International Journal of Computational Linguistics and Applications vol. 7, no. 1, 2016, pp. 161–173
Received 08/02/2016, accepted 07/03/2016, final 20/06/2016
ISSN 0976-0962, http://ijcla.bahripublications.com

Grapheme-to-Phoneme Transcription in Hungarian

Attila Novák[1] and Borbála Siklósi[2]

[1] *MTA-PPKE Hungarian Language Technology Research Group, Hungary*
[2] *Pázmány Péter Catholic University, Hungary*

Abstract

A crucial component of text-to-speech systems is the one responsible for the transcription of the written text to its phonemic representation. although the complexity of the relation between the written and spoken form of languages varies, most languages have their regular and irregular phonological set of rules. In this paper, we present a system for the phonemic transcription of Hungarian. Beside the implementation of rules describing default letter-to-phoneme correspondences and morphophonological alternations, the tool incorporates the knowledge of a Hungarian morphological analyzer in order to be able to detect compound and other morpheme boundaries, and it contains a rich lexicon of entries with irregular pronunciation. It is shown that the system performs well even on texts containing a high number of foreign names.

1 Introduction

In the research reported about in this study, our goal was to implement a system that can automatically transform written Hungarian to its phonemic representation. The original intent of the system was to transcribe a database of Hungarian geographic terms. However, due to certain design decisions, our system proved to perform well also on texts containing a high ratio of foreign names and suffixed forms.

Even though units in a written alphabet might correspond to a phonemic unit of the spoken language, the complexity of this mapping varies among languages. Even if we consider only languages using the Latin alphabet, there are significant language-specific differences. The applicability of certain methods depends on both morphosyntactic and phonological characteristics and the type of correspondence orthography and phonology has in the given language.

In English, orthographic standards were fixed quite early, while its sound system has further evolved [1]. Thus, it is often quite difficult to predict the correspondence between written and spoken forms. However, since the number of word forms is limited, either a manually created, or an automatically generated lexicon – containing both written and transcribed word forms – can cover almost the whole vocabulary of the language. The main problem in English is (in addition to eventual OOV items, like names) massive homography with items belonging to different part of speech often having different pronunciation.

In the case of some other languages, such as Hungarian, the relation between written and spoken forms is much closer; the orthography is basically phonemic. In most cases, pronunciation is predictable from the orthographic form. Still, there are many general and exceptional assimilation phenomena many of which are conditioned by morphological structure. Moreover, agglutination yields a huge number of word forms, making the inclusion of the full vocabulary in a lexicon impossible [2].

Thus, lexical lookup must be combined by some procedural method mapping orthographic strings to sequences of phonemes, exploiting processing capabilities instead of just storing large amount of offline data in the form of lexicons.

The structure of this paper is as follows: In Section 2 related approaches are overviewed briefly. Then, our method for transcription is described, including detailed arguments about the language-specific difficulties of Hungarian phonology. Finally, an evaluation of our tool is presented with an error analysis of the most significant errors revealed during the experiments.

2 Related Work

There are three main branches of grapheme-to-phoneme transcription methods [3]:

- dictionary look-up,
- rule-based approaches,
- data-driven approaches.

Dictionary look-up is used when the mapping between the orthographic and phonological representation is based on conventions, and rules or generalizations are not applicable. The advantage of such methods is that other information (e.g. lexical stress, part-of-speech) can also be stored in the dictionary. However, the creation of such dictionaries by hand is very expensive and tedious.

No matter how limited the agglutinating behaviour of a language is, there will always be new words or word forms, which are not covered by a predefined lexicon. Rule-based approaches overcome this limitation by applying a set of predefined grapheme-to-phoneme transcription rules. These rules are language-specific and have to be manually defined by linguists, then these can be formulated for example in the framework of finite-state automata [4]. Such rule-based methods also require an exception lexicon for irregular word forms.

Machine learning methods are also applied to grapheme-to-phoneme transcription. In [5], it has been shown that the generalization capability of such methods is better than that of rule-based approaches (at least for English). One of the most successful implementations is based on the idea of Pronunciation by Analogy

(PbA) [6]. The theory behind this approach is based on psycholinguistic models, i.e. predicting the pronunciation of a word by finding similarities to words for which the phonological representation is known. Joint-sequence models [3] aim at finding the most likely pronunciation for an orthographic form by using Bayes' decision rule. For all data-driven approaches, a dictionary or a transcribed corpus is needed for training the system or building statistical models.

