Rule-based grapheme-to-phoneme method for the Greek

A. Chalamandaris, S. Raptis,, P. Tsiakoulis.

Department of Speech Technology
Institute for Language and Speech Processing, Athens, Greece
{achalam, spy, ptsiak}@ilsp.gr

Abstract

This paper describes a trainable method for generating letter to sound rules for the Greek language, for producing the pronunciation of out-of-vocabulary words. Several approaches have been adopted over the years for grapheme-to-phoneme conversion, such as hand-seeded rules, finite state transducers, neural networks, HMMs etc, nevertheless it has been proved that the most reliable method is a rule-based one. Our approach is based on a semi-automatically pre-transcribed lexicon, from which we derived rules for automatic transcription. The efficiency and robustness of our method are proved by experiments on out-of-vocabulary words which resulted in over than 98% accuracy on a word-base criterion.

1. Introduction

The pronunciation of a word is an important task for both speech recognition and speech synthesis. The use of pre-transcribed lexicons can be considered to be the most robust and reliable method, but its use is limited since it is generally unable to cope with other words such as proper names, morphological variants etc.

Several methods and approaches have been proposed for performing automatic conversion from graphemes into phonemes and vice-versa. A direct hand-crafted lexicon access had been proposed initially [1]. The use of neural networks [2] has also been proposed, but with limited efficiency. Probabilistic approaches is one of the most commonly used methods for several languages ([3], [4]), the efficiency of which however is controversial and highly dependant on the target language. Two or more-level rule based approaches have also been proposed ([5], [6], [7], [8]), nevertheless, their implementation is not a trivial task since it relies on ad hoc empirical rules. Finally, other statistical methods have been tested for the same task with satisfactory performance such as HMM ([9], [10], [11]), finite transducers and automata ([12], [13], [14], [15], [16], [17]), or decision trees ([18], [19]).

2. Grapheme-to-Phoneme Conversion for the Greek Language

For the Greek language there have been several attempts for automatic grapheme-to-phoneme conversion, using different approaches. In [8] a set of empirical rules has been proposed while in [16] and [18] different methods like automata and decision trees have been proposed.

Grapheme-to-phoneme conversion for the Greek language is considered to be straightforward since for example the stress always given without ambiguity. Nevertheless this remains an open issue since no system has achieved 100% accuracy. For the Greek language, there are about 200 rules that generally apply for automatic transcription ([20], [8], [16]), which however cannot deal with exceptions of coarticulation phenomena, which are context- or word-specific. Namely, the cases of *C/i/V*, i.e. the cases where a consonant is followed by the /i/ phoneme and a vowel, the cases of the phonemes /b/, /d/, and /g/, and cross-word coarticulation phenomena are the most problematic ones.

In the case of *C/i/V*, the phoneme /i/ can arise from 6 different letters or set of letters (ι, η, υ, οι, ει, υι). This is a very frequent case in Greek since it has been estimated that the words containing a *C/i/V* phenomenon are the 8.5% of the total number of words that appear in a regular written text in modern Greek, as this was measured in HNC [21].

HNC is ILSP's corpus that has been acquired over the years and currently contains more than 34,000,000 words of written text. It is a web-based database where one can run queries on parts of speech or lemmas, selecting different types of sub-corpora.

The corresponding percentage for the words containing one of the /b/, /d/, or /g/ phonemes is 4.3%. Thus, it is obvious that if a letter-to-sound system for the Greek that has not foreseen to specially manipulate these cases is bound to perform rather poorly (less than 90% success rate on a word-base criterion).

3. The Proposed Method

The aim of the proposed method is to automatically produce grapheme-to-phoneme conversion rules based on a lexicon containing phonetic transcriptions prepared manually. The total number of words in the lexicon is 890,000, including the different morphological variants for each lemma.

The method consists of the following steps:

- The first step is to define all the possible matching rules for Greek, i.e. associating graphemic patterns with the respective phonemic patterns. A graphemic pattern may include one or more letters and a phonemic pattern one or more phonemes.
- Next, an alignment stage follows, where each word is aligned to its transcription based on the defined matching rules.
- Finally, the method tries to define rules for each graphemic pattern that are consistent throughout the corpus.

3.1. Pre-transcribed Lexicon Preparation

The preparation of the pre-transcribed lexicon, that has been used for deriving the conversion rules, is one of the most painstaking and time-consuming tasks. Its completion was carried out in two stages:

- Initially, a rough transcription was produced automatically for all the words of the lexicon using a prototype grapheme-to-phoneme engine that was based on empirical rules for the Greek language ([20], [8], [16]).

