Cross-Language Phonemisation In German Text-To-Speech Synthesis

Jochen Steigner [1], *Marc Schröder* [2]

[1] Intelligent User Interfaces Lab
[2] Language Technology Lab
DFKI GmbH, Saarbrücken, Germany
{[1]steigner, [2]schroed}@dfki.de

Abstract

We present a TTS component for transcribing English words in German text. In addition to *loan words*, whose form does not change, we also cover *xenomorphs*, English stems with German morphology. We motivate the need for such a processing component, and present the algorithm in some detail. In an evaluation on unseen material, we find a precision of 0.85 and a recall of 0.997.

Index Terms: speech synthesis, text-to-speech, phonemisation, loan words, xenomorphs, cross-language

10.21437/Interspeech.2007-531

1. Introduction

Interactive NLP systems like TTS have attained high levels of quality in recent years. Nevertheless, non-native words represent a major difficulty. In an increasingly internationalised world, TTS systems, and also language recognisers, should be designed to cover non-native items in current languages in order to meet the expectations of users towards a high quality NLP system. In the present paper, we present an approach for the phonemisation of non-native words within the framework of the German TTS synthesis system MARY[1]. As everyday experience suggests that English represents the language with the strongest influence on current spoken German, we decided to focus our research on the phonemisation of *anglicisms*.

The paper is organised as follows. First, we give a definition of the term *anglicism* and illustrate how anglicisms can be categorised. Then we present a corpus study that was used to collect examples of anglicisms in German. We describe the algorithm for the pronunciation of English and cross-language words implemented in the MARY TTS system. Finally, an evaluation is presented that illustrates the performance of this strategy on unseen material.

2. Definitions

Anglicisms are words that contain one or more lexemes originating from American English or British English, used in a language that is not English. This definition does not distinguish between words that stem from British English and words stemming from American English. We subdivide anglicisms into three categories:

1. **Loan words** are lexemes borrowed from English that are used in unmodified form (neither morphologically nor semantically) in another language. The word "Homepage" can be taken here as an example that is relatively new to the German language. On the other hand, "Manager" is an example that has existed in German linguistic usage for many years, as well as the loan word "Jeans".
2. **Xenomorphs** are words that are built from combinations of English morphemes and morphemes from the respective native language [6]. The morphology of xenomorphs is governed by the morphological rules of the native language. Numerous examples belonging to this group have entered the German language within the last years, mainly influenced by the distribution of the World Wide Web. We consider "downloaden" (*to download*) as one of the most popular examples for this group. We also categorise compounds like "Marketingabteilung" (*marketing department*) or "Outdoorkollektionen" (*outdoor collections*) as members of this group.
3. **Pseudo Anglicisms** are words in German linguistic usage that have English phonology and morphology but are not used in their original meaning. Probably the most popular example of this category is the word "Handy", which is the German term for 'mobile phone'. "Smoking" represents an older example of a pseudo anglicism – it has the meaning of 'dinner jacket' in German.

Language purists can become very emotional when confronted with the use of anglicisms in German, claiming that they are not "real German". For the purpose of speech synthesis, we can leave that normative question unanswered. Instead, we follow a *descriptive linguistics* mindset: if people use anglicisms, our TTS system should be able to pronounce them.

3. An Informal Corpus Investigation

For the purpose of collecting examples of anglicisms that are used in German, we built a corpus of approximately 2,500,000 words from the online forum of the German computer magazine *Chip* (http://www.chip.de). Our idea was that in a forum of a computer magazine, users talk about information technology, computers and related issues. Thus, necessarily a lot of anglicisms like loan words and xenomorphs will be used in the postings of the users. Moreover, we assumed that the language used in the forum would more closely resemble spoken dialog than language used in other web sources like reports, news or technical descriptions. Users normally make their postings quickly, and thus behave more like in a spoken language discourse than it would be the case when writing a report or any other "official" content for a website.

We used the pronunciation lexicon of the MARY TTS system to filter out known words. Out of the unknown tokens, about 1900 occurred three or more times. Indeed, a high proportion of unknown tokens were *loan words* and *xenomorphs* (approximately 25%). The majority of non-native items were loan words of English (approximately 90%). The xenomorphs found in the corpus exhibited a large variety of different forms, includ-

ing inflected verbs with an English stem like “downloaden” (to download), as well as compounds like “Mehrfachpostings”.

