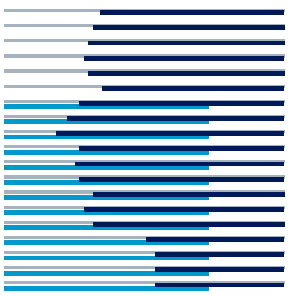
How to Integrate Phonetic and Linguistic Knowledge in a Text-to-Phoneme Conversion Task: A Syllabic TPC Tool for French

Nicole Beringer



Technical Report No. IDSIA-18-04
October 2004

IDSIA / USI-SUPSI
Dalle Molle Institute for Artificial Intelligence
Galleria 2, 6928 Manno, Switzerland

This work was published on ICSLP 04 and is part of the conference proceedings.

How to Integrate Phonetic and Linguistic Knowledge in a Text-to-Phoneme Conversion Task: A Syllabic TPC Tool for French

Nicole Beringer

October 2004

Abstract

The goal of this work is to present a text-to-phoneme conversion (TPC) check tool for generating dictionaries usable for Speech Synthesis and Speech Recognition which includes phonetic and linguistic knowledge[1].

The aim is to improve and accelerate the task of producing canonic pronunciation dictionaries in languages where no simple text-to-phoneme conversion is possible.

The prototype was developed for a French syllable-based text-to-phoneme task but it is portable to other languages.

1 Introduction

In Speech Synthesis as well as in Automatic Speech Recognition (ASR) the phonetic transcript of the orthographic word form is needed. Therefore, dictionaries are generated which map the orthographic form to the canonical phonetic form of a word. Despite sounding easy this is in fact quite difficult because the corresponding phonetic transcripts have to be a hundred percent correct for further use.

In order to guarantee that there will not be errors in speech synthesis or ASR caused by an incorrect TPC all entries of the dictionaries have to be corrected manually - a very time-consuming job.

Of course, the better the TPC transcript the less correction is needed, which, as a consequence minimizes the realtime factor of that work.

One way to improve the TPC transcript and to minimize the realtime factor of the manual correction is to use phonetic and linguistic knowledge within the TPC.

Former studies [1] on German show that there are numerous problems in mapping German orthography to a high quality transcript without using any linguistic background information.

Including knowledge of the morphological structure of words [2] achieves a high accuracy but it has to be considered that the morphological decomposition is difficult and has a high error rate. It was recently shown [1] that the performance of a syllable-based TPC is comparable to the morpheme based one but that it is not as error prone in its implementation.

This consideration was taken into account for this work to develop a TPC tool and a TPC checker for French. The work resulted in two main parts:

- a TPC tool for French, which includes syllabic knowledge of French (SAM-PA Version [4])

[1] In the here presented paper we mostly refer to speech synthesis. But using the TPC checker also in an ASR environment is possible

- a TPC check tool, which checks whether a phoneme string is a possible syllable for French or not.

The former was developed analogously to the P-TRA German system [3, 1](see section 2); the latter includes a new method for checking whether a syllable is possible in a language or not. Of course, this is the main problem for the TPC tool.

In section 3 the problems for TPC in French are shown. An analysis of French syllable structure is given in section 4 and we describe our prototype for the French TPC checker in section 5. Finally, results and future work are discussed in the last section.

2 Syllable-based TPC

2.1 Basics

As a starting point, translating the orthographic form of an utterance to its phonetically transcribed equivalent may be done by a simple text-to-phoneme conversion, where each letter is translated into an IPA [6] or SAM-PA [5] symbol to obtain the canonic form of the word.

Each orthographic form has one standard phonetic reference which is called the standard phonetic or canonic form of the word.

This conversion presents little or no difficulty in languages where the letters of the orthographic form map one-to-one to the sounds of the canonic form. Languages which mostly map the letters one-to-one to the sounds are for example Spanish, Italian and Portugese.

Other languages which adhere to complex orthography like German, English or French exhibit a non-trivial conversion. Within this paper complex orthography is defined as an orthographic form which is based on etymologic word forms and therefore may differ to a greater or lesser extent in the phonetic realization.

2.2 Including Phonotactics

To deal with complex orthographic word forms we have to know something about the phonotactics of the language to convert. This is often done by including the morphological structure of words but although it is highly acccurate for TPC the required morphological decomposition is tedious and error-prone.

[1] describes an alternative approach where using the orthographic syllable instead of morphemes is proposed to deal with the multiple choice of spellings for the same phoneme in German.

As will be described below, the main problem is the ambiguity of orthographic representation of the same phoneme in all languages which do not map letter to sound one-to-one.

