

# Multi-Context Rules for Phonological Processing in Polyglot TTS Synthesis

Harald Romsdorfer and Beat Pfister

Speech Processing Group  
Computer Engineering and Networks Laboratory  
ETH Zurich  
{romsdorfer,pfister}@tik.ee.ethz.ch

## Abstract

Polyglot text-to-speech synthesis, i.e. the synthesis of sentences containing one or more inclusions from other languages, primarily depends on an accurate morpho-syntactic analyzer for such mixed-lingual texts. From the output of this analyzer, the pronunciation can be derived by means of phonological transformations which are language-specific and depend on various contexts. In this paper a new rule formalism for such phonological transformations is presented, which complies also with the requirements of the mixed-lingual situation.

## 1. Introduction

Following the approach of generative phonology in [1], text-to-speech (TTS) synthesis requires the underlying syntactic structure in order to derive the correct pronunciation, as it was shown e.g. in [2] and [3]. This underlying structure is even more important in the case of polyglot TTS where mixed-lingual input text has to be processed. Such texts can contain various inclusions from other languages.

The morphological and syntactic analysis of mixed-lingual text presented in [4] provides an appropriate solution to this issue. The syntax tree from this analysis not only describes the morphological structure of the words and the syntactic structure of the sentence, but also includes a language label for each constituent.

In this paper we describe in Section 2 the requirements to a component which performs two tasks, namely phonological transformations on the one hand and assimilation and pronunciation modification on the other hand. Section 3 illustrates such transformations by means of an example and describes how they can be specified by means of so-called multi-context rules. In Section 4 we show the realization of a component for such phonological transformations. This realization is based on finite state automata (FSA).

The result of these transformations is a complete phonetic representation of a sentence to be synthesized. From this phonetic representation the phono-acoustical model derives the sentence prosody and generates the speech signal, as shown in [5].

## 2. Requirements of a mixed-lingual phonological component

The requirements of a mixed-lingual phonological component are evidently determined by the way of pronunciation wanted. For our polyglot TTS system we are aiming at the “Swiss manner” of pronunciation of mixed-lingual text. This means that foreign inclusions, especially English, French and Italian ones, are pronounced in accordance to the rules of the originating language. Assimilation to the base language, i.e. to German, is rather marginal and happens primarily close to language switching positions.

Consequently, the phonological component of our polyglot TTS system must be able to cope with the different phonological phenomena of each language involved. This requires a rule formalism that is flexible enough to describe all possible context restrictions of such phonological rules. The following section gives some examples of phonological transformations and the corresponding context restrictions.

### 2.1. Context restrictions of phonological phenomena

Phonological phenomena are language specific and depend on various contexts, as the following examples illustrate:

**German aspiration** In word-initial position, the German unvoiced plosives [p], [t] and [k] preceding a vowel are aspirated, denoted as [p<sup>h</sup>], [t<sup>h</sup>] and [k<sup>h</sup>], resp. They are also aspirated in word-final position before a break.

**French liaison** Within French noun groups, liaison is forbidden between a singular noun and a consecutive adjective, e.g. “un bruit effroyable” [œ-brɥi-e-frwa-jabl]; between a plural noun and a following adjective it is optional, e.g. “les amis agréables” [le-za-mi-(z)a-gre-abl]; liaison is mandatory between a preceding adjective and a noun, e.g. “un bon ami” [œ-bɔ-na-mi].

**French liaison consonant realization** The phonetic liaison consonant can be directly derived from the

corresponding graphemic consonant: “s”, “x” or “z” result in [z]; “c”, “q” or “g” in [k], etc.

**English linking “r”** Word-final “r” is usually pronounced only, if the following word begins with a vowel, e.g. “four eggs” [fɔ:r egz] but “four pounds” [fɔ: paʊndz].

These examples show that phonological phenomena depend on various contexts: The German aspiration rule needs only phonetic context, whereas the English linking rule requires both phonetic and graphemic contexts. The French liaison rules need phonetic, graphemic and syntactic contexts. Furthermore, since all of these phonological phenomena are language-specific, language forms another context.

## 2.2. Cross-lingual assimilation phenomena

As mentioned above, foreign inclusions in German sentences virtually keep the pronunciation prescribed by the originating language (see begin of Section 2). Even if assimilation of foreign inclusions to the base language is very weak, it is clearly present and must be handled correctly.