For Hungarian, there is an online dictionary containing 1.5 million word forms and their phonetic transcription [7]. The construction of this dictionary included several main steps. First, word forms from a large, written corpus were collected and the list of the resulting words were cleaned (i.e. foreign and misspelled words removed). Then, transformation rules were applied. Finally, exceptions were defined and corrected manually. The authors state, that their dictionary can be considered as a reference dictionary, providing the largest coverage of Hungarian word forms and their IPA transcriptions. However, only word forms appearing in the original corpus are included, not providing the possibility either for transcribing other inflected forms (coming across such forms is a rather frequent event in the case of agglutinating languages like Hungarian) or including new or foreign words and names.

3 METHOD

In the case of languages with phonemic orthographies, such as Finnish, Estonian or Hungarian, the transcription of a written word form is usually rather straightforward. For example, the word *ablak* ('window') is pronounced as [ɔblɔk] in Hungarian. (Table 1 shows the transcription of the standard pronunciation of letters in the International Phonetic Alphabet, which is used in this research to represent phonemic transcription.) However, there are two types of phenomena that make the transcription nontrivial: assimilation phenomena (many of which are conditioned by morphological structure), and words having a traditional or foreign orthography. Another problem is the normalization of semiotic systems (such as numbers, abbreviations etc.).

Table 1: The phonemes of Hungarian

letter	IPA	letter	IPA	letter	IPA	letter	IPA
á	aː	b	b	n	n	zs	ʒ
a	ɔ	p	p	ny	ɲ	s	ʃ
o	o	d	d	j	j	cs	t͡ʃ
u	u	t	t	h	h	l	l
ü	y	g	ɡ	v	v	r	r
i	i	k	k	f	f	dz	d͡z
é	eː	gy	ɟ	z	z	dzs	d͡ʒ
ö	ø	ty	c	sz	s		
e	ɛ	m	m	c	t͡s		

Our method is based on three components: a morphological analyzer, a lexicon for irregular stems and the implementation of phonological rules defined using XFST (Xerox Finite-State Tool) [8].

3.1 *Morphological Analysis*

First, the morphological structure of each word is identified using the Humor morphological analyzer [9, 10]. This is necessary in order to find morpheme boundaries to which certain morpho-phonological rules refer. Lexical palatalization, for example, applies only to some specific inflectional suffixes. In addition, certain phonemes are represented by digraphs (*cs, gy, ty, ny, sz, zs, dz, dzs*, and their long forms). However, if a morpheme boundary intervenes, the individual consonants of these digraphs are pronounced as consonant clusters (other rules might affect their behaviour resulting in partial or full assimilation). For example, for the compound word *eszközsáv*, ‘toolbar’ the correct transcription is [ɛskøsʃaːv] instead of what we would get if it were a monomorphemic word: [ɛskøʒaːv].

Compounds, written in Hungarian as single (often long) word forms like in German, might also contain components that have an irregular pronunciation. These should also be recognized by the morphological analyzer to avoid their transcription by the regular phonological rules.

3.2 *Lexicon of Irregular Stems*

In all natural languages, there are word forms with irregular pronunciation. These are usually proper names and words of foreign origin. Words of the latter category might adapt to the adopting language to some extent. For example, the English word *file* might be written in Hungarian as in English: *file* or using the standard Hungarian transliteration of the adapted pronunciation of the word, i.e. *fájl*. In both cases, the phonetic form is [faːjl]. However, the phrase *New York* is used only in its original form in written texts and is pronounced as [ɲuːjork]. In Hungarian, however, not only foreign, but some traditionally spelled words also fall into this category. Such irregularities occur in quite a few family names, geographical names, etc. In addition, there are cases where standard pronunciation deviates from what orthography suggests in terms of vowel and/or consonant length. For example the word *egyesület* 'association' is pronounced as [ɛɟːɛʃylet] rather than [ɛɟɛʃylet], as suggested by the orthographic form.

Another group of words included in the lexicon are members of the semiotic system. These use the same set of characters and symbols as the writing system of the language, but render meaning to such units of text in a different manner. In order to be able to produce the phonological transcription, these units must be normalized in a preprocessing step. Examples are numbers, abbreviations, acronyms, units of measure, dates, mathematical expressions, e-mail addresses, etc.

