J Braz Comput Soc (2013) 19:127–134
DOI 10.1007/s13173-012-0088-0

ORIGINAL PAPER

Generating a pronunciation dictionary for European Portuguese using a joint-sequence model with embedded stress assignment

Arlindo Veiga · Sara Candeias · Fernando Perdigão

Received: 30 December 2011 / Accepted: 23 August 2012 / Published online: 12 September 2012

Abstract This paper addresses the problem of grapheme to phoneme conversion to create a pronunciation dictionary from a vocabulary of the most frequent words in European Portuguese. A system based on a mixed approach funded on a stochastic model with embedded rules for stressed vowel assignment is described. The implemented model can generate pronunciations from unrestricted words; however, a dictionary with the 40k most frequent words was constructed and corrected interactively. The dictionary includes homographs with multiplepronunciations. The vocabulary was defined using the CETEMPúblico corpus. The model and dictionary are publicly available.

Keywords G2P conversion · Grapheme–phoneme converter · Pronunciation dictionary · Joint-sequence model · Stress assignment rules

1 Introduction

The grapheme to phone(me)[1] conversion (G2P), also called letter-to-sound conversion, maps a written text into a string of symbols which represent the speech sounds exactly and unequivocally. Several frameworks have been proposed to tackle the G2P conversion, among which linguistically rule-based modules [18] and statistical approaches [10] can be mentioned. Mainly in the languages in which orthography is roughly phonologically based, such as the Portuguese and other Romanic Languages, linguistic rule-based systems should provide a good coverage of the association between letters and sounds [6,25,29]. However, probably no natural human-language satisfies this assumption exactly, because exceptions from the G2P conversion can be found perhaps in every language. The most common irregularity covers situations when the association between grapheme and phoneme is not quite one-to-one but can be, to some extent, ambiguous and greatly dependent on the neighbor-contexts.To deal with this problem, rule-based systems have been adopted along with a list of exceptions to cover the unruled situations. But this solution turns the development and the maintenance of the system very complex, hard and tiresome. Moreover, the rule-based G2P is more likely to make mistakes for new words. In contrast to the rule-based systems outlined above, a number of authors have addressed the G2P conversion from a stochastic perspective. This approach to G2P conversion is based on the idea that using pronunciation examples it could be possible to predict the pronunciation of unseen words by analogy. This method was already implemented by [8] and

This is a revised and extended version of a previous paper that appeared at STIL 2011, the 8th Brazilian Symposium in Information and Human Language Technology http://www.ufmt.br/stil2011/.

A. Veiga · S. Candeias (✉) · F. Perdigão
Instituto de Telecomunicações-polo de Coimbra,
Coimbra, Portugal
e-mail: saracandeias@co.it.pt

A. Veiga · F. Perdigão
Department of Electrical and Computer Engineering,
FCTUC, Universidade de Coimbra, Coimbra, Portugal
e-mail: aveiga@co.it.pt

F. Perdigão
e-mail: fp@co.it.pt

[1] Phone(me) signifies either phone or phoneme. Since the studies on the G2P often alternate between the terms phone and phoneme (as we will see with more detail in the Sect. 4), here we propose a mixed term just to highlight the problem.

[2], among others, for Portuguese. In this paper, we use a new statistical approach for which outstanding results have been reported, named the joint-sequence model, [5]. In this model, graphemes and phonemes are combined into a single state, giving rise to "graphonemes".

Although the joint-sequence model has shown to be a powerful tool, we also shown in this paper that for the case of the Portuguese language the determination of the stressed vowel leads to a substantial improvement in the system performance, as was also reported in [8]. Thus, we included a linguistically rule-based pre-processing stage, for stress assignment,which marks and disambiguates most of the pronunciations.