4. The Phonemisation of Loan Words

The work of [2] and [3] demonstrates that speakers of German include *xenophones* (non-native phones) to a quite high degree when pronouncing loan words in spoken language. The studies suggest that these xenophones that were used in pronunciation by almost all speakers need to be integrated into the phoneme inventory of a TTS system. The phonemisation of loan words in MARY involves two steps:

- The *retrieval* of the pronunciation for the loan word from an English lexicon.
- The *mapping* of the pronunciation to the MARY phoneme inventory including xenophones.

4.1. Lexicon Retrieval

In MARY TTS, we have the facility to access the CMU US English lexicon which is part of the FreeTTS speech synthesis [4]. Using this lexicon, we can retrieve the transcriptions for English words in German text.

4.2. Pronunciation of Loan Words

For the purpose of appropriately pronouncing English loan words, the German components of the MARY system were equipped with the xenophones depicted in Table 1.

θ	**Th**read
ð	Bro**th**er
ɹ	**r**emote
w	**W**eb
eɪ	M**ai**l
ʤ	**J**ob

Table 1: English phones in the German MARY phone set.

This set of xenophones overlaps to a high degree with the xenophones that are suggested to be included in the phoneme inventory of a German TTS [2]. The German MARY voices are capable of using these xenophones in the pronunciation of loan words. Xenophones that are not part of the MARY phoneme inventory (e.g., the [ʌ] in ‘b**u**t’ or the [æ] in ‘tr**a**p’) are mapped to phonetically similar German phones or phone sequences.

It should be noted that these mappings inevitably result in slightly different pronunciations. But the studies of [2] show that not every xenophone that is contained in a loan word is actually realised by German native speakers. The findings of [2] illustrate that speakers prefer some xenophones to be realised in their nativised variant. Moreover, it is suggested that choosing a too high level of non-native pronunciation could lead to a rejection of the synthesised output by the user because the output sounds somewhat conceited or high-browed.

Using the described technique, an appropriate pronunciation for loan words can be achieved in MARY. The majority of anglicisms taken from our corpus investigation can be processed in this way. If the input word is not a loan word according to our definition and therefore cannot be retrieved from the English lexicon, it is handed over to the *Xenomorph Phonemiser* which we illustrate in the next section.

5. The Phonemisation of Xenomorphs

In Section 4 we illustrated how the pronunciation of loan words is established within MARY by using an English lexicon, xenophones and mappings of xenophones to German phones. In this section, we present an approach that makes use of this technique in the phonemisation of *xenomorphs*.

5.1. Xenomorphology

[5] implemented a lexical component capable of handling xenomorphs in Swedish in a two-level formalism consisting of a transcribed lexicon with a morphophonological description of Swedish nouns and a unification based grammar formalism. According to their study, a lexical TTS component designed to solve the task of pronouncing xenomorphs should be capable of handling foreign sounds, inflected xenomorphs, xenomorph compounds and interactions between foreign items and native prosody. Therefore, the pronunciation of xenomorphs represents a more complex task than the pronunciation of loan words as described in the previous section. It can be seen from the study of [5] that xenomorphs can be subdivided into

- Inflections with an English stem,
- and compounds or derivations containing English words.

Analogically, the Xenomorph Phonemiser implemented in MARY has a two-level architecture consisting of an *Inflectional Analysis* and a *Compound Analysis* component. These are described in the following.

5.2. Inflectional Analysis

The first stage in the xenomorph phonemisation process is the inflectional analysis. Input graphemes are investigated here with respect to German inflection endings. The inflectional analysis consists of three successive steps:

1. *Inflection Search:* The input grapheme is searched for inflectional endings.
2. *Morphological Analysis:* The morphological analysis tries to derive an English root morpheme from the input grapheme. Additionally, information about the morphological composition of the input is achieved by morphological tests.
3. *Composition of surface transcription:* The surface transcription is built from the decomposed morphemes according to the morphological information found by the morphological analysis.

5.2.1. Inflection Search

The inflection search algorithm tries to find a German inflection ending as a suffix of the input grapheme by employing a *longest match* strategy. The endings used by our analysis are represented in an ending lexicon. If the search for an inflection ending succeeds, it is stored and removed from the rest-word. The rest-word is handed over to the morphological analysis. To illustrate our algorithm, we refer to the example of the xenomorph *“connectenden”* (present participle with accusative ending of “to connect”). In the inflection search, the suffix ’**en**’ is identified as a legal German inflection ending in the input grapheme. For the morphological analysis, the suffix ’en’ is then removed from the input word.

If no inflection ending can be found in the input grapheme, it is also transferred to the morphological analysis.

5.2.2. Morphological Analysis

It should be noted that at this point we still do not know if the input grapheme actually represents a xenomorph. Two tasks have to be carried out at this stage of processing:

1. Derive an English root morpheme from the remaining input word.
2. Determine the morphological composition of the word as a basis for the surface transcription.