Although handling each of these examples poses a number of problems of its own, it is out of the scope of this paper to go into details. You can refer to [11] instead. However, it is worth mentioning the case of abbreviations, where we shall differentiate forms

- where the abbreviated form can be pronounced as if it were a word (e.g. *NATO* [naːtoː]),
- or the abbreviated form is substituted by the original form in speech (e.g. *du.* [deːlutaːn] 'afternoon')
- or the abbreviation is spelled in speech (e.g. *USB* [uːeʃbeː]).

In our system, abbreviations are first matched against the lexicon that includes the transcription for those that are pronounced as words. If there are no matches, then the default rule is to spell the abbreviated form.

3.3 *Phonological Rules*

The morphophonological rules in our system were implemented using XFST. The description is based on [12]. The order of rules is shown in Table 2. The order of rules is determined by the following factors. 1) Orthographic peculiarities of consonant notation must be handled before other rules. 2) Lexical rules are applied before those describing postlexical processes. 3) There are a few feeding constraints between specific processes detailed below where we provide some details about each process.

Handling orthographic peculiarities

1. Certain palatal and sibilant consonants and affricates are denoted by digraph letters in Hungarian orthography, as shown in Table 1. Geminate consonants are in general denoted by doubling the corresponding letter. However, geminates of sounds denoted by digraphs are denoted by doubling only the first letter of the digraph. This rule handles these cases. Letter sequences that look like the geminate form of digraph-denoted consonants may also be cases of clusters, e.g. *ssz* may be a sequence of *s*+*sz* [ʃs], but this may occur only if there is an intervening morpheme boundary. In addition, the (partly context-sensitive) pronunciation of the letters *q, w, x* and *y,* used only in loan words and names, is defined in this group of rules.

Table 2: Phonological rules in the order of their application

#	**rule**
1.	convert long digraphs, *x, w, qu, y, ly*
2.	lexical *h*-deletion
3.	lexical palatalization
4.	lexical palatal merging (lex. palatalization must feed it))
5.	shortening of high final vowels of polysyllabic stems (optional)
6.	lengthening of intervocalic and word-final *dzs* and *dz*
7.	first syllable of every word stressed
8.	voicing assimilation (regressive, right context checked on the output)
9.	adaffrication (voicing assim. must feed it)
10.	nasal assimilation
11.	degemination
12.	j: at the and of phon. phrase: friction and devoicing after voiceless obstruents; friction after voiced consonants at the end of phon. phrase
13.	postlexical alternation of *h* (post sonorant voicing; palatalization and velarization in coda)
14.	postlexical palatalization
15.	stops, fricatives, nasals, liquids: gemination over all boundaries
16.	affricates: gemination over suffix boundaries
17.	convert vowels

Lexical processes

2. The final *h* of a subset of *h*-final words (e.g. *méh*, ‘bee, uterus’) is not pronounced unless a vowel-initial suffix follows.
3. The initial *j* of inflectional suffixes palatalizes preceding stem-final dental stops and *l*. The rule applies only at inflectional suffix boundaries.
4. The initial *j* of inflectional suffixes merges with a preceding stem-final palatal consonant. Lexical palatalization feeds this process.
5. Polysyllabic stems the orthographic form of which ends in a long high final vowel (*í, ú, ű*) are in general pronounced with a short final vowel [i, u, y] except in highly polished speech. We implemented this optional shortening.

6. Intervocalic and word-final *dzs* and *dz* are pronounced long. There are a handful of lexical exceptions with a short intervocalic *dzs* (e.g. *fridzsider* ‘fridge’ [fridʒider]).

Stress

7. Stress assignment is rather trivial in Hungarian: it always falls on the first syllable. The only complication is unstressed words like determiners and other clitics, but these are handled outside the rule set.