Common errors in the conversion procedure occur with the heterophonic homographs. Some theoretical frameworks with experimental results were recently proposed, e.g. [32,33] and [34] for European Portuguese (EP); and [35–38] and [39] for Brazilian Portuguese (BP). The study by Braga and Marques [34] proposed algorithms to deal with this problem of the homograph ambiguity in EP, using a linguistic rule-based methodology. Working with a part-of-speech (PoS) parser to disambiguate homographs which belong to different PoS, and a semantic analyser to disambiguate homographs that belong to the same PoS, the authors extended the approach proposed in [35,36]. In fact, PoS categorization is insufficient to disambiguate entries in a pronunciation dictionary. Our solution consists in including PoS as well as pronunciation information for each dictionary entry.

The vocabulary used to generatethe pronunciation dictionary is in its previous form of the current "Acordo Ortográfico" (AO).[2] However, we think that this mixed-based G2P can also achieve good performance for EP with the AO. The inherent flexibility in dealing with the EP could be extended to other Romanic languages, which make this an advantageous approach.

The remainder of the paper is organized as follows. In Sect. 2, the joint-sequence model is briefly discussed. Section 3 presents how the vocabulary and dictionary were generated while Sect. 4 describes the linguistic model. In Sect. 5, experimental results are presented and the methodology used to deal with the heterophonic homographs is explained in Sect. 6. Then, the main conclusions are summarized and future work directions are foreseen.

2 Joint-sequence model

Given a sequence of N graphemes defined by $G = G_1^N = \{g_1, g_2, ..., g_N\}$, the goal is to find a sequence of M phonemes, $F = F_1^M = \{f_1, f_2, ..., f_M\}$, that best describes the phonetic transcriptionof the original sentence. The statistical approach to this problem corresponds to the determination of the optimal sequence of phonemes, F^*, that maximizes the conditional probability of phonemes, F, given a sequence of graphemes, G:

$$F^* = \arg\max_F P(F|G). \tag{1}$$

It is difficult to determine F^* directly by calculating $P(F|G)$ for all possible sequences F. However, using the Bayes theorem, we can rewrite the problem as:

$$F^* = \arg\max_F P(F|G) = \arg\ \max_F \{P(G|F).P(F)/P(G)\}. \tag{2}$$

Since $P(G)$ is common to all sequences F, the problem can be simplified in the following way:

$$F^* = \arg\max_F P(G|F).P(F). \tag{3}$$

Using a phonological dictionary, previously created, it is possible to estimate $P(G|F)$ and the a priori probability, $P(F)$, for all sequences F and G found in this dictionary. The Markov-based approaches estimate a model for each phoneme and use n-gram models to compute$P(F)$. These approaches model the dependency between graphemes and phonemes and the dependency between phonemes, but do not model dependencies between graphemes [12,17,28]. Due to these constraints, other statistical approaches emerged proposing joint probability models $P(F, G)$ to determine the optimal sequence of phonemes [4,14], directly using the expression of the joint probability in (1) in place of the conditional probability. In this approach, all the dependencies present in the dictionary were modeled, resulting in improved performances than those obtained by the other models.

2.1 Alignment between graphemes and phonemes

Some graphemes have a univocal correspondence with the phonemes. However, for other graphemes the correspondence to phonemes depends on several factors, such as the grapheme context and the part-of-speech. There are also cases where several graphemes may lead to a single phoneme, and where a single grapheme can lead to several phonemes. All statistical approaches face this problem, being necessary, during the training process, to segment and align the two sequences (a phoneme sequence and the corresponding grapheme sequence) with an equal number of segments. The solution is not always trivial or unique and depends

[2] The "AcordoOrtográfico" or "AO" (Portuguese Language Orthographic Agreement) is an international treaty signed in 1990, with the purpose of creating a unified orthography for the Portuguese language, to be used by all the countries that have Portuguese as their official language (see also http://www.portaldalinguaportuguesa.org/acordo.php

on how the alignment algorithms associate graphemes to phonemes of a given word. Alignment can be classified as follows [16]:

1) "*one-to-one*" Each grapheme relates with only one phoneme (segments with one symbol only). A null symbol ('_') is used to deal with the cases in which a grapheme can originate more than one phoneme (the insertion of phonemes), or the cases where more than one grapheme originates only one phoneme (the deletion of phonemes). This alignment is easy to implement using the Levenshtein algorithm [22]. In the literature, these algorithms are called alignment "01-01" if insertions and deletions of phonemes are allowed, or "1-01" if only deletion of phonemes is allowed. This last case corresponds to the alignment used in this work.
2) "*many-to-many*" The segments are composed of various symbols, which allow the association of several graphemes to several phonemes. This alignment is more generic and can be used without any prior knowledge of mapping between graphemes and phonemes. It handles insertions and deletions of phonemes without using any special symbol. On the other hand, the resulting model is more difficult to estimate and its performance is generally lower than the model with alignment "one-to-one". These alignments are also known as "*n*-to-*m*".

2.2 Statistical model

After the alignment, the sequences of graphemes and phonemes have the same number of segments. So, a new entity, born from the association of a segment of graphemes and phonemes can be defined, and is called "graphone(me)" [4]. A sequence of K graphonemes is annotated as $Q(F, G) = \{q_1, q_2, \ldots, q_k\}$. Given a sequence of K graphonemes, $Q(F, G)$, rather than assuming independence between symbols, the probability of the joint-sequence, $P(Q(F, G))$, can be estimated using the so-called "n-grams" [5] (sequences limited to n symbols).

2.3 Model estimation

The n-gram models are used to estimate the probability of symbols knowing the previous $n - 1$ symbols (history). The estimation of the probability of an n-gram is based on the number of its occurrence. This probability is easy to compute, but there is a problem in assigning a zero probability to the n-grams not seen or with limited number of training examples. To overcome this limitation, it is necessary to model unseen examples (using a discount) or uncommon examples (using smoothing). Thus, a small probability mass must be reserved from the most frequent n-grams to the absent or uncommon n-grams. There are several proposed algorithms to solve this problem of probability mass redistribution, such as Good-Turing [15], Witten-Bell [31], Kneser-Ney [20], Ney's absolute discount [23] and Katz's smoothing [19]. In this work, we have adopted the algorithm implemented by [13], which uses a modified version of Kneser-Ney algorithm [9].

3 Pronunciation dictionary

In this work, we intend to create a pronunciation dictionary from a given vocabulary. The vocabulary derives from the CETEMPúblico corpus [26], that corresponds to a collection of newspaper extracts published from 1991 to 1998, annotated in terms of sentences and containing 180 million words in European Portuguese. The process of generating the vocabulary starts by taking all the strings annotated as words, which obey simultaneously to the following criteria: (1) start with a letter (a–z, A–Z, á–ú, Á–Ú); (2) do not contain digits; (3) are not all upper case (e.g. acronyms); (4) do not have the character '.' (e.g. URLs); (5) end with a letter (e.g. not A4, UTF-8); (6) the corresponding lemmas do not contain '=' (e.g. compound nouns).

From the resulting list, we took the sub-list of words that occur more than 70 times in the corpus, totaling about 50k different words. Foreign words were then removed, using an automatic criteria followed by manual verification. This process results on a vocabulary of 41,586 words.

3.1 Transcription

The transcription of the vocabulary words is a result of an iterative procedure. First, a statistical model was estimated, as described in 2.2, using the SpeechDat pronunciation dictionary [27]. This dictionary contains about 15k entries, from which foreign words were deleted. Some SAMPA transcriptions [30] were substituted according to the following directions: (1) we did not use the velar /l∼/ and the semivowels /j/ and /w/; and (2) some standardization in the pronunciations was done, such as considering /6i/ as the pronunciation of all < ei > grapheme sequences (e.g. <l **ei**te> /l6it@/ and <alh **ei**a> /6L**6i**6/).