- At the second stage the lexicon was manually corrected. special care was given to the cases where *C/i/V* appear, and was carried out by experts phoneticians, who had a priori agreed upon one pronunciation of words of which more than one pronunciations were valid.

3.2. Alignment of the Lexicon

The alignment problem is maybe the most important task in the framework of letter-to-sound conversion. Although at a first sight it appears to be as simple as a sole string-matching procedure, such an approach would not be meaningful since it consists of a mapping between sequences over dissimilar alphabets.

There are several different approaches for performing the mapping between the grapheme and the phonetic transcription of every word, such as Dynamic Time Warping techniques, force alignment ([19]), regression and decision trees, HMMs ([11]), *epsilon* scattering ([19], [22]) i.e. mapping of one grapheme to either a phoneme or *epsilon*, or other machine learning techniques. Although these methods perform well enough and are good to use for unsupervised training, their accuracy is well below perfect and highly dependent on the complexity and the variety of the target language.

In order to achieve optimal results, a supervision approach is necessary. By having the latter in mind, we designed our system in such way that no error could be imposed during the alignment process. This was achieved by predefining all possible mappings from one grapheme to the corresponding phoneme, or from many graphemes to one or many phonemes. By doing so we disambiguated the cases where a grapheme corresponds to *epsilon* [22] (i.e. does not correspond to any phoneme but disappears) and managed to derive one single best mapping path for every pair of words, without the possibility of alternative, equally optimal or conflicting paths.

The total number of the mapping rules defined is about 12,000 and is a superset of all possible patterns for grapheme-to-phoneme conversion for the modern Greek. It should be noted that not all of these mappings are possible, since several of them cannot be found in real-life Greek words, while on the other hand, many of these rules are conflicting since they are not 1-1 mappings, but one class can have more than one mappings. This is an alternative way for performing the mapping between the words, since a large percentage of the reported bibliography makes use of the *epsilon* mapping, where a single letter can be transcribed into either a phoneme or into nothing. In Greek the mapping of a letter into alternative phonemes can be attributed to the context, therefore it is better if the alternative transcriptions are assigned to longer contexts. Our feeling was that although the *epsilon* cluster facilitates the automatic unsupervised alignment between word-pairs, it can lead to contradicting alignments. Therefore we decided to avoid its use and elaborate with longer clusters instead, where the disappearance or the alteration of a grapheme happens only within the limits of a cluster.

For example the Greek letter *iota (ι)* is transcribed into the phoneme *i* ,while the same letter in the context of *'δια'* the same letter can be transcribed into either */i/* or */j/*. The same letter can disappear, but only, for example, in the context of *για* for example and can never disappear without the context.

As soon as all possible transcription patterns are derived, the alignment process is simple to carry out. The algorithm is depicted in pseudo-code:

Algorithm:

1. *i=1 /* initialize the cursor */*
2. *Starting from the letter at position i find the longest transcription pattern that can match with the next n letters (n=sizeof(best_unit))*
3. *i=i+n /* proceed with the parsing */*
4. *if (i!=sizeof(string)) goto Step 2*

For example the word 'διαβαίνω' is aligned to its transcription as follows:

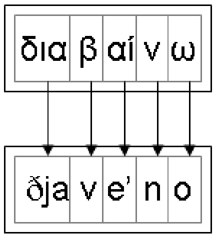

Figure 1: Grapheme to phoneme mapping using longer than single-sized classes.

while for the same word, using the approach of the epsilon phoneme results in the following alignment:

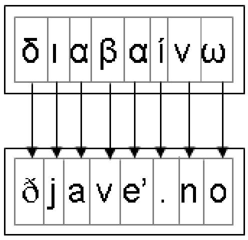

Figure 2: Grapheme to phoneme mapping using the "epsilon" approach.

By searching for the longest unit every time the cursor moves on within the word boundaries, we achieve to follow the best path according to the transcription patterns that we have predefined. The *epsilon* approach, although is convenient to use when attempting a dynamic and self-trained mapping between graphemes and phonemes, often leads to ambiguous and erroneous mappings, which will also lead to the derivation of erroneous rules.