Postlexical rules

8. There is a regressive (right-to-left) voicing assimilation affecting obstruents. The peculiarities are: *v* is devoiced, but it does not trigger voicing; *h* triggers devoicing, but it is not voiced. This process must feed adaffrication.
9. Adaffrication: certain stop + fricative and stop + affricate clusters merge into corresponding affricates. We did not implement optional adaffrication processes characterizing only very casual speech, like Stop + fricative adaffrication across word boundaries or palatal + stop or palatal + affricate adaffrication.
10. Nasal *n* assimilates to the place of articulation of a right-adjacent stop or nasal, *n* and *m* are realized as a labiodental nasal [ɱ] preceding a labiodental obstruent.
11. There are a number of degemination processes, which are conditioned on different contexts. Monomorphemic geminates degeminate in the context of any other consonant: CC-X → C-X, X-CC → X-C (where - can be any boundary or none at all). Degemination across boundaries XC-C → X-C, C-CX → C-X is obligatory if X is an obstruent (and we implemented the process in nasal contexts as well). C-CX → C-X degemination affects a restricted subset of obstruents only. Degemination following a liquid L, LC=C → L=C, occurs only across inflectional suffix boundaries.
12. At the end of a word, *j* is realized as a voiceless [ç] or voiced fricative [ʝ] if it follows a voiceless/voiced consonant.

13. There is a postlexical alternation of *h*. It is voiced in intervocalic position and between a sonorant and a vowel. It is palatalized to [ç] in coda after front vowels, and, in other codas, it is velarized to [x].
14. Postlexical palatalization: dental *t, d, n* are palatalized before a palatal *ty, gy, ny*.
15. Stops, fricatives, nasals and liquids geminate over any type of boundaries.
16. In not-very-casual speech, affricates geminate only over suffix boundaries.
17. Finally, we convert the orthographic representation of long vowels to the standard IPA Vː notation.

4 EVALUATION

Our system was evaluated on the 80206-word Hungarian version of George Orwell's 1984. We used the Hungarian model of the eSpeak speech synthesizer [13] as a baseline system, the only freely available tool capable of performing grapheme-to-phoneme conversion for Hungarian we found. eSpeak can output an IPA transcription of its input. We also considered using the on-line pronunciation database [7] available at `http://beszedmuhely.tmit.bme.hu/mksz/` as another baseline. This dictionary contains 1.5 million word forms, including inflected forms and is supposed to be both representative and 99% correct. However, the database is not available for download, and even the function mentioned in the user guide of the site that would allow the user to download the first 1000 hits returned for a query is not implemented. So we did not manage to use it either as a reference or as a baseline vocabulary-based system.

We measured word error rate on the whole corpus. In the case of optional alternations, we accepted all correct variants. The eSpeak output lacks indication of any postlexical assimilation processes (of obstruent voicing, palatalization, nasals, /h/ and /j/), fails to clearly distinguish the IPA representation of affricates from obstruent clusters (e.g. /t͡ʃ/ vs. /tʃ/) and often erroneously represents geminate consonants as e.g. /tt/ instead of /tː/. We postcorrected these errors in the eSpeak output in order to make it comparable to our output (and correct). Another discrepancy between the two systems was that we implemented the optional shortening of stem-final long high vowels, which is typical even in non-casual standard Hungarian speech, while eSpeak outputs these in the their somewhat stilted long form. The word error rates of the two systems are shown in Table 3.

Table 3: Evaluation. u/i: ratio of words with shortening of stem-final long high vowels; assim/h/j/N/voic: ratio of words erroneously lacking marking of voicing/palatal/nasal/j/h assimilation but otherwise correct; WER: residual word error rate.

system	WER
our system WER	0.35%
eSpeak u/i	0.98%
eSpeak WER	2.26%
eSpeak assim/h/j/N/voic	14.81%

The errors in eSpeak's output not mentioned before are mainly due to lexical gaps (including the numerous English names in the text), its inability to resolve some common abbreviations, errors concerning geminate /r/'s and the pronunciation of the digraph *ch*, some idiosyncratic errors concerning the representation of certain words and the overapplication of lexical palatalization to morphemes that should not be affected. The latter type error is caused by the lack of morphological analysis in eSpeak: lexical palatalization is handled in a pattern-based manner, that also matches at wrong places. Our system is much better at pronouncing English names; its errors are mainly due to lexical gaps (different from those in eSpeak), wrong resolution of abbreviations and overanalysis of certain bogus compounds. The numerous Newspeak words in 1984 made up by Orwell, which a 'Hungarian' translation in the text, did not cause much trouble for either system, as they generally contain easy-to-convert letter sequences, and both systems have a productive transcription component instead of relying solely on a dictionary.