The result of applying the statistical model to CETEMPúblico vocabulary was fairly accurate, although with some significant flaws. Then, we followed a long procedure of manual verification and correction of the transcriptions. The next step was to compare the transcriptions with other ones, generated by a commercial speech synthesizer. This comparison allowed us to rely on our results since the majority of the transcriptions agreed. All different transcriptions were analyzed one by one and we found that the transcriptions from our dictionary were the right ones most of the times. This has led to the phonological transcription dictionary referred to as "dic_CETEMP_40k".

With the "dic_CETEMP_40k", a new statistical model was built. The test of this model on the training dictionary, allowed us to correct some remaining errors as well as to standardize and regularize some transcription procedures. Throughout the development of this work, the dictionary had been revised and corrected. Although it may still contain some errors, we are confident on its accuracy. We think that this dictionary could be an interesting resource for studies about phonetics and phonology of Portuguese.

3.2 Graphoneme alignment

An important step for establishing the statistical model is the alignment between graphemes and phonemes in the form "1-01" (one grapheme leads to zero or one phoneme; see Sect. 2.1).The option"1-01" was chosen from the beginning, because we had identified only six cases where a grapheme could originate more than one phoneme. Some cases had the insertion of a yod in some words beginning with <ex->; others had the cases of non-common pronunciations such as <põem>→/po~i~6~i~/ and <têm>→/t 6~i~6~i~/. Defining symbols corresponding to more than one phoneme solved this problem of phoneme insertion. The problem of the phoneme deletions still remains, because there are always graphemes that do not originate any phoneme.

The alignment between graphemes and phonemes was then obtained using the known edit distance or Levenshtein algorithm [22]. This required defining a distance between each phoneme and grapheme. This distance or cost of association was defined using the log probability of this association, which was estimated from an aligned dictionary.

4 Phonetic–phonological restrictions

Since the EP is a language with much phonological regularity, we added to the G2P module some linguistic restrictions, which were pertinent to convert graphemes into phonemes. Before any regard on the linguistic rules, an aspect concerning the phonetic/phonological binomial must be clarified. While phonetics gives us the physical and articulatory properties of the sound pronounced (it means the surface structure), phonology studies the sound that has a given role in the pronunciation (the underlying structure). However, any methodological perspective concerning the speech transcription links these two linguistic fields since it deals with the inter-relationship between the units and its distinctive character (phonemes) and the physical reality of those units (phones and allophones) [11].

The studies on the G2P often alternate between the term phone [8,24] and the term phoneme [2], without any clarification on the perspective followed. We justify our option to adopt the term phoneme mainly, because the procedure to convert the letter into the sound brings us information that derives from the structure of the language (such as both left and right context which implies the choice of a single unit excluding all other units available in the language). The phoneme that corresponds to the grapheme is well accepted as a class to which may group all allophonic realizations able in EP (which could include all the multi pronunciations). We also considered that the phoneme conversion corresponds to the EP-standard. The phonological neutralization of oppositions is not described in this study and phonemes do not represent any archiphonemes.

Algorithms have been constructed based on practical linguistic rules, such as stress marking of the vowel (the syllable nuclei) of any single word and by identifying short contexts in which the correspondence between grapheme and phoneme has a good stability.

4.1 Rules for stress assignment

Following the theoretical assumptions discussed in [21], we adopted to mark all vowels, which are stressed (the syllable nuclei) within a word. The importance of the stressed vowel (V_{stressed}) has been recognized in previous G2P works, such as in [8]. Since the n-grams context is short and cannot, most of the times, retain information about the syllable structure, marking the V_{stressed} improves the statistical model by expressing graphoneme classes unequivocally. As in [1], our proposal considered to mark the V_{stressed} (with the symbol ’ ” ’) and did not require the identification of the syllabic unit. However, the process of identifying the V_{stressed} that is described in this study was achieved in a very simple way. In the following Table 1, a set of rules for stressing vowels is presented with examples. All contexts were considered, including those without a stressed vowel, such as the prepositions <com>, <de>, <em>, <sem>, <sob>, <do(s)>, <no(s)>; the personal pronouns <me>, <te>, <se>, <nos>, <vos>, <lhe(s)>, <o(s)>,<a(s)>, <lo(s)>, <no(s)>, <vo(s)>, <mo(s)>, <to(s)>, <lho(s)>; the relative pronoun <que>; and the conjunctions <e>, <nem>, <que>, <se>, which are often added to a stressed nuclei within the prosodic unit.

