Mixed-Lingual Text Analysis for Polyglot TTS Synthesis

Beat Pfister and Harald Romsdorfer

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

Abstract

Text-to-speech (TTS) synthesis is more and more confronted with the language mixing phenomenon. An important step towards the solution of this problem and thus towards a so-called polyglot TTS system is an analysis component for mixed-lingual texts. In this paper it is shown how such an analyzer can be realized for a set of languages, starting from a corresponding set of monolingual analyzers which are based on DCGs and chart parsing.

1. Introduction

It has been shown, e.g. in [1], [2] and [3], that high-quality text-to-speech (TTS) synthesis requires a syntactic analysis of the input text for two reasons: Firstly, in order to determine the correct pronunciation of homographs that belong to different word classes (or languages; cf. Section 2.2). Secondly and more important, the syntactic structure is necessary for sentence accentuation and phrase boundary assignment, which is used to derive the sentence prosody.

Our German TTS synthesis system *SVOX*, which comprises such a text analysis component, has been realized some years ago. It produces high-quality speech, as long as the input text is pure German. Real-life texts contain numerous inclusions from foreign languages, however. Therefore, a TTS system which has to read aloud such mixed-lingual texts needs the corresponding capability to analyze the words and sentences and derive the appropriate pronunciation and prosody.

It has to be emphasized that such a polyglot TTS system is not the same as a so-called multilingual TTS system. Existing multilingual TTS systems can be switched to operate in one of several independent language modes, but in general each language is treated by an independent subsystem and synthesized with a language-specific voice. Therefore, these multilingual TTS systems cannot be used for mixed-lingual texts.

In this paper, we present the text analysis component of our polyglot TTS system, which has the same architecture as the original monolingual TTS system *SVOX*. The speech production part of this polyglot TTS system has been shown in [4].

2. Review of the requirements

The requirements of the text analysis component of a polyglot TTS synthesis arise from the texts, of course, which have to be converted into speech. Therefore, we first illustrate the language mixing phenomena typically encountered in published texts and then derive the specifications for the analysis of such texts.

2.1. Types of inclusions

Investigations of various real-life texts, mainly from Swiss newspapers, have shown that inclusions from other languages are quite frequent. Most inclusions are English but many French ones have also been found (see [5]). The size of such inclusions is widely varying and ranges from a part of a word (e.g. a stem) up to a whole phrase.

A small collection of German example sentences with English inclusions is shown in Table 1. These sentences illustrate three major groups of foreign inclusions: (1) There are mixed-lingual word forms that can be produced from English stems by means of German declension or conjugation.[1] Additionally, there exist various mixed-lingual compounds. (2) Full foreign word forms that follow the foreign morphology, but possibly do not syntactically agree. (3) Multi-word inclusions which are syntactically correct foreign constituents.

Inclusion type:	**Example sentence:**
noun stem	Den *Manager*n wird misstraut. [deːn ˈmænɪd͡ʒərn vɪrt mɪsˈtra͡ut]
verb stem	Er *surf*t gerne im Meer. [|eːɐ̯ sɜːft ˈɡɛrnə |ɪm meːɐ̯]
adjective stem	Das ist ein *smart*es System. [das |ɪst |a͡in ˈsmɑːtəs zʏsˈteːm]
compound noun	*Manager*gehälter sind umstritten. [ˈmænɪd͡ʒəɡəˌhɛltɐ sɪnt |ʊmˈʃtrɪtən]
noun(s)	Die *Fans* lieben ihr *Team*. [diː fænz ˈliːbən |iːɐ̯ tiːm] Der *Laser* ist eine Lichtquelle. [deːɐ̯ ˈle͡ɪzə |ɪst ˈ|a͡inə ˈlɪçtˌkvɛlə]
uninflected adj.	Sie ist *happy*. [ziː |ɪst ˈhæpɪ]
proper name	Der Konkurs von *Swiss Dairy Food* ist noch nicht abgewendet. [deːɐ̯ kɔnˈkʊrs fɔn swɪs ˈde͡ərɪ fuːd |ɪst nɔx nɪçt ˈ|apɡəˌvɛndət]
noun group	*Human Touch* kommt an. [ˈhjuːmən tʌt͡ʃ kɔmt |an]

Table 1: *Examples of various English inclusions (in italics) in sentences of the base language German*

[1] There are even more exotic mixed-lingual words such as "*out*ge*sourc*t", i.e., some German past participle construction of the English verb "*to outsource*". Even such special forms are correctly analyzed in our system. Although they occur relatively rarely, there is a clear tendency, however, that such forms are getting more and more frequent and common and thus have to be processed appropriately.

