

# On deriving rules for nativised pronunciation in navigation queries

Isabel Trancoso\*, Céu Viana\*\*, Isabel Mascarenhas\*\* and Carlos Teixeira\*  
INESC / IST\*, CLUL\*\*

INESC, R. Alves Redol, 9, 1000 Lisbon, Portugal

Phone: +351.1.3100268, Fax: +351.1.3145843, Email: Isabel.Trancoso@inesc.pt  
<http://www.speech.inesc.pt>

## ABSTRACT

Navigation queries are typical examples of contexts in which a recognizer may have to deal with non-native names. In order to build a pronunciation lexicon with these names, special GtoP rules may be derived. The paper addresses this problem in the context of navigation queries in French including German names and vice-versa. The special GtoP rules were mostly based on statistics derived from cross-lingual spoken corpora.

Keywords: non-native, pronunciation, lexicon, grapheme-to-phone, navigation

## 1. INTRODUCTION

This paper describes some of the problems faced with the use of foreign names in navigation queries in the framework of the TELEMATICS project VODIS (Vocal Interfaces for Driver Information Systems)<sup>1</sup>. The project's aim is the development of robust spoken interfaces for use inside a car, namely, for controlling the radio, the CD changer, the phone and, most important for the current work, the navigation system. The two official languages of the project are German and French, according to the interests of the car and equipment manufacturers involved in the project. Two project demonstrators for these two languages were developed and tested: the prompted approach and the mixed initiative approach. The development of these two demonstrators, per se, raises enough complex and innovative issues in this particular environment; however, it was felt that the issue of cross-linguality should be approached as well, given its importance in the context of navigation systems.

In fact, a German driver may find it very useful to be able to address his own German navigation system in France, in a query such as *Ich möchte nach Toulouse in die Rue des Lois.*, or conversely, for a French driver in Germany: *J'aimerais connaître le chemin le plus rapide pour aller dans la Zaehringstrasse.*

Hence, although it is not planned that a cross-lingual system would be fully integrated and tested in any of the demonstrators, the issue of recognition of foreign names

was addressed in the project. In order to take this into account, two spoken cross-lingual corpora were built: German subjects speaking French and French subjects speaking German. The two cross-lingual corpora are not as significant and as balanced as desired. Nevertheless, we felt that the amount of data allowed us to attempt deriving some relevant statistics on second language pronunciation.

Research on second language acquisition has shown that although the phonological inventory of the native language may condition the perception and performance of L2 at least during the first learning stages, it is not possible to build a clear pattern of mispronunciations on the basis of a contrastive analysis of phonological inventories only. Difficulties may also be observed for categories that are distinctive in both languages but differ in their contextual realizations. As some contrasts are easier to acquire than others and contextual realizations may considerably differ with training (see [4] for a review), a strong variability in mispronunciation patterns may be expected.

Some approaches have been recently presented to tackle the problem of non-native pronunciations in speech recognition. Deriving a set of alternative pronunciations is a typical one, either by hand, based on the knowledge of phonetic phenomena involving segmental perception and production in second language acquisition/learning, either by some automatic method. In [5], for instance, the set of alternative pronunciations for each word was integrated into a single probabilistic transcription network. Other approaches use a multi-phonetic acoustic space to represent the expanded phonetic space covered by the alterations in non-native phoneme pronunciations. These approaches collapse acoustic models of phonemes from different languages, on the basis of the articulatory/acoustic similarities among these sounds [1]. These multi-lingual phone models can be created by mapping to the IPA based phone set, for instance, or by some type of automatic data-driven clustering [2].

The VODIS context imposes some limitations in what concerns the speech recognition system, which restrict our type of approaches to deriving alternative pronunciations to be included in the lexicon. The main issue is, therefore, how to derive these pronunciations automatically.

<sup>1</sup><http://isl.ira.uka.de/VODIS/>

Besides the native and non-native material collected, we had two other sources of knowledge: the LexTool GtoP system from the speech recognition provider (L&H) and the ONOMASTICA inter-language lexicon for both languages [6].

This paper starts by a brief description of the cross-lingual corpora developed in the framework of VODIS (section 2). The main part (section 3) is devoted to the pronunciation of foreign names and the derivation of simple rules for building alternative pronunciations. The last section summarizes our work.

For easier reference, we have adopted SAMPA phonetic symbols for transcribing the French and German words.