In a word like e.g. “Dufourstrasse”, which is composed from the French proper name “Dufour” and the German noun “Strasse” (*street*), the French [ʀ] has to be replaced by the German [r]. It would sound rather affectedly to pronounce [dʏfʊʀ[tra:sə] instead of [dʏfʊr[tra:sə]. Such assimilations occur only near the language switching position, however, and only in short inclusions.

## 2.3. Consequences for phonological rule formalism

Considering phonological and cross-lingual assimilation phenomena as presented above, a rule formalism has to comply with the following requirements:

- It must be possible to define rule contexts on the phonetic, graphemic and syntactic level.
- The formalism must allow the specification of language dependent contexts. Accordingly, it must be possible to constrain rules to specific cross-lingual contexts.
- Additionally, it is highly desirable that for a mixed-lingual phonological component for a set of languages, the superset of the rule sets of the individual languages (eventually extended with some rules for cross-lingual phenomena) can be used.

Therefore we have designed a new formalism for phonological rules, that extends our existing two-level formalism described in [6], and that meets all the requirements stated above. This formalism allows to specify multiple context constraints, i.e. graphemic, phonetic and syntactic constraints in a cross-lingual manner. Thus we called it a “multi-context rules” formalism.

## 3. Multi-context rules

Before introducing the multi-context rule formalism, we illustrate the function of phonological transformations by means of a mixed-lingual example sentence.

### 3.1. Mixed-lingual example

Let’s consider the mixed-lingual sentence “Anciens Amis sind keine Amis anciens.” (“Ex-friends are no old friends.”). This German declarative sentence contains two incomplete French noun groups (i.e. the article is missing which corresponds to the German indefinite plural form). The syntax tree of this sentence is shown in Figure 1.

This syntax tree specifies the pronunciations on the word level. The application of phonological transformations produces the standard pronunciation of the sentence which is (to simplify matters, syllable stress and prosodic phrase information has been removed):

[ʔäs-jĕ-z a-mi- zɪnt- k<sup>h</sup>aj-nə- ʔa-mi-(z) äs-jĕ]

Note that phonetic symbols in parentheses are optional. The following phonological transformations have been applied in this example:

The aspiration of [k<sup>h</sup>] in the German word “keine” follows the German aspiration rule defined in Section 2.1. The plosive [t] in the word “sind” is not aspirated, however, because there is no break after this word.

The standard pronunciation of the French partial noun groups is obtained by applying the French rules for mandatory and optional liaison, resp. The first French inclusion “Anciens Amis” is pronounced [äs-jĕ-za-mi]. Here the liaison consonant [z] is inserted. In the second incomplete French noun group “Amis anciens”, the liaison consonant [z] is optional (as defined in Section 2.1) which results in [a-mi-(z)äs-jĕ]. The actual realization of this optional consonant depends on the style of pronunciation wanted.

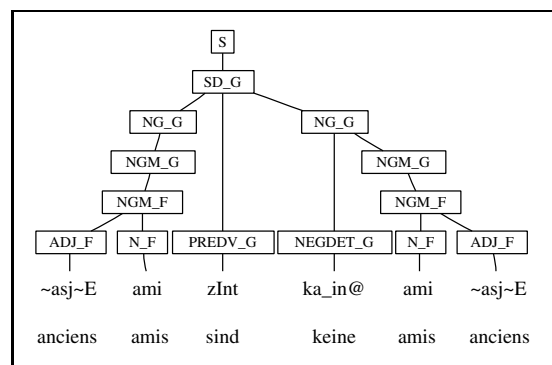


Figure 1: Syntax tree of the sentence “Anciens Amis sind keine Amis anciens.”, including graphemic and phonetic terminals. The phonetic symbols largely follow the SAMPA definition. The suffixes <sub>F</sub> and <sub>G</sub> of the constituent identifiers indicate the languages French and German resp.

Furthermore, a cross-lingual phenomenon has to be considered which arises from the German glottal stop rule. According to this rule, potentially every initial vowel of a word or a stem morpheme is preceded by a glottal stop. A similar rule applies also for foreign inclusions, i.e. a glottal stop has to be assigned to the initial vowel of the inclusion. “Anciens Amis” therefore has to be pronounced as [ʔãsjẽz a-mi] in our example sentence.

From this example it is obvious that only the application of both the German and the French phonological transformations can produce the desired standard pronunciation for a German sentence with French inclusions.

