ##
	|				
Grapheme layer :	<t>				

To model this type of error, the zero phoneme is used as rule input and will be substituted according to its graphemic representation.

3.3.2 RECONSTRUCTING REDUCED VOWELS

According to our speech data, it is a characteristic trait of native speakers of French and German pronouncing English material to ignore the English rule of 'vowel reduction' (full vowels become schwa in unstressed positions). Instead, these speakers tend to produce full vowels, as in *Stanford* *[stænfɔɐ̯t] or *Aston* *[æstɔn].

It is a widely held view in phonological theory that reduced vowels in English have an underlying representation as full vowels and that vowel reduction is a derivational process whose output becomes evident only in the phonetic surface representation. Pronunciation dictionaries for speech processing, however, do not necessarily contain underlying representations; more commonly, the pronunciation of a word is given in terms of its phonetic surface form. Since a surface schwa does not reveal the underlying full vowel it is derived from, a simple allophone mapping cannot capture this type of pronunciation error. Yet these errors can be modelled by a graphically constrained rule. For instance, the following rule will generate the above cited mispronunciation of *Aston* as *[æstɔn] :

Phone layer : [ə] → [ɔ]
|
Grapheme layer : <o>

Since the graphemic constraint is an obligatory condition for the application of the rule, any instance of schwa that is not derived from a full vowel will be left unchanged.

3.3.3 GRAPHEMIC CONTEXTS

In many languages there are spelling pronunciation rules which are not defined by an immediate grapheme-phoneme association; instead, they are triggered by the surrounding graphemes. In German, for instance, orthographic double consonants indicate shortening of the preceding vowel. It was found in our speech data that native French speakers reading German material show the opposite tendency, namely to *lenghten* vowels before orthographic double consonants (e.g. in the town name *Ammensleben* *[aːmənsleːbən]).

Thus, an additional rule type is required in order to tie phone replacements to *graphemic contexts* (as opposed to the *graphemic constraints* outlined above). E.g., a rule for 'vowel lengthening before orthographic double consonants' must be formulated as follows:

Phone layer : [a] → [aː]
|
Grapheme layer : < > / __ <CC>

Here, the right context <CC> of the (arbitrary) grapheme representing [a] triggers the phoneme substitution [a] → [aː]. Since the graphemic representation of the substituted phone is irrelevant for this rule, the grapheme slot can be left unspecified.

3.3.4 LANGUAGE-SPECIFIC GRAPHEMES

In cases where L2 orthography contains language-specific characters or diacritics, non-native speakers reading L2 words frequently replace these characters by the most similar native L1 characters. Characters like German 'Umlaute' (vowels marked by a diaresis) <ä, ü, ö> will thus be pronounced just like their non-marked counterparts <a, o, u>. Such misreadings caused by 'graphemic cognates' can be modelled by restricting phoneme substitutions to the relevant diacritically marked graphemes. Here is an example of a rule that models the mispronunciation of German <ü> as [ʊ]:

Phone layer : [ʏ] → [ʊ]
|
Grapheme layer : <ü>

This rule generates variants such as *Düsseldorf* *[dʊsəldɔʁf], which are quite common e.g. among native English speakers. Note that the rule application will be blocked if [ʏ] is graphemically represented by <y>, as in *Stadtkyll* [ʃtatkʏl]. Rules of this type will also be useful for various target languages not included in this study, particularly if the diacritic marks a notable phoneme shift (e.g. <ø, å> vs. <o, a> in Scandinavian languages).

4 FUTURE EXTENSIONS

In its present status, the rule system includes sets of post-lexical accent rules for English, French, and German. Currently, the number of rules per language direction is 80-100. Future extensions of the rule system will focus on two issues: (i) Modelling shifts in word stress patterns that can frequently be observed in non-native pronunciation variants (L1 patterns transferred to L2); (ii) the role of morphemes and lexemes which are part of the learned vocabulary (of speakers with some L2 proficiency). The data indicates that these elements (e.g. *-stein*, or *-mühle* in German city names) are less susceptible to accented pronunciation and may thus escape the effects of the substitution rules. Furthermore, an extension to additional (native and target) languages is scheduled. Rule sets for Italian (as L1 and L2) and Dutch (as L2 only) will be formulated.

ACKNOWLEDGMENT

This study is funded by the *Deutsche Forschungsgemeinschaft* (DFG).

REFERENCES

[1] Cremelie, N. and ten Bosch, L., 2001. "Improving the Recognition of Foreign Names and Non-Native Speech by Combining Multiple Grapheme-to-Phoneme Converters." *Proceedings ISCA ITRW Workshop Adaptation Methods for Speech Recognition*, Sophia Antipolis, France [on CD-ROM].

[2] Kenstowicz, M., 1994. *Phonology in Generative Grammar*. Cambridge/MA, Oxford: Blackwell.

[3] 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.

[4] Schaden, S., 2003. "Rule-based lexical modelling of foreign-accented pronunciation variants." To appear in: *Proceedings 10th Conference of the European Chapter of the Association for Computational Linguistics* (EACL '03). Budapest, Hungary.

[5] Schane, S. A., 1968. *French phonology and morphology*. Cambridge, Mass.: MIT Press.

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

[7] Strik, H., 2001. Pronunciation Adaptation at the Lexical Level. *Proceedings ISCA ITRW Workshop Adaptation Methods for Speech Recognition*, Sophia Antipolis, France [on CD-ROM].

[8] Ternes, E., 1976. *Probleme der konstrastiven Phonetik*. Hamburg: Buske Verlag.