3.3. Letter-to-Phoneme Rules

Given the mapped transcribed lexicon, we make an exhaustive search for every conversion rule that has been defined in the mapping process. At a first stage, by parsing the entire lexicon for every rule we identify the rules that apply without contradiction to the entire lexicon. These rules are set as the default rules that apply generally, such as for

Level	Num															
0	890								(0,0) 890							
1	1,145							(1,0) 561		(0,1) 584						
2	1,558						(2,0) 1226		(1,1) 220		(0,2) 112					
3	1,259					(3,0) 1089		(2,1) 55		(1,2) 50		(0,3) 65				
4	541				(4,0) 442		(3,1) 19		(2,2) 19		(1,3) 19		(0,4) 42			
5	244			(5,0) 176		(4,1) 8		(3,2) 11		(2,3) 4		(1,4) 17		(0,5) 28		
6	103		(6,0) 49		(5,1) 2		(4,2) 3		(3,3) 1		(2,4) 11		(1,5) 5		(0,6) 32	
7	33	(7,0) 15		(6,1) 1		(5,2) 4		(4,3) 0		(3,4) 0		(2,5) 8		(1,6) 0		(0,7) 5
	5,773															

Figure 3. The distribution of rule contexts. Each line corresponds to a specific context size. Parentheses include the left and right context sizes respectively. Below them is the number of such rules.

example for the letter alpha α which is always transcribed into the phoneme ‘a’. During a second stage, we identify the rules that do not apply always but are context dependent. These are the rules that have the same seed but different mapping (e.g. the unit ‘δια’ can be transcribed into either ðja or ðia). In order to handle these rules, we try to identify the shortest possible context for every rule to apply, with no contradiction within the entire lexicon. The lengthening of the context is performed on either sides (right and left context), one each time, until the compatibility of this rule is met throughout the lexicon.

This is a simple process that is responsible for identifying the rules that apply within the given lexicon and their, dependencies among the ones predefined during the mapping process. By this one would expect that the deriving rules would not exceed the number of the predefined mapping rules, however this is not absolutely true, since not all rules that have been manually predefined apply within the lexicon, many of them do not apply at all in Greek words, while on the other hand, the same rule can be the exception to the default rule with several different contexts.

The above process has resulted initially in 5,773 different rules with context up to 7 units. Due to the size of the hand-crafted lexicon, there are several errors that have been imposed during the manual transcription of the lexicon. However, after the production of the first set of rules, it became clear that some rules were attributed to erroneous entries in the lexicon. These entries were easy to track by the irregular context length and the low frequency of the rule that they derived. The same process was repeated on the corrected data in order to produce more accurate rules for letter-to-sound conversion.

4. Evaluation

In order to evaluate the efficiency of the system we run three different tests, one on the pre-transcribed lexicon, after a delineation 90%-10% split, we used 90% of the words for training of the system and the 10% of the words for evaluating it. The 10% of the lexicon was picked for one experiment every 9 words, in order to ensure random distribution of the words within the lexicons and to include as many as possible morphological variants, while for a second experiment the 10% was picked randomly following a Gaussian distribution in the indexes within the lexicon.

Finally, for further evaluation we manually checked the performance of the system on the 1,500 most frequent words of HNC [21] that were not included in the training lexicon, and projected the measured success rate on the overall size of the corpus by taking into account the frequencies of these words.

The results, summarized in the following table, were exceedingly well and accurate.

	Num. of Words	**Success Rate (%)**
Uniform Split	88,928	99,15%
Random Split	88,928	99,20%
Out of Vocabulary Words	1,500	99.43%

Table 1: The evaluation results of the proposed system

At this point we must stress that an important factor for the high success rate of our system is the fact that the lexicon we have used is long enough to cover most of the coarticulation phenomena met in the Greek language, therefore the training of the system is complete.

A typical test for measuring the efficiency of such systems is the evaluation with proper names or that have not been included during the training stage of the system [15]. The evaluation of the system with 12,856 Greek proper names that were not included in the training lexicon gave similar results, with an overall success rate of more than 98.5%.

Another evaluation test that would also be meaningful for such systems is the assessment with nonsense words, especially designed so that they cover most of the issues mentioned before as problematic for the Greek pronunciation. Such an evaluation is planned for the furute.

5. Conclusions

In this paper we presented a method for building an efficient and robust system for letter-to-sound transcription for the Greek language. Using manually transcribed lexicons and with hand-seeded mapping rules, we have achieved to build a rule-based grapheme-to-phoneme system that performs exceedingly well with out-of-the-vocabulary words, with a success rate of over 98% on word-base criterion. In its current state, the system cannot handle cross-word coarticulation phenomena, which often occur. It is in our short-term plans to investigate and incorporate rules that will be able to handle such cases, mainly as a post-processing module.