### 3.2. A multi-context rule formalism

We propose a rule form to specify phonological transformations on the syntax tree with restrictions on specific syntactic, graphemic and phonological contexts. A multi-context rule is specified by a subtree pattern plus an associated phonological rule separated by the symbol ‘:’.

$$\text{SubtreePattern} : \sigma/\rho \Leftrightarrow L \_ R ;$$

The application of this phonological rule is triggered at each matching position of the subtree pattern within the syntax tree.

The phonological rule is an extension to the well-known two-level formalism in [7] by including graphemic, phonetic and phonological symbols in the alphabet. The subtree pattern specify the syntactic context and define for each constituent if the graphemic and/or phonetic representation is subject to the phonological transformation defined by the rule. These patterns may be specified using constituent symbols plus additional wildcard symbols listed in Table 1.

For a successful match of a subtree pattern with the syntax tree two conditions must be fulfilled: first, all constituents dominated by a specific constituent  $K$  in the pattern must also be dominated by the corresponding constituent  $K$  in the syntax tree. Second, two consecutive constituents  $L$  and  $R$  of the subtree pattern must match with neighboring constituents in the syntax tree. Two constituents,  $L$  and  $R$ , are neighbors if there is no right branching in the path from  $L$  to the common immediate dominator of  $L$  and  $R$ , and if there is no left branching in the path from  $R$  to this dominator. Examples of subtree patterns are shown in Figure 2.

## 4. Implementation

The input to the phonological component is the syntax tree with graphemic and phonetic terminals. Every successful match of a subtree pattern on this input triggers

*	any sequence (0...n) of constituents including their (possibly empty) subtrees
?	any constituent (exactly one)

Table 1: Wild-card symbols used within syntax patterns.

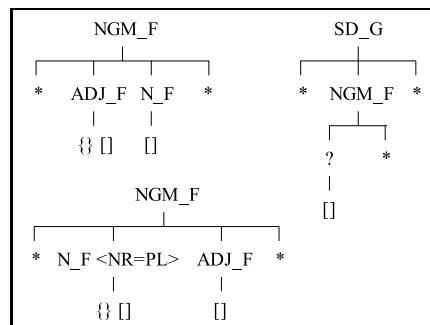


Figure 2: Examples of subtree patterns: the top left pattern specifies the syntactic context of French mandatory liaison between a preceding adjective and a noun within a French noun group. The operator ‘[]’ selects the underlying phonological representation of both constituents for application by the associated phonological rule. Also, denoted by the operator ‘{}’, the graphemic representation of the first one is selected. The bottom left pattern specifies the syntactic context of French optional liaison between a plural noun and a subsequent adjective. The pattern specifies the noun with an additional feature-value pair, i.e. <NR=PL>, which only matches plural nouns. The pattern to the right specifies a mixed-lingual syntactic context for glottal stop assignment within French noun groups as foreign inclusions in German declarative sentences.

the application of the associated phonological rule to a sequence of phonetic and/or graphemic terminals as specified by the operators ‘[]’ and ‘{}’ within this subtree pattern.

The associated phonological rules adhere to the standard two-level formalism as defined for our TTS system in [6]. These phonological rules can be compiled into a FSA using the same two-level compiler as standard two-level rules.

Figure 3 shows two examples of multi-context rules for French liaison that use the two left subtree patterns of Figure 2 as syntactic context. Applying the first one of these rules to the syntax tree of the example sentence in Figure 1 matches the last French inclusion, i.e. “Amis anciens”. The operators ‘[]’ and ‘{}’ select the graphemic and phonetic representation of the noun plus the phonetic representation of the adjective as input to the phonological transformation rule. This input sequence is shown on the left side of (1). The right side denotes the same sequence after insertion of the optional liaison [(z)] by the associated phonological rule:

$$\{\text{amis}\}[\text{ami}][\text{ãsjẽ}] \Rightarrow \{\text{amis}\}[\text{ami}(z)][\text{ãsjẽ}] \quad (1)$$

The second multi-context rule of Figure 3 matches the first French inclusion, i.e. “Anciens Amis”. The selected phonological input sequence is shown on the left side of (2). The corresponding output sequence on the right side shows the insertion of mandatory liaison [z]:

$$\{\text{anciens}\}[\text{ãsjẽ}][\text{ami}] \Rightarrow \{\text{anciens}\}[\text{ãsjẽz}][\text{ami}] \quad (2)$$

The mixed-lingual glottal stop rule using the left subtree pattern of Figure 2 as syntactic context is shown in Figure 4.