2.3 TPC Problems in German - Some Examples

In German we have the problem of vowel lengthening vs. spoken /@/ or /h/ depending on the syllabic surrounding, a problem which is easy to solve for human beings but impossible for a naive TPC system without any further description.

This problem is multiplied by the possibility of compound words. In German it is possible to combine almost every word with another form to a new one. Consonant or vowel clusters typical for word boundaries are no longer spoken as in the isolated word. To solve this problem one has to consider the syllabic structure.

Another typical problem for German is the use of glottal stops before vowels at the beginning of a word or after another vowel; this can be solved by taking the syllabic structure of a word into account.

However, having modelled these problems by syllable structure, there still exist exeptions, e.g. anglicisms in German. A TPC system has to model the "denglish" pronunciation instead of using the right native form in order to obtain constistent results. This problem arises in all languages regardless of the transcription method.

For German we could use the P-TRA (**P**honetische **TRA**nskription) TPC system [3] to translate text-to-phoneme. Syllable information is directly included in rewrite rules of the form:

left context | orthographic searchstring | right context => Phonemic correlate [1].

Contrary to our TPC for French presented here the German P-TRA needs the orthographic syllable structures in the input to achieve good conversion results.

3 TPC for French

The TPC for French presented here was developed to automatically transcribe French orthographic words into their canonical form for synthesizing and recognition purposes.

In contrast to the German P-TRA, which derives syllabic information from the orthography we have to include it - as well as the intonation information - in the **canonic** form.

3.1 The TPC-Tool

The text-to-phoneme conversion is done in three main steps:

- letter to phonemic conversion of the relevant sound. Here, syllabic information is included to avoid ambiguity.
- syllable structure of the canonic form: syllables within the canonic form are marked by "|".
- main accent is set to the relevant syllable of the canonic form.

The following paragraphs give a detailed description on how the three steps are realized:

TPC: Starting with a syllable-based conversion the system modifies orthographic representation in order to get the phonetic transcript. This is done as follows:

- The longest mapping orthographic searchstring is changed into the phonetic equivalent using conversion rules. The searchstring can be the whole word or a syllabic structure.
- While converting this searchstring into its phonetic representation, syllable markers are set according to the syllabic structure. Note, that whole words are devided into syllables and therefore suffice the syllabicility.
- For those orthographic parts where no syllabic structure is given yet, the system maps letter to sound one-to-one depending on French phonotactics.

Note as well, that we do not consider the environment the syllables are found in like in P-TRA.

Once the canonic form has been produced, "liaison" information - information on how word boundaries should concatenate - is given according to [7]. We distinguish between an obligatory "liaison" and an optional one. The latter may be included, the former must be included. Besides the "liaison" there are word boundaries which are never concatenated. This can be found in the rules as well.

Syllabic Structure: If the results of a conversion is known to be a syllable, the boundaries to the adjacent elements are marked with "|".

Accentuation: In French, normally the last syllable of a word is accentuated if its nucleus is not /@/. This information is included in the canonic form.

3.2 TPC Problems in French

It is clear from the TPC description for French that there are problems - mainly of ambiguity - in the conversion processing.

Similarly to German there are ambiguities in vowel realization:

There are many-to-one conversions in each case where the letter has an accent. Furthermore, the same phoneme can have different orthographic strings.

Consonants vary in their representation depending on their environment they are found in. Before front vowels velars are represented as fricatives, before back vowels they are mapped one-to-one.

4 French Syllable Structure

French syllables [7] usually have an onset of mostly two, maximally four consonantal phonemes. In the case of three or four consonantal phonemes we have the sequence /st/ followed by /R/ followed by the semivowel /H/. Otherwise, the second preferably is a liquid or semivowel.

The nucleus can be any kind of vowel. Note that the centralized Schwa-vowel never bears the word accent!

The coda consists of at most three consonantal phonemes. If the syllable has three coda consonants they are concatenated as follows:

- plosive fricative plosive.
- /R/ plosive /R/

In principal, if consonantal clusters in the phoneme sequence are interrupted by /@/, this vowel is skipped if the new sequence does not have more than two consonants. This is known by the term “mot phonétique”.

“Liaison” is blocked in the following cases:

- the next word begins with an “h aspiré”
- after real pauses - after a “mot phonétique”

It is obligatory in case of assimilations:

- regressive: voiceless becomes voiced and vice versa
- progressive: “e instable”, after or in a “mot phonétique”

It is optional in any other case.

5 TPC Check Tool for French

In section 3.2 we showed some of the problems we have in the text-to-phoneme conversion which makes the TPC result somewhat error prone. Including known French syllables in the TPC tool improves the result but we still have unspecified sequences which could consist of impossible phoneme concatenations or which are not segmented on syllabic basis.