A problem arises with words, which are morphologically derived, such as the adverbs ending in <mente>, especially when the adjectival form, from which they derive, has a stress mark (e.g.<rápido>→<rapidamente>;<dócil>→ <docilmente>). The solution adopted was the following: we implemented an algorithm that divides the word into two parts, <*ROOT*> and <mente>. The <*ROOT*> part undertakes a specific module, which compares it with a list of graphematic patterns which have the V_{stressed} identified. This

Table 1 Rules for stress assignment of the vowels (V)

	Rules	Example
1.	**If** the word has a V with a graphic stress mark, **Then** $V \rightarrow V_{\text{stressed}}$	aux”ílio, an”álise, avaliaç”ão, ”às, s”ót”ão
2.	**If** the word has not a graphic stress mark and ends in <a>, <e> or <o> followed (or not) by <m|n|s>, **Then** prior V to <a>, <e> or <o>→ V_{stressed}	c”arta, d”ançam, cont”ente(s), h”omem, h”omens, est”udo(s)
3.	**If** the word has not a graphic stress mark and ends in C <l>, <r>, <x> or<z>, **Then** the last $V \rightarrow V_{\text{stressed}}$	defens”or, cant”ar, emit”ir, dev”er, can”al, pap”el, fun”il, telef”ax, dupl”ex, cab”az, fel”iz, arr”oz
4.	**If** the word has not a graphic stress mark and ends in V <i> or<u>, followed (or not) by <m|n|s>, **Then** <i> or<u>→ V_{stressed}	delf”im, bot”ins, par”is, alg”um, com”uns, jes”us
5.	**If** in 2, 3 and 4, the V <i> or<u>arepreceded by other V, **Then** $V \rightarrow V_{\text{stressed}}$	p”ai(s), r”ei(s), m”au(s), l”eu, decid”iu, c”aixa(s), ad”eus, p”eixe(s), p”auta(s), l”ouça(s), natur”ais
6.	**If** in 5 V <i> or<u>are followed by <ch>, <nh>, <m + C|#> or<n + C >, **Then** <i> or<u>→ V_{stressed}	sandu”iche, vento”inha, amendo”im, co”imbra

method solved all the cases present in the dictionary of 40k words.

This pre-processing module attributes a special symbol to all stressed vowels generating a univocal graphoneme.

5 Model results

All experiments were based on the pronunciation dictionary of 41,586 Portuguese words as described in Sect. 3.1. There are two cases, corresponding to the dictionary with and without stress marking.

To train and test the statistical model, each one of these two dictionaries was partitioned into fivefold for a cross-validation procedure. The initial dictionary is divided into fivefold, each one with 8,317 (20 %) randomly chosen words. The words are mutually exclusive in each of the five folds. Each fold is used to perform a training and a testing run. Final results were obtained by evaluating the average of the five partial results.

The performance of the G2P conversion system was expressed in two average error rates (over the fivefold): average error rate of phonemes (PER) and average error rate of words (WER). The following figures summarize the results obtained using n-grams with n between 2 and 8.

As it can be seen in Fig. 1, the marking of the stressed vowel contributed to a significant improvement in the system performance. Note that, on the contrary to what we would expect, the use of n-grams with large contexts (n greater than 5) did not improve the system. The best results were achieved with 5-grams, attaining 0.32 % of PER and 2.44 % of WER using stress marking. In fact, we observe an increase in the error rates with contexts larger than 5-grams. This can be explained by the lack of samples to estimate properly the n-grams with large contexts. The optimal length of n-grams was 5 in this case, but it depends on the size of the training dictionary. For example, the optimal context for the SpeechDat pronunciation vocabulary was $n = 4$.