Reprint from Proceedings of Eurospeech 2003, September 1-4, Geneva, Switzerland

2.2. Foreign language inclusions vs. loan words

We have to distinguish between inclusions from other languages as shown in Table 1 and loan words. The latter are strongly assimilated to the base language, not only in morpho-syntactic terms, but also with respect to the pronunciation, whereas the former roughly keep their foreign pronunciation.

Loan-words in mixed-lingual text may, however, raise an additional issue concerning homographs in places where their pronunciation depends on the language context. Consider e.g. the word “argument” that can either be pronounced in German as [argʊˈmɛnt] or in English as [ˈɑːgjumənt] or in French as [aʀgym ã].

The solution to this problem is outlined in Section 4.2.

2.3. Consequences for the text analysis

Clearly, a TTS synthesis system that has to pronounce sentences as shown in Table 1 must include a morphological and syntactic analyzer in order to process such inclusions appropriately. This means:

First of all, a mixed-lingual text analysis can hardly be based on a lexicon with full word forms only. The number of possible mixed-lingual word forms is huge, because almost arbitrary combinations of stems, endings and prefixes have to be considered.[2] Therefore, a mixed-lingual text analysis must include a morphological analysis component.

Secondly, foreign inclusions often do not meet the generally required syntactic agreement of the base language. In the fifth sentence in Table 1, e.g., the gender of the English noun *“Laser”* (neuter) and the gender of the German masculine article “der” do not match.

Thirdly, the type of morphemes, the word classes, the constituents and their number and values of the morpho-syntactic features differ largely between individual languages. A superset of classes and features, even for two languages only, is impractical. The same holds for the word and sentence grammars.

Additionally, homographs are much more frequent in mixed-lingual texts than in monolingual ones (see Section 2.2).

3. Approach to mixed-lingual text analysis

3.1. The *SVOX* text analysis

The *SVOX* TTS synthesis system has got a modular system concept, where voice-independent parts like the morpho-syntactic analysis, the sentence accentuation, the prosodic phrasing, etc. are strictly separated from voice-dependent parts, i.e., from all speech signal related processing. These two main parts are called transcription and phono-acoustical model, resp.

The transcription comprises a morphological and syntactic analyzer, realized as a bottom-up chart parser, plus a subsequent rule-based phonological generation module, determining sentence accent levels and prosodic phrase boundaries. The morphological analysis is either a lookup in the full-form lexicon[3] or, if not found, the word is decomposed into correct sequences of lexemes from the morpheme lexicon that contains stems, endings and prefixes. A sequence of morphemes is correct, if it meets the restrictions imposed by the word grammar. After the word analysis, the sentence structure is determined by means of the sentence grammar. All grammar rules and lexical entries are written in a penalty-extended DCG (definite-clause grammar) formalism. Additionally, two-level rules mediate between lexical and surface forms (cf. [2]).

```
VS_G (V1,A,NON)        "web"   ['ve:b]
VE_G (V1,A,IND,PRES,NON,S3) "t#"  [t#]
AS_G (P,K,S)           "weit+" ['va_it+]
NS_G (SC3,PC7,N)       "netz+" ['nEt_s+]
NES_G (SC3,G,S3)       "es#"   [@s#]
NEP_G (PC7,N,P3)       "e#"    [@#]

PRON_G (N,S3,S3,M)     "er"    ['?e:^6]
PREPC_G (D,S3,M,D,NON,L) "im"  ['?Im]
```

Table 2: *Some entries of the German morpheme (top) and the full-form (bottom) lexicon with graphemic and SAMPA-like phonetic representation*

3.2. From mono to mixed-lingual text analysis

From the considerations in Section 2.3 we conclude that a practical approach to mixed-lingual text analysis for a certain set of languages $\{L_1, L_2, L_3, \dots\}$ must be achieved in two steps: First we have to design the corresponding set of monolingual analyzers. Each monolingual analyzer includes its own lexica (i.e. a full-form and a morpheme lexicon) and its own word and sentence grammars. It is important to note, that for all the grammars the same DCG formalism is used (see Section 4). This allows to apply the same chart-parser for all of these monolingual analyzers.[4]

In the second step, we have to design for each language pair $\{L_i, L_j\}$ a so-called inclusion grammar. The inclusion grammar G_{ij} defines the elements of language L_j that are allowed as foreign inclusions in language L_i. In order to get a mixed-lingual analyzer for the languages L_i and L_j, we have to load the lexica and the grammars of these languages and additionally the inclusion grammars G_{ij} and G_{ji}.