6. Acknowledgements

The work reported on here was carried out within ILSP (Institute for Language and Speech Processing), in Athens. We would like to thank S. Karabetsos, E. Spais C. Kalimeris and C. Alexandris for their valuable help during this work, Dr. A. Protopapas for his support and all the people who have supported and actively contributed to this work.

7. References

[1] Schmidt M., et al, "Phonetic Transcription Standards for European Names (Onomastica)", *European Conference on Speech Communication and Technology (Eurospeech), vol. 1, pp. 279—282, 1993.*

[2] Michael D. Riley, "A statistical model for generating pronunciation networks", *International Conference on Acoustics Speech, and Signal Processing, p. 737-740, 1995.*

[3] Andersen O. and Dalsgaard P., "Multi-lingual testing of a self-learning approach to phonemic transcription of orthography", *Eurospeech95, p. 1117-1120, 1995*

[4] Andersen O., Kuhn R., et al., "Comparison of two tree-structured approaches for grapheme-to-phoneme conversion", *Proc. of ICSLP, pp. 1808-11, 1996.*

[5] Antal van den Bosch and Walter Daelemans. "Data-oriented methods fro grapheme-to-phoneme conversion", *6th European Conference of the Association for Computational Linguistics, proc. p45-53.*

[6] Ravishankar M., Eskenazi M., "Automatic Generation of Context-Dependent Pronunciation", *ESCA Eurospeech97, 1997.*

[7] Niklas Torstensson, "Grapheme-to-phoneme conversion, a knowledge-based approach", *Speech Music and Hearing TMH-QPSR-Fonetik, vol. 44, p.117-120, 2002.*

[8] Bakamidis S. and Carayannis G., "Phonemia: a phoneme transcription system for speech synthesis in modern Greek", *Speech Communication 6 p. 159-169, 1987.*

[9] Yeon-Jun Kim, Ann Syrdal and Alistair Conkie, "Pronunciation Lexicon Adaptation for TTS Voice Building", *ICLSP, 2004.*

[10] Ananlada Chotimongkol and Alan W. Black., "Statistically trained orthographic to sound models for Thai", *ICSLP, 2000.*

[11] Rentzepopoulos P. and Kokkinakis G., "Efficient multi-lingual Phoneme-to-Grapheme conversion on HMM", *Computational Linguistics, vol. 22, p.319-376, 1996.*

[12] Walter M., P. Daelemans and Antal P. J. van den Bosch. "Language-Independent data-oriented grapheme-to-phoneme conversion", *Progress in Speech Synthesis, p. 77-89, Springer, New York.*

[13] Dan Gildea and Dan Jurasfsky, "Automatic induction of finite state transducers for simple phonological rules", *33rd Annual Meeting of the Association for Computational Linguistics, proc. p 9-15, 1995.*

[14] George Anton Kiraz, "Compressed storage of sparse finite-state transducers", *Workshop on Implementing Automata, 1999.*

[15] Martin Jansche, "Re-Engineering Letter-to-Sound Rules", *The Second Meeting of the North American Chapter of the Association for Computational Linguistics, 2001.*

[16] Sgarbas K., Fakotakis N. and Kokkinakis G., "A PC-KIMMO-Based Bi-directional Graphemic/Phonetic Converter for Modern Greek", *Literary and Linguistic Computing, Oxford University Press, vol.13 No.2 pp. 65-75, 1998.*

[17] Caseiro D., et al. "Grapheme-to-phone using finite-state transducers", *IEEE Workshop on Speech Synthesis, 2002.*

[18] Dermatas E. and Kokkinakis G., "A Language-independent probabilistic model for automatic conversion between graphemic and phonemic transcription of words", *Eurospeech99, p. 2071-2074, 1999.*

[19] Pagel V., Lenzo K. and Black A. W., "Letter to sound rules for accented lexicon compression", *ICSLP, vol. I, p 252-255, 1998.*

[20] Petrounias E., "Grammatical and comparative analysis of modern Greek", *University Studio Press, Thessaloniki, 1993.*

[21] Hellenic National Corpus™ (HNC) ILSP, http://hnc.ilsp.gr/ - Web Version 2.0, 2005

[22] Alan W. Black, Kevin Lenzo and Vincent Pagel. "Issues in Building general letter to sound rules.", *3rd ESCA Workshop on Speech Synthesis, proc. 77-80, 1998.*