5 Conclusion

In this paper, an automatic tool was described that is able to transcribe Hungarian text to its phonetic representation. The system is more than a look-up tool for individual words, but is able to transcribe whole sentences, taking into account sound assimilations appearing at word boundaries as well. This is achieved by the incorporation of a morphological analyzer capable of detecting morpheme and compound boundaries, and by a set of transcription rules. Moreover, as the system is not limited to the vocabulary of a prebuilt dictionary, it is capable of transcribing any word forms, which is of crucial importance in languages like Hungarian, where agglutination and compounding can produce an unlimited number of words. It has been shown that evaluating our system on a dataset containing a high number of word forms not available in dictionaries, our system resulted in much lower error rate than a commercial tool, even if the latter was evaluated with a less strict attitude.

References

1. Németh, G., Olaszy, G.: A magyar beszéd (Hungarian Speech). Akadémiai Kiadó, Budapest, Hungary (2010)
2. Kurimo, M., Puurula, A., Arisoy, E., Siivola, V., Hirsimäki, T., Pylkkönen, J., Alumäe, T., Saraclar, M.: Unlimited vocabulary speech recognition for agglutinative languages. In: Proceedings of the Main Conference on Human Language Technology Conference of the North American Chapter of the Association of Computational Linguistics. HLT-NAACL '06, Stroudsburg, PA, USA, Association for Computational Linguistics (2006) 487–494
3. Bisani, M., Ney, H.: Joint-sequence models for grapheme-to-phoneme conversion. Speech Commun. **50**(5) (May 2008) 434–451
4. Kaplan, R.M., Kay, M.: Regular models of phonological rule systems. Comput. Linguist. **20**(3) (September 1994) 331–378

5. Damper, R., Marchand, Y., Adamson, M., Gustafson, K.: Evaluating the pronunciation component of text-to-speech systems for English: A performance comparison of different approaches. Computer Speech and Language **13**(2) (1999) 155 – 176
6. Dedina, M.J., Nusbaum, H.C.: PRONOUNCE: A program for pronunciation by analogy. Computer Speech and Language **5**(1) (1991) 55 – 64
7. Abari, K., Olaszy, G., Zainkó, C., Kiss, G.: Magyar kiejtési szótár az interneten [Hungarian Online Pronunciation Dictionary]. In: IV. Magyar Számítógépes Nyelvészeti Konferencia, Szeged, SZTE (2006) 223–230
8. Beesley, K., Karttunen, L.: Finite State Morphology. Number 1 in CSLI studies in computational linguistics: Center for the Study of Language and Information. CSLI Publications (2003)
9. Novák, A.: What is good Humor like? [Milyen a jó Humor?]. In: I. Magyar Számítógépes Nyelvészeti Konferencia, Szeged, SZTE (2003) 138–144
10. Prószéky, G., Kis, B.: A unification-based approach to morpho-syntactic parsing of agglutinative and other (highly) inflectional languages. In: Proceedings of the 37th annual meeting of the Association for Computational Linguistics on Computational Linguistics. ACL '99, Stroudsburg, PA, USA, Association for Computational Linguistics (1999) 261–268
11. Taylor, P.A.: Text-to-speech synthesis. Cambridge University Press, Cambridge, UK, New York (2009)
12. Siptár, P.: A magánhangzók [Consonants]. In Kiefer, F., Bánréti, Z., Ács, P., eds.: Fonológia. Number 2 in Strukturális magyar nyelvtan. Akadémiai Kiadó (1994)
13. eSpeak: eSpeak. `http://espeak.sourceforge.net/` Accessed: 2015-04-10.

ATTILA NOVÁK
MTA-PPKE HUNGARIAN LANGUAGE TECHNOLOGY
RESEARCH GROUP,
UNIVERSITY,
ADDRESS, COUNTRY
E-MAIL: <NOVAK.ATTILA@ITK.PPKE.HU>

BORBÁLA SIKLÓSI
FACULTY OF INFORMATION TECHNOLOGY AND BIONICS,
PÁZMÁNY PÉTER CATHOLIC UNIVERSITY,
50/A PRÁTER STREET, 1083 BUDAPEST, HUNGARY
E-MAIL: <SIKLOSI.BORBALA@ITK.PPKE.HU>