In the case of our example "connectenden", we have successfully found the inflection ending **en**. The rest-word ("connectend") is then looked up in the English lexicon. As this token cannot be found in the English lexicon, a test is performed to determine the morphological form of the input. For our example, we test if the suffix of the token matches the present participle suffix **end**. If the test succeeds, the suffix **end** is removed from the token and the remainder is again looked up in the English lexicon. This time, the lookup is successful – we have found a valid English root morpheme. The knowledge about the morphological form of the input grapheme is then stored as a source of information for the composition of the surface transcription.

The main idea with the tests is to examine the input token for orthographical variants that avoid a direct retrieval of the input from the English lexicon. The mere removal of inflection endings may not be sufficient. For example, imagine the English infinitive "to scan". Within our data, we found a nativised variant of the root morpheme 'scan' that was modified to end on a geminate ("sca**nn**en"). When the inflection ending 'en' is removed from the input word, the morphological analysis fails to retrieve the rest-token on the English lexicon. Therefore, a test is carried out that investigates the token for ending on a geminate. If this test succeeds, the second 'n' is removed from the token and again, the remainder is looked up in the English lexicon. When the lookup succeeds, the token can be transferred to the composition of the transcription. If no root morpheme can be derived by the morphological analysis, the inflectional analysis is started again with the length of the separated suffix reduced by one. This is useful in some cases when the ending search has matched a substring that partially consists of letters from the root morpheme (e.g., updat**et**). If this also fails, the input word is transferred to the *Compound Analysis*.

5.2.3. Compose Surface Transcription

When the Morphological Analysis has successfully derived a root morpheme, the processing of the input word finishes with the composition of the surface transcription. Here, the transcription of the root morpheme is taken from the English lexicon and the transcription of the ending is retrieved from the ending lexicon. The surface transcription is built from the partial transcriptions with respect to German syllabification rules.

The morphological information provided by the morphological analysis is used in the composition of the surface transcription. In the case of the input word "connectenden", we have to include the transcription for the present participle suffix **end** between the transcription of the root morpheme and the transcription of the ending **en**. The transcriptions for the particular morphemes are retrieved from the lexicons and the surface transcription is built with respect to German syllabification rules: /kə.n'ɛk.tən.dən/.

5.3. Compound Analysis

If no inflected form can be identified by the Inflectional Analysis, the input word is processed further by the *Compound Analysis* module. This component attempts to identify maximum-length prefixes for which a phoneme transcription is known. For this purpose, prefixes of decreasing length are looked up, in sequence, in a dedicated German prefix lexicon, in the standard German lexicon, and in the English lexicon. When a known prefix is found, it is split off, and the remainder of the input word is looked up first in the German, then in the English pronounciation lexicon. If found, the transcriptions of the two parts are concatenated to provide the full phoneme string.

If the remainder is not found in one of the lexicons, a number of additional analysis steps are attempted. Any German noun declination endings are detached from the end of remainder before a renewed lexicon lookup (e.g., **n** as in "Internetprovidern", the dative form of *internet providers*, or **s** as in "Downloadfensters", the genitive form of *download window*). Similarly, any letters corresponding to potential *fuge morphemes* (semantically meaningless filler morphemes between some German compounds: **-es-**, **-er-**,**-en-**, **-e-**, **-n-**, **-s-**) are detached from the head of the remainder.

If an analysis could still not be found after these steps, the remainder is analysed by the *Inflectional Analysis* module described above. At this stage, words like "einchecken" (*to check in*) can be analysed. It is composed of the German prefix "ein-" (*in*), found as a prefix in the prefix lexicon, and the English stem "check" with the German infinitive ending "-en" (infinitive), which can be analysed by the Inflectional Analysis.

The next escalation step is the recursive application of the *Compound Analysis* to the remainder. An example of a word which can be analysed in this way is "downgeloadet" (past participle of *to download*). The English adjective "down" is identified as a known prefix; in the recursive application of the compound analysis of the remainder, "ge-" is identified as a German prefix, and "-load+et" is analysed by the Inflectional Analysis.

If none of these analyses yields a result, the procedure is repeated with the next shorter known prefix of the input word. If no analysis can be found, letter to sound rules are applied.

5.4. Stress Assignment

Loan words come with stress markers when retrieved from the English lexicon. They do not need to be modified by a subsequent stage of processing. The same holds for inflected xenomorphs as well as derivations like the past participle, where the English root morpheme is retrieved from the lexicon including stress markers. German inflection endings and derivation affixes cannot be stressed, therefore the surface transcription inherits the stress of the English root morpheme.