4. The implementation

4.1. Lexica and grammars for individual languages

Since for obvious reasons[4] all grammatical names (i.e. non-terminals) have to be unique, not only within single languages, but also across the languages, all names have been taged with a language suffix. Appart from this, the naming is completely unconstrained. In order to improve readability, however, some standard names have been introduced for all grammars and lexica (e.g. VS stands for verb stem and the suffixes _G and _E distinguish between German and English, resp.).

Comparing entries of the German and English lexica (Tables 2 and 3) exhibits such differences: e.g. while German noun

[2]In contrast to monolingual grammars, no rules exist that restrict the combination of stems, prefixes and suffixes with foreign stems, prefixes and suffixes. Despite most English verb stems are declined according to the weak German verb class and thereby restricting possible verb endings, still combinations like “ge*scann*t”, “ge*scan*t”, “ge*scanned*”, even “einge*scann*t”, “durchge*scann*t”, “über*scann*t”, etc. are variants of the English past participle *“scanned”* that can be found in German texts. Furthermore and even worse, there is a virtually unlimited number of compound words with several stems possibly from different word classes and languages.

[3]This full-form lexicon contains mainly grammatical or function words and some irregular forms.

[4]If the names of all word classes and all rule heads are disjoint, we can load the lexica and the grammars of several languages and get a multilingual analyzer for these languages, i.e., an analyzer for sentences from any of these languages. Note that such an analyzer is a language detector at the same time.

```
VS_E (S,PRES)          "surf"  ['s3:f]
VE_E (S,PRES,IND,PRES,S3) "s#"    [z#]
AS_E (TYPE4)           "wide"  ['wa_Id]
ASE_E (TYPE4,POS)      "#"     [#]
NS_E (NC3,N)           "web"   ['web]
NS_E (NC3,N)           "world" ['w3:ld]
NE_E (NC3,S3)          "#"     [#]
NE_E (NC3,P3)          "s#"    [z#]
```

Table 3: *Entries of the English morpheme lexicon*

stems NS_G embody three morpho-syntactic features (singular class, plural class and gender), English noun stems NS_E have got only two (noun class and gender). Accordingly, the German lexicon lists the singular and the plural noun endings as two different types of lexical entries (NES_G and NEP_G), whereas the English lexicon contains only one type of noun endings NE_E.[5]

Likewise, the grammars are independent, which can easily be seen from the Tables 4 and 5, that show some rules for German and English verb forms, resp.

```
V_G (?PERS,?MOOD,?TENSE,?IMP,?) ==>
    VST_G (?VCL,?VSTYP)
    VE_G (?VCL,?VSTYP,?MOOD,
          ?TENSE,?IMP,?PERS) * 1

VST_G (?VCL,?VSTYP) ==>
    REP_SIMPX_G (?)
    VSIMP_G (?VCL,?VSTYP) * 1

VSIMP_G (?VCL,?VSTYP) ==>
    REP_BVPR_G ()
    VS_G (?VCL,?VSTYP,?) * 1

REP_BVPR_G () ==> * 0  :INV
REP_BVPR_G () ==>
    BVPR_G ()
    REP_BVPR_G () * 10  :INV

REP_SIMPX_G (SEP) ==> * 0  :INV
REP_SIMPX_G (SEP) ==> HYPHEN () * 0
```

Table 4: *Rules from the German word grammar*

The grammar rule penalty values, the integer values behind the “*”, as shown in Table 4, are used to select the optimal solution out of a number of ambiguous solutions.[6] These penalty values have been set manually by a linguistic expert.