## 2. CROSS-LINGUAL CORPORA

As mentioned above, two cross-lingual corpora were collected in the scope of the VODIS project:

- The German cross-lingual collection involved 28 speakers. Each of these German subjects was asked to speak, besides some native material, a list of 38 French keywords (out of a list of 63 command & control words, e.g. "raccrocher"), and 10 spontaneous navigation queries in German with French destinations. The know-how of French of each speaker is summarily indicated. Almost all the speakers had a very good knowledge of French or were "taught" by the recording monitor how to pronounce the names if that was not the case.
- The French cross-lingual collection involved 150 speakers. Each of these French subjects was asked to speak, besides some native material, up to 6 spontaneous navigation queries in French with German destinations, and 3 German city names (also spelled in French). The know-how of the non-native language is not indicated.

Both corpora were recorded in a lab environment. The unbalance between these two corpora results from the fact that the German partners preferred to separate their spoken data collection into two parts - a native one (with 135 speakers and a number of digits, C&C keywords, phone queries, radio commands and native navigation queries per speaker) and a non-native one described above, whereas the French partners preferred to merge the two collections. This lack of balance and the fact that very few material is actually spoken in the two corpora (just a subset of the French C&C keywords, as most of the proper names are different), conditioned the work that was done in this framework.

## 3. NON-NATIVE PRONUNCIATION

Given the limitations imposed by the speech recognition system adopted in the project, we have attempted to derive a set of simple rules in order to automatically generate

alternative pronunciations to be included in the lexicon.

The first question we addressed was, naturally, the adequacy of the LexTool GtoP system in accounting for part of the observed variability. Let us deal first with the case of French subjects speaking German names. It seemed a reasonable expectation that rules for French could correspond to the most probable pronunciations of French speakers with no knowledge of German and the rules for German to those of French speakers with a very good knowledge of that language. In our database, however, these two sets of rules account only for 2.9% and 3.1% of the cases, respectively. A more close analysis of the available data showed that even when the know-how of the foreign language is very limited, the speaker can typically do a better job of processing non-native grapheme sequences than the GtoP conversion system.

The following examples illustrate the large variability observed with the non-native speakers (we have selected isolated word names since we only had access to GtoP systems for isolated words):

Proper name (German)	Pronunc. by French GtoP	Pronunc. by French
Aachen	aaSA~	aS9n
Brandenburg	brA~dA~byr	brA~d9nburg
Burgbernheim	byrgbernEm	burgBERnajm
Heidelberg	EdElbEr	ajdElbErg
Neuplatendorf	n9platA~dOrf	n9plat9ndOrf
Oberstenweiler	ObErstA~wEle	ob9rstA~vajl9r
Schmilau	Smilo	Smilaw
Schwaig	SwE	SwEg
Strachau	straSo	straSo
Velgen	vElZA~	vElZEn

**Table 1:** Examples of pronunciations of German proper names (by French GtoP and by French speakers).

A close observation of the data has shown us that most speakers know that final consonants are pronounced in German or, at least, they are aware of the fact that silent final consonants are specific of French. Thus, the French GtoP rules according to which "g" is silent after "r" (e.g. "Heidelberg") or "r" is silent after "e" (e.g. "Oberstenweiler") do not apply. In several cases, the observed pronunciation coincides with the one expected for French words, since with few exceptions endings in "Consonant-en" are pronounced as "Consonant-[En]" (e.g. "Velgen"). It alternates, however, with [9n] (e.g. "Aachen"), which is much more frequent (77% against 20%). As 80% of the words end in consonant and only 2% of those are silent, several other GtoP rules for French fail to account for the observed correspondences.

Word internal "n" and "m" may combine with the preceding tautosyllabic vowels to indicate their nasality, but nasal/non-nasal realizations are also found (29% and 71%, respectively). Sequences such as "au", "ei", and "ai" constitute another important source of variability, correspond-

ing either to a single vowel or a diphthong. Although [S] is certainly the preferred reading for "ch", it may alternate with [k], as in French, with acceptable German [x] and with [R], as the closest approximation of this sound that does not belong to the French inventory. Another non-existent sound in this inventory is the aspirated "h". Nevertheless, aspiration may occur, but most often the "h" is simply omitted, or it prevents liaison with the following vowel. Other major sources of variability were the graphemes "w" (either as a [v] or as [w]), "g" when followed by "e" or "i" (pronounced either as [g] or as [Z]) and "u" (either as [u] or as [y]).