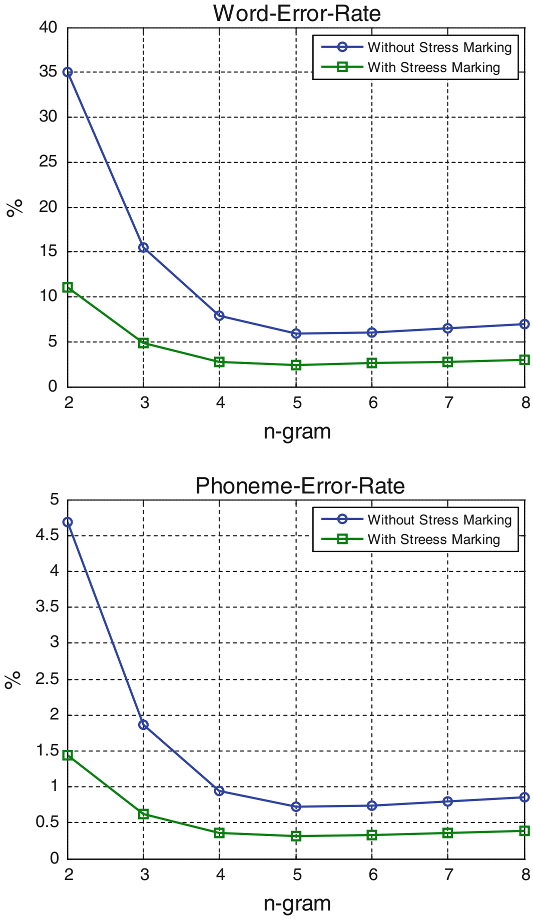

Fig. 1 Word and phoneme error rates (WER and PER) for the two models.

We cannot compare directly our results with other systems’ for Portuguese, since the data or the systems are not publicly available. However, the results presented here are the best reported in similar works. For instance, in [7] a PER of 99.11 % is achieved in 1,000 sentences of CETEMPúblico (8–12 words per sentence), but the total number of words is not reported. In [8] a WER of 3.94 % and a PER of 0.59 % were indicated with 7-grams and with stress assignment. In this work, it was already reported a significant performance improvement with the stress assignment. The database has more than 200k words automatically transcribed. In [2], a value of performance of about 89 % is reported.

Although we cannot compare directly these results with ours, we think that the joint-sequence model has achieved very good results. In fact, by inspecting the test errors, we observed that most of them resulted from uncommon grapheme patterns or compound words without graphic stress marks. The most frequent errors resulted from the ambiguity of the pronunciation of the stressed <e> and <o>, since they could be pronounced as /E/ vs. /e/ and /O/ vs. /o/ without any systematic rule.

Other errors are due to the multiple pronunciations of some homographic words. Although this kind of errors is not the most frequent in the results presented here, cases of heterophonic homographs are very important to consider. To solve this problem of multiple pronunciations, we had to change our G2P system to include additional information appended to each pronunciation in the dictionary.

6 Heterophonic homographs

When two words have the same spelling but different pronunciations, they are called heterophonic homographs. They can belong to different PoS, such as in <*dobro*>, pronounced as /dobru/ 'double' (noun) or as /dObru/ 'fold' (verb in the 1st person of the present tense, indicative mood), in <*poça*>, pronounced as /pOs6/ 'puddle' (noun) or as /pos6/ 'damn!' (interjection), and in <*esmero*>, pronounced as /@Smeru/ 'care' (noun) or as /@SmEru/ 'I perfect' (verb in the 1st person of the present tense, indicative mood). Heterophonic homographs can also have the same PoS, such as in <*aposto*>, pronounced as /6poStu/ 'appended' or as /6pOStu/ 'I bet' (both verbs); in <*travesso*>, pronounced as /tr6vesu/ 'naughty' or as /tr6vEsu/ 'transverse' (both adjectives); and in <*bola*>, pronounced as /bol6/ 'meat pie' or as /bOl6/ 'ball' (both nouns).