4.2. The inclusion grammar

In order to analyze sentences as listed in Table 1, an additional grammar, that defines which English inclusions can occur in German sentences, has to be added to the analyzer. This inclusion grammar G_{GE} is shown in Table 6. It consists of rules that map a foreign constituent and its features to the corresponding constituent of the base language. E.g., the third rule defines that a German noun N_G with a certain number ?N and gender ?G can be replaced by an English noun with the same number and gender. The fourth rule states that this can also happen without matching genders, but at a higher penalty. The rule penalty values have to be chosen such that interlingual ambiguities are handled correctly.

[5]This is due to individual preferences of the grammar writers.

[6]It is important to note, that a full parse is always performed which keeps all ambiguities in the chart. Only the final sentence solution is selected from the chart according to the minimum accumulated penalty criterion.

```
V_E (?MOOD,?TENSE,?PERS) ==>
    BVPR_OPT_E ()
    VS_E (?VT,?STEM)
    VE_E (?VT,?STEM,?MOOD,
          ?TENSE,?PERS) * 1

BVPR_OPT_E () ==> * 1 :INV
BVPR_OPT_E () ==> BVPR_E () * 1
```

Table 5: *Rules from the English word grammar*

Due to the fact that the constituents, which are mapped by an inclusion grammar rule, often have different feature sets, unification is not always possible. It is still possible, however, to provide a *feature mapping* within the inclusion grammar rules. E.g., the rules mapping an English noun stem NS_E to a German noun stem NS_G map also the number of features, and the rules for the inclusion of noun groups adapt the positions of the features. Furthermore, restrictions can be specified as shown in the seventh rule: only English present verb stems may be included and they have to be conjugated according to the German verb class V1 which means weak conjugation.

```
ADJU_G (?) ==> AJH_E (?) * 100

AS_G (?,?,?) ==> AS_E (?) * 150

N_G (?,?N,?G) ==> N_E (?N,?G) * 100
N_G (?,?N,?) ==> N_E (?N,?) * 120

NS_G (?,?,?G) ==> NS_E (?,?G) * 150
NS_G (?,?,?) ==> NS_E (?,?) * 170

VS_G (V1,A,?) ==> VS_E (?,PRES) * 150

NG_G (?,?N,?G) ==> NG_E (?N,?,?G) * 200
NG_G (?,?N,?) ==> NG_E (?N,?,?) * 300

NGN_G (?,?N,?G) ==> NG_E(?N,?,?G) * 250
NGN_G (?,?N,?) ==> NG_E(?N,?,?) * 350
```

Table 6: *Grammar G_{GE} defines the possible English inclusions in German sentences.*

It is easily seen that the inclusion grammar principle provides also an appropriate solution to the problem of interlingual homographs, no matter if they are due to loan words or not. The rules of the inclusion grammars basically allow all variants of a homograph, but the high penatly values prioritize that one which matches the language of the including constituent.

It has to be mentioned that generally inclusion grammars contain some 20 rules only and thus are very compact compared to grammars of individual languages which need typically more than 500 rules.

5. Results

In order to get an analyzer for German sentences with English inclusions, the German and the English lexica and grammars have to be loaded together with the inclusion grammar G_{GE}. Now the analyzer is ready to process a sentence like ”Er *surf*t im *World Wide Web*.” (”He surfs the world wide web.”)