The sequence 'NGM\_F ( ? [] \* 0 )' matches any left-most subconstituent of a French partial noun group NGM\_F. Thus, this rule matches twice in the example sentence of Figure 1. Only the phonetic representation of that subconstituent is subject to further phonological transformations. The left sides of (3) show these input sequences, while the right sides present the corresponding outputs.

$$\begin{aligned} [\tilde{a}s\tilde{j}\tilde{e}z] &\Rightarrow [?\tilde{a}s\tilde{j}\tilde{e}z] \\ [ami(z)] &\Rightarrow [?ami(z)] \end{aligned} \quad (3)$$

## 5. Discussion and Conclusions

The presented multi-context rule formalism allows to describe phonological phenomena which are inherently language-specific and may depend on graphemic, phonetic, syntactic and other context. A set of such rules therefore can be used in the phonological transformation component of a TTS system.

For a polyglot TTS system that has to process mixed-lingual text of a certain set of languages, it is advantageous that the mixed-lingual phonological component can easily be constructed from the monolingual ones of the languages concerned. By means of multi-context rules this is possible, provided that the rules have been defined in a language-dependent manner. In this case, the rule sets of the individual languages can just be put together.

In our polyglot TTS system, multi-context rules are always language-specific, because all constituent names

```

%P set of all phone symbols
%V set of vowel symbols

NGM_F ( * N_F <NR=PL> { } [ ] ADJ_F [ ] * ) :
@/' (z)' <=>
's' '}' ' [' { %P } %V _ ']' ' [' %V ;

NGM_F ( * ADJ_F { } [ ] N_F [ ] * ) :
@/' z' <=>
's' '}' ' [' { %P } %V _ ']' ' [' %V ;
```

Figure 3: Multi-context rules specifying phonological transformations for French liaison: the first rule inserts an optional liaison [(z)] between French plural noun and subsequent French adjective within the nominal part of a French noun group. The second rule inserts a liaison [z] between preceding French adjective and French noun within the nominal part of a French noun group. Both rules have the same graphemic context, i.e. the grapheme “s” preceding a graphemic word boundary ’}’, and the same phonetic context, i.e. two neighboring vowels in front of and after a phonetic word boundary ’]’ ’[’.

```

%V set of vowel symbols

SD_G ( * NGM_F ( ? [ ] * ) * ) :
@/' ?' <=> ' [ ' _ %V ;
```

Figure 4: Multi-context rule specifying a mixed-lingual phonological transformation for glottal stop assignment within French inclusions in German sentences: this rule inserts a glottal stop at the beginning of the first subconstituent of French noun groups only if they are inclusions in a German declarative sentence.

of the syntax tree from the morpho-syntactic analyzer are language specific (all of them have got a language suffix) and thus the language dependence is automatically given by the subtree pattern of the rule.

It has to be mentioned that multi-context rules cannot only be used for phonological transformations as shown in this paper, but they are also well-suited to describe particular (but regular) pronunciations such as specific dialectal or stylistic variants.

## 6. Acknowledgment

This work was partly supported by the Swiss National Science Foundation in the framework of NCCR IM2 and by CTI.

## 7. References

- [1] N. Chomsky and M. Halle. *The Sound Pattern of English*. Harper and Row, New York, 1968.
- [2] C. Traber. *SVOX: The Implementation of a Text-to-Speech System for German*. PhD thesis, No. 11064, Computer Engineering and Networks Laboratory, ETH Zurich (TIK-Schriftenreihe Nr. 7, ISBN 3 7281 2239 4), March 1995.
- [3] R. Sproat. Multilingual text analysis for text-to-speech synthesis. In *Proceedings of the ICSLP'96*, Philadelphia, October 1996.
- [4] B. Pfister and H. Romsdorfer. Mixed-lingual text analysis for polyglot TTS synthesis. In *Proceedings of the Eurospeech 2003*, pages 2037–2040, Geneva, 2003.
- [5] C. Traber, B. Pfister, et al. From multilingual to polyglot speech synthesis. In *Proceedings of the Eurospeech*, pages 835–838, September 1999.
- [6] C. Traber. *Improvements of the Morpho-Syntactic Analysis of the SVOX Text-to-Speech System*. Projektbericht, Institut für Technische Informatik und Kommunikationsnetze, ETH Zürich, Mai 1997.
- [7] K. Koskeniemi. *Two-Level Morphology: A General Computational Model for Word-Form Recognition and Production*. PhD thesis, Department of General Linguistics, University of Helsinki, Finland, 1983.