To deal with the problem of deciding what pronunciation the converter should present for a given heterophonic homograph, we integrated into the G2P system a list of 591 homographs which contains 1,182 different pronunciations.[3]

The homographs were taken from several databases, namely the CETEMPúblico, the Orthographical Vocabulary of Portuguese[4] and dictionaries for Portuguese available online[5]. Each homograph has associated both PoS category and pronunciation form. We have focused on heterophonic homographs which have the vowels <e> and the <o>, since they could be pronounced, respectively, either as /e/-/E/ or /o/-/O/ regardless of the phonological context. The most frequent cases of heterophonic homographs exemplify the multi-pronunciation of the vowel located in the stress position; however, some pairs with the vowel located in a non-stress position were found in the corpora, such in <*pregar*> /pr@gar/ 'to nail' (verb) vs. /prEgar/ 'to preach' (verb) or <*pegada*> /p@gad6/ 'glued' (verb, adjective) vs. /pEgad6/ 'footprint' (noun).

Although the PoScategory is enough to clarify the pronunciation of the most of homographs, there are some cases of different pronunciations for the same PoS, as was already observed in [34]. In our dictionary, the ambiguity of the pronunciation remains in 228 homographs with the same PoS. For this reason, we associated to a dictionary entry not only the PoS, but also the indication of the alternative vowel sound.

In terms of implementation of a practical G2P system, an off-line dictionary can be developed and incorporated in it. In fact, our final system includes the developed dictionary as an "exception list" using a hash table. Only the words not included in the dictionary are converted by the statistical model. This turns the G2P system with a very low latency, since the vocabulary has the most frequent words. It is the user responsibility to indicate the desired pronunciation with PoS or/and the alternative vowel sound; otherwise a default pronunciation is returned.

7 Conclusions and future work

The generation of a pronunciation dictionary for European Portuguese is described in this work. The technique used for the grapheme to phoneme conversion is based on a stochastic model, the joint-sequence model, which uses the concept of graphonemes and in which rules for stressed vowel assignment were embedded. The vocabulary includes the most frequent words that occur in Portuguese, as found in the CETEMPúblico corpus. A list of about 600 homographic words was also included, to disambiguate the cases of multiple pronunciations occurring on Portuguese. The results presented here are the best reported in similar works, although not directly comparable due to the use of different databases.

The G2P system is freely available on the website http://lsi.co.it.pt/spl/ in the "resource" section, which contains the models, dictionaries and the G2P module.

There is an ongoing study on the analysis of the phonological behavior of the foreign words. Morphological information in terms of masculine/feminine, singular/plural and the inflection of the verbs can also be included in future developments. We also intend to enlarge the dictionary to other varieties of Portuguese.

[3] See http://lsi.co.it.pt/spl/resources/dic_homografas_heterofonas.txt.

[4] http://www.portaldalinguaportuguesa.org.

[5] http://www.infopedia.pt/; http://www.priberam.pt/dlpo/.

Acknowledgments The two first authors acknowledge Instituto de Telecomunicações (Arlindo Veiga) and Science and Technology Foundation-FCT (Sara Candeias, SFRH/ BPD/36584/2007) for their scholarships. This work was also fundedby FCT under the Project (PTDC/CLE-LIN/11 2411/2009) and partially supported by FCT (Instituto de Telecomunicações multiannual funding PEst-OE/EEI/ LA0008/2011).