The analysis result, i.e. the syntax tree of this sentence is shown in Figure 1 as indented list. Each node is labelled with the constituent name, the names and values of the features, and the assigned penalty value. Terminal nodes additionally contain the grapheme and phoneme strings. By means of the lexical entries and the grammar rules given in the Tables of Section 4, the verb branch V_G of this tree can easily be verified.

```
S  461
+- DS_G  460
   +- NG_G  <C N> <N S3> <G M>  6
   |  +- NG_SIMP_G  <C N> <N S3> <G M>  5
   |     +- PRON_G "er" ['?e:^6] <C N> <N S3> <PERS S3> <G M>  1
   +- V_G  <PERS S3> <MOOD IND> <TENSE PRES> <IMP NON> <REF REF>  155
   |  +- VST_G  <VCL V1> <VSTYP A>  153
   |  |  +- VSIMP_G  <VCL V1> <VSTYP A>  152
   |  |     +- VS_G  <VCL V1> <VSTYP A> <REF ?_0>  151
   |  |        +- VS_E "surf" ['s3:f] <VCL S> <VSTYP PRES>  1
   |  +- VE_G "t#" [t#] <VCL V1> <VSTYP A> <MOOD IND> <TENSE PRES> <IMP NON> <PERS S3>  1
   +- AN_G  <ANG ANG>  289
      +- PG_G  <C D> <N S3> <G N>  288
         +- PG_SIMP_G  <C D> <N S3> <G N>  287
            +- PREPC_G "im" ['?Im] <C D> <N S3> <G N> <R D> <AG NON> <KO L>  1
            +- NGN_G  <C ?_0> <N S3> <G N>  284
               +- NG_E  <NUM S3> <PRON NON> <GEN N>  34
                  +- MOD_REP_E  21
                  |  +- MOD_REP_E  11
                  |  |  +- MOD_E  8
                  |  |     +- N_E  <NUM S3> <GEN N>  4
                  |  |        +- NS_E "world" ['w3:ld] <NCL NC3> <GEN N>  1
                  |  |        +- NE_E "#" [#] <NCL NC3> <NUM S3>  1
                  |  +- MOD_E  8
                  |     +- ADJ_E  <GRADE POS>  7
                  |        +- AJ_E  <GRADE POS>  4
                  |           +- AS_E "wide" ['wa_Id] <ADJTYP TYPE4>  1
                  |           +- ASE_E "#" [#] <ADJTYP TYPE4> <GRADE POS>  1
                  +- N_E   <NUM S3> <GEN N>  4
                     +- NS_E "web" ['web] <NCL NC3> <GEN N>  1
                     +- NE_E "#" [#] <NCL NC3> <NUM S3>  1
```

Figure 1: *Syntax tree from the analysis of the sentence:* ”Er *surf*t im *World Wide Web*.” *It shows a German declarative sentence with two English inclusions: the English verb stem “surf” has been accepted as a German one at a penalty of 150 (lines 9 and 10) and the English noun group “World Wide Web” has got a penalty of 250 (lines 16 and 17).*

Most interestingly, despite most words of this sentence are English, it is still correctly analyzed as a German sentence with the correct syntactic structure. In particular, the English verb stem `VS_E` and the final English noun group `NG_E` have been correctly identified. The graphemic ambiguity of ”web” (it can be a German verb stem `VS_G` as well as an English noun stem `NS_E`) has also been resolved correctly. Moreover, the analysis not only detects the correct sentence base language but even marks each morpheme with the corresponding language.

6. Conclusions

In this paper, we have presented a new and very accurate analyzer for mixed-lingual German and English sentences while maintaining a strict separation of the databases for each language. This type of morpho-syntactic analyzer meets exactly the requirements of a polyglot TTS system that has to pronounce mixed-lingual text in the “Swiss manner”, which means: foreign inclusions in German sentences are generally not assimilated to the base language. Each inclusion, even if it is a part of a word only, is pronounced according to its originating language instead. With our approach to morpho-syntactic analysis we achieve both, precise language detection and accurate structure determination.

The first version of this mixed-lingual analyzer copes already with fairly tricky sentences. But both, lexica and grammars need further refinement. Furthermore, we have started to include additional languages, primarily French and Italian.

7. Acknowledgement

This work was partly supported by the Swiss National Science Foundation in the framework of NCCR IM2. The authors gratefully acknowledge also the support of the Swiss authorities within COST 258.

We also thank Christof Traber for his valuable suggestions.

8. References

[1] R. Sproat. Multilingual text analysis for text-to-speech synthesis. In *Proceedings of the ICSLP’96*, Philadelphia, October 1996.

[2] C. Traber. *SVOX: The Implementation of a Text-to-Speech System for German*. PhD thesis, No. 11064, ETH Zurich (ISBN 3 7281 2239 4), March 1995.

[3] M. Riedi. *Controlling Segmental Duration in Speech Synthesis Systems*. PhD thesis, No. 12487, ETH Zurich (ISBN 3-906469-05-0), February 1998.

[4] C. Traber, B. Pfister, et al. From multilingual to polyglot speech synthesis. In *Proceedings of the Eurospeech*, pages 835–838, September 1999.

[5] B. Pfister, E. Wehrli et al. *Lexical and Syntactic Analysis of Mixed-Lingual Sentences for Text-to-Speech*. Final Report of SNSF Project No 21-59396.99. Institut TIK, ETH Zürich, November 2002.