Graphemes "e" and "o" may be pronounced as [E]/[e] and [O]/[o], respectively. In both cases, the first pronunciation of each pair is preferably observed in closed syllables, and the second in open ones. The grapheme "e" is often interpreted as a *schwa* and an important part of [e]/[ə] alternations may be accounted for with a rule similar to the one that governs vowel-zero alternations. However, when the grapheme consonantal sequences do not occur in French, or when the resulting cluster is difficult to pronounce, an increased variability is found, concerning not only the cluster resolution but also that of the preceding vowel.

Based on the observations described above, we have derived a very simple set of rules for producing alternative pronunciations for French subjects speaking German. This set of rules was written as a *sed* script and includes around 120 commands. It assumes the orthographic entry is written in capital letters and produces SAMPA symbols for French. Alternative pronunciations are indicated in between brackets (i.e., [aw;o] indicates the two possibilities of pronouncing "au"). We have considered only some of the cases described above, avoiding the [E]/[e] and [O]/[o] alternations, in order not to include too many alternative pronunciations.

The rules have been trained on a subset (80%) of the ONOMASTICA inter-language lexicon (French speaking German, around 1000 entries), and were tested on the remaining subset (20%). 73% of the observed pronunciations in the ONOMASTICA test set were accurately described by the rules. This lexicon was built on the basis of what the authors considered to be the "average" knowledge of German in the country and therefore is more or less coherent. The pronunciation of the vowel "e" and the voicing assimilation in consonant clusters accounted for the most systematic errors. This type of clusters, however, was not coherently treated in the training set. When tested on the spoken entries of the cross-lingual corpus (around 400), the results were much worse. A close observation of the errors has shown us once more that the inter-speaker variability due to the different knowledge of the foreign language is too large to be described just by the cases we have selected as most frequent. Increasing too much the number of cases of alternative pronunciations, however, would lead to a combinatorial explosion.

In what concerns the opposite problem, i.e. German subjects speaking French, the analysis is more difficult. The number of speakers collected in VODIS is much more limited, and we do not know if the database can be considered as illustrative of the general know-how of the French language in the country. In fact, their relative high familiarity with French can perhaps justify the fact that in the pronunciation of the command words, 81% of the entries could be considered as adequate pronunciations in French. In the pronunciation of proper names, the percentage was at least of the same magnitude.

Although pronunciation errors were not frequent at all, some of the most common are illustrated below with a few examples from command words:

Command words	Pronunciation by Germans
guidage silencieux	(in "gui" (as [gaj]) and "len" (w/o nasalization)
autoroutière	(in "è" (as [a])
mode manuel	(the "u" was not pronounced as [y])
curiosité	(the "u" was not pronounced as [y])
aéroport	(speakers pronounced the final [t])

**Table 2:** Examples of pronunciations of French navigation C&C words by German speakers.

The example of "guidage" is one of the most interesting ones since it illustrates a common situation: if the speaker is not very familiar with the foreign language, he/she may pronounce the word as in the foreign language he/she is most familiar with (English, in this case).

The ONOMASTICA inter-language lexicon for German speaking French [3] does not give us a very valuable information on typical pronunciations since it looks as if it was built assuming zero knowledge of French (see Table 3 for examples).

Proper name	ONOMASTICA
AIX EN PROVENCE	alks En pRo:fEnts@
BOULOGNE BILLANCOURT	bu:lOgn@ bllaNku:6t
CHÂTELET	ka:t@l@t
HAUTS DE SEINE	haUts de: zaIn@
PONT L'ÉVÊQUE	pOnt le:fEkv@
ROCHEFORT	ROx@fORt
SAINT PAUL	zaInt paUl
VENDÔME	fEndo:m@

**Table 3:** A few examples taken from the ONOMASTICA inter-language lexicon (German speaking French).

Hence, the best we can do with the available databases is to generate two alternative pronunciation lexica: one assuming zero knowledge of French, as in ONOMASTICA, and another one assuming a very good knowledge of French, as in the recorded database. The first can be generated using the German GtoP system and the latter using the French GtoP system post-processed by a suitable conversion between the phonetic symbols in the two languages. The two alternatives will be illustrated with a few examples of proper names from the spoken database:

Proper name (French)	Pron. by German GtoP	Pron. by French GtoP*	Pron. by Germans
Lusignan	lUzIɣna:n	lyziJA	lyziJA
Pleurtuit	plOIRtuIt	pl9Rtwi	pl9Rtui
Lois	lOIs	lwa	lua
Fontaine	fOntaIn@	fO tEn	fO tEn
Germain	geRmaIn	ZERmE	ZeRmE
Georges	gejORg@s	ZORZ@	ZORZ
Journées	jouRnes	ZuRne	ZuRne

**Table 4:** Examples of pronunciations of French proper names.

The conversion between phonetic alphabets that was used in the above table (marked with an asterisk) implied not only the conversion of the French "w" to the German "u" (which we regarded as closest), but also the addition of nasal vowels which do not exist in the German phonetic inventory, but were clearly pronounced by the German speakers. The distinction between "A" and "E", however, was very small, so a single phonetic symbol could be used for both.

In between the two extreme pronunciation alternatives, one could consider the same type of pronunciation problems that has been found for French subjects speaking German, although in the opposite direction. The evidence of these mistakes in our database was, however, very reduced. Hence, although a similar set of rules could have been derived, these rules could not be validated by the available data.

## 4. CONCLUSIONS

This paper summarizes the work done on non-native speech recognition within the framework of VODIS. The emphasis of the report was on the derivation of alternative pronunciation lexica for German and French cross-lingual experiments. The derivation was based on statistical data collected by the VODIS partners and on previous know-how from the ONOMASTICA project.

From a speech recognition point of view, the usability of these pronunciation variants is still an open question which must be studied next in conjunction with the use of either mono-lingual or multi-lingual phone models.

From a text-to-speech synthesis point of view, the subject of pronunciation of non-native names is also very important. In this context, the issue of inter-speaker variability which was our main concern in this paper is not relevant, as we are interested in deriving a single pronunciation. However, the cross-lingual data collected in the VODIS project may be very useful to derive what can be considered the most common pronunciation for many proper names, and also to get some insight on the need to use an expanded phone set, not only for recognition, but also for synthesis purposes. In fact, we have verified that a large percentage of speakers is able to pronounce sounds which do not belong to their language phone inventory.

Our previous data collection efforts in terms of non-native pronunciations concerned first Portuguese subjects speaking different languages and later subjects from different countries speaking English. The VODIS cross-lingual data collection did not involve any vocabulary in English, which is undoubtedly the most current second-language learned nowadays on a world-wide scale. Nevertheless, it was very interesting to notice that when subjects know very little about a foreign language they frequently use their knowledge of English to pronounce the unknown words. This trend was often verified in the VODIS collection with French and German speakers, and it would be very interesting to study its existence in other cross-lingual data not involving English.

## Acknowledgements

Although this work was mainly done at INESC, it would have been impossible without the cooperation and help of many VODIS colleagues. Special thanks go to: Robert Grudszus, for the definition of the cross-lingual corpora and its collection in Germany, Philippe Doignon for the collection in France, Uwe Meier for all the help with the tkvodis tool for cross-lingual collection, and Johan Smolders and Jan Odijk for providing us with the phonetic transcriptions of the L&H GtoP tool. Last but not the least, many thanks to Luis Arevalo for his great help throughout the project.

## REFERENCES

- [1] P. Bonaventura, F. Gallochio, J. Mari, G. Micca (1998), "Speech recognition methods for non-native pronunciation variation", Proc. Workshop on Modeling Pronunciation Variation for Automatic Speech Recognition, pp. 17-22, Rolduc.
- [2] Kohler, J. (1996), "Multi-lingual phoneme recognition exploiting acoustic-phonetic similarities of sounds", Proc. Int. Conf. on Spoken Language Processing, Philadelphia.
- [3] Mengel, A. (1993), "Transcribing names - a multiple choice task: mistakes, pitfalls and escape routes", Proc. 1st ONOMASTICA Research Colloquium, pp. 5-9, London.
- [4] Strange, W. (1995), "Cross Language studies of speech perception: A historical review", In W. Strange (Ed) *Speech Perception and Linguistic Experience: Issues in Cross-Language Research*, York Press, Timonium, Maryland.
- [5] Teixeira, C., Trancoso, I., and Serralheiro, A. (1997), "Recognition of non-native accents", Proc. of the European Conf. on Speech Comm. and Tech., pp. 2375-2378, Rhodes.
- [6] Trancoso et al (on behalf of the ONOMASTICA Consortium) (1995), "The ONOMASTICA interlanguage pronunciation lexicon", Proc. of the European Conf. on Speech Comm. and Tech., Madrid.