For compound xenomorphs and German compounds, any morpheme that is retrieved from the English or German lexicon is stressed on one syllable. Compounds that consist of two root morphemes inherit the stressing of the first root morpheme [8]. Compounds that consist of more than two root morphemes can also inherit the stress of the first root morpheme in certain cases – exceptions to this case are given in [8]. As these exceptions would represent an object for further research, we employed a simple rule for the stress assignment of compounds: If a compound transcription has more than one stress marker, the left-most one is taken over to the surface transcription.

6. Evaluation

The performance of an NLP system can be measured in terms of *precision* and *recall* [7]. To measure the precision and recall of the Xenomorph Phonemiser, another corpus of roughly 80000 words was collected in the way described in section 3.

Our test corpus contained 1027 xenomorphs. Ignoring misspellings, the xenomorph phonemiser returned transcriptions for 1202 words from the test corpus. From that list, 178 words have been counted as not being xenomorphs. This results in a total of 1024 xenomorphs that were correctly recognised and transcribed by our system. The xenomorph phonemiser failed to recognise 3 words that can actually be categorised as xenomorphs according to our definition. Among the 178 words that were falsely recognised and transcribed by the system were mainly

- German words that could not be found in the German lexicon but were falsely composable from English and German morphemes by the Compound Analysis.
- Proper names and brand names not contained in the lexical sources.
- Abbreviations that were not identified by a preceding module in MARY.

Based on the results presented above, the following performance scores were calculated for the xenomorph phonemiser:

$$Precision = 0.852$$
$$Recall = 0.997$$

Using these scores, we can calculate a harmonic mean of

$$F = 0.919$$

for the xenomorph phonemiser.

7. Discussion

The xenomorph phonemiser was tested on numerous English root morphemes such as *download, upload, boot, scroll, scan, mail*. The algorithm successfully computed pronunciations for these English morphemes appearing in all conjugational and derivational forms presented in Section 5.1.

As there is no normative phonology that determines the pronunciation of xenomorphs, the pronunciations produced by the algorithm cannot easily be judged in terms of *correctness*. A manual investigation of the transcriptions evidenced the ability of our approach to produce *appropriate* pronunciations for a large variety of phenomena.

Since we pursue the objective to enhance the quality of a TTS system, it can still be asked to what extent the transcriptions derived by our system fit the requirements of a text-to-speech synthesiser. As far as the integration of xenophones is concerned, only those xenophones are used by MARY that were shown in the literature to be highly accepted by native speakers of German. However, a listening acceptance test would shed light on the aspects of sound quality with and without our xenomorph phonemiser.

8. Conclusion

In this paper, we have presented a method for cross-language phonemisation in German TTS. We have shown how our algorithm covers English loan words and xenomorphs combining an English word stem with German prefixes, suffixes or compound words. In an evaluation, we have shown that the recall of our algorithm is close to 100%, but that the algorithm currently provides a non-negligible number of false positives. In a production system, it will be desirable to shift the precision vs. recall balance to a state where less false positives are found, e.g. by requiring some minimal length of English words in the compound analysis, or by consulting lists of proper names before starting a compound analysis.

9. Acknowledgements

Preparation of this paper was partially supported by the EU projects TALK (IST-507802) and HUMAINE (IST-507422).

10. References

[1] Schröder, M. and Trouvain, J., "The German text-to-speech synthesis system MARY: A tool for research, development and teaching", International Journal of Speech Technology, 6, 2003.

[2] Abresch J. and Breuer S., "Assessment of Non-native Phones in Anglicisms by German Listeners", Proc. ICSLP, 2004.

[3] Abresch, J., "The pronunciation of Anglicisms and English Proper Names in German: A Corpus Study", Proc. of the 16th ESSP, 2005.

[4] Walker, W. and Lamere, P. and Kwok, P., "FreeTTS - A performance case study", http://freetts.sourceforge.net, 2002.

[5] Lindström, A. and Kasaty, A., "A two-level approach to the handling of foreign items in Swedish speech technology applications", Proc. ICSLP, 2000.

[6] Lindström, A. and Eklund, R., "Xenophenomena: Studies of foreign languge influence at several linguistic levels", Proc. of 24. Jahrestagung der Deutschen Gesellschaft fr Sprachwissenschaft, 2002.

[7] Melamed, I.D. and Green, R. and Turian, J.P., "Precision and Recall of Machine Translation", HLT-NAACL, 2003.

[8] Wagner, P., "Vorhersage und Wahrnehmung deutscher Betonungsmuster", PhD Thesis, Rheinische Friedrich-Wilhelms-Universität Bonn, 2002.