Therefore, we analyzed the phonetic and syllable structure of French and defined “possible” phonetic syllables.

5.1 "Possible" syllables

Definition:
'*"Possible" syllables are phonetic strings between "|" which have not yet been covered by an existing translation rule (orthography translated into phoneme string) and have a syllable structure following French phonotactics.*

Table 1 shows the structure "possible" syllables have to follow.

onset	nucleus	coda
{}	vowel (oral or nasal)	{}
semivowel		semivowel
liquid		semivowel liquid
plosive		liquid
fricative		plosive
nasal		plosive liquid
plosive liquid		fricative plosive
fricative liquid		fricative
nasal liquid		plosive fricative
liquid semivowel		nasal
plosive semivowel		nasal fricative
fricative semivowel		fricative nasal
nasal semivowel		liquid nasal
/st/ liquid		liquid plosive liquid
/st/ liquid semivowel		plosive fricative plosive

Table 1: *classification of syllable structure*

Note that we obtain possible syllables by concatenating the entries of the columns. Per column only one line is possible.

5.2 TPC Checker - Functionality

The main function of the TPC checker is to find out whether a phoneme sequence is a syllable or not.

Based on table 1 all phoneme strings between "|" are checked. If the sequence differs from all given possible syllables, the TPC string is first looked up in a list of "impossible" syllables.

If a "syllable" is found in the list of impossible syllables, an automatic phoneme string correction according to former encountered manual corrections is performed to obtain better results in the checking part (self-learning system).

Otherwise, the "syllable" and its word environment is echoed to be checked by a human.

In case that the "syllable" in question is considered by the expert to be indeed correct, the phoneme string is included in the conversion rules of the TPC tool and the TPC checker, otherwise it is included in a list of impossible syllables together with the manual correction.

Based on the word accent rule

> Word accent is set on the last syllable if its nucleus is not /@/

the position of the word accent is checked and corrected if wrong.

5.3 TPC Checker - Results

Table 2 shows the TPC error rate of the simple one-to-one conversion, the syllabic TPC for French and the TPC checked strings compared to the manually corrected form as well as the word accent error rate. The number of known syllables is 5021; the TPC checker includes 225 aditionally possible syllables. The number of words to be converted is 16249.

	correct words	word accent error %	correct %
one-to-one	2637	9.10	16.23
syllable-based TPC	10760	5.23	66.13
TPC checker auto	11414	4.73	70.25
TPC checker manual prompt	15669	3.57	96.43

Table 2: *Comparison of the TPC error rate of the three conversion methods for French*

Although the one-to-one conversion performance is out of question its word accent error is not too bad. This proofs that the accented ultima in French is very regular.

With the syllable-based TPC we gain almost 50% in word correctness. Accents are 4% better as well.

With the checker the word correctness improves about 4% only having included automatic corrections out of the list of impossible syllables (TPC checker auto), the word accent error gains 0.5%.

The hit rate of the unknown syllables prompted for manual correction (TPC checker manual prompt) covers another 26.18% and reaches a conversion correctness of 96.43%.

6 Conclusion

It was shown that if phonetic and linguistic knowledge is included in TPC the performance of the phonetic transcript improves regardless of the language.

We have further shown that syllable-based TPC for French reaches a moderate performance. This can be drastically improved by using syllabic schemes in a checking sequence.

Although we still need a human correction to obtain phonemic representations for speech synthesis and speech recognition dictionaries, the realtime factor in the correction step minimizes by every new learned syllable.

Finally, we will include an environment check for the defined syllable structure following the P-TRA TPC to better manifest the right syllables.

References

[1] Libossek, M.; Schiel, F.: “Syllable-based Text-to-Phoneme Conversion for German“ in: Proc. of the International Conference on Spoken Language Processing, Beijing, 2000, pp. 283-286.

[2] Wothke, K.: “Automatic phonetic transcription taking into account the morphological structure of words” in: Technical Report, IBM Germany, Scientific Center, Heidelberg, 1991.

[3] Stock, D.: "P-TRA - Eine Programmiersprache zur phonetischen Transkription," in: Hess, W.; Sendlmeier, W. (eds): Beiträge zur Angewandten und Experimentellen Phonetik, Franz Steiner, Stuttgart, 1992.

[4] Heuven, V. J. Van; Pols, L.C.W. (eds): "Analysis and Synthesis of speech" Walter de Gruyter, Berlin, 1993.

[5] www.phon.ucl.ac.uk/home/sampa/home.htm.

[6] Handbook of the International Phonetic Association

[7] Martinet, Walter, Carton, Lucci: "Etude phonétique du français contemporain à travers la variation situationelle" Grenoble 1983.