References

1. Andrade E, Viana MC (1985) Corso I—Um Conversor de Texto Ortográfico em Código Fonético para o Português. Technical Report, CLUL-INIC, Lisboa
2. Barros MJ, Weiss C (2006) Maximum entropy motivated grapheme-to-phoneme, stress and syllable boundary prediction for Portuguese text-to-speech. IV Jornadas en Tecnologías del Habla, Zaragoza
3. Eckhard B (2000) The parsing system “Palavras”: automatic grammatical analysis of Portuguese in a constraint grammar framework. Dr.phil. thesis, Aarhus University Press, Aarhus
4. Bisani M, Ney H (2002) Investigations on joint-multigram models for grapheme-to-phoneme conversion. In: Proceedings of the 7th international conference on spoken language processing (ICSLP’02), Denver, USA, pp 105–108
5. Bisani M, Ney H (2008) Joint-sequence models for grapheme-to-phoneme conversion. Speech Commun 50(5):434–451
6. Braga D, Coelho L (2006) A rule-based grapheme-to-phone converter for TTS systems in European Portuguese. VI International Telecommunications Symposium, Fortaleza
7. Braga D (2008) Algoritmos de Processamento da Linguagem Natural para Sistemas de Conversão Texto-Fala em Português. PhD thesis, Universidade da Coruña
8. Caseiro D, Trancoso I, Oliveira L, Viana C (2002) Grapheme-to-phone using finite-state transducers. In: Proceedings of the IEEE 2002 workshop on speech synthesis, California USA, pp 215–218
9. Chen S, Goodman J (1998) An empirical study of smoothing techniques for language modeling. Technical Report TR-10-98, Center for Research in Computing Technology (Harvard University)
10. Chotimongkol A, Black A (2000) Statistically trained orthographic to sound models for Thai. In: Proceedings of ICSLP, vol 2. Beijing, China, pp 551–554
11. Crystal D (2002) A dictionary of linguistics and phonetics, 5th edn. Blackwell, Oxford
12. Demberg, V. (2006), Letter-to-Phoneme Conversion for a German Text-to-Speech System, Stuttgart University, published as book by Verlag Dr. Müller (VDM), ISBN: 978-3-8364-6428-4 (from Amazon.com).
13. Demberg V, Schmid H, Möhler G (2007) Phonological constraints and morphological preprocessing for grapheme-to-phoneme conversion”. In: Proceedings of the 45th annual meeting of the association for computational linguistics (ACL-07), Prague, Czech Republic, pp 96–103
14. Galescu L, Allen J (2001) Bi-directional conversion between graphemes and phonemes using a joint N-gram model”. In: Proceedings of the 4th ISCA workshop on apeech aynthesis, Perthshire, Scotland
15. Good I (1953) The population frequencies of species and the estimation of population parameters. Biometrika 40(3,4):237–264
16. Jiampojamarn S, Kondrak G, Sherif T (2007) Applying many-to-many alignments and hidden markov models to letter-to-phoneme conversion”, HLT-NAACL, Rochester, New York, pp 372–379
17. Jiampojamarn S, Kondrak G (2009) Online discriminative training for grapheme-to-phoneme conversion. In: Proceedings of INTERSPEECH, Brighton, UK, pp 1303–1306
18. Kaplan RM, Kay M (1994) Computational linguistics. In: Regular models of phonological rule systems, vol 20, issue 3. MIT Press, Cambridge, pp 331–378
19. Katz S (1987) Estimation of probabilities from sparse data for the language model component of a speech recognizer. IEEE Trans Acoust Speech Signal Process 35(3):400–401
20. Kneser R, Ney H (1995) Improved backing-off for M-gram language modeling. In: Proceedings of ICASSP, vol 1. pp 181–184
21. Mateus, MH, d’Andrade E (2000) The phonology of Portuguese. Cambridge University Press, USA 18(2):309–312
22. Navarro G (2001) A guided tour to approximate string matching. ACM Comput Surveys 33(1):31–88
23. Ney H, Essen U, Kneser (1994) On structuring probabilistic dependences in stochastic language modelling. Computer Speech Lang 8(1):1–38
24. Oliveira C, Moutinho L, Teixeira A (2004) Um Novo Sistema de Conversão Grafema-Fone para PE Baseado em Transdutores”, Actas do II Congresso Internacional de Fonética e Fonologia, Maranhão, Brazil.
25. Oliveira LC, Viana MC, Trancoso IM (1992) A rule-based text-to-speech system for Portuguese. In: Proceedings of ICASSP, vol. 2. San Francisco, USA, pp 73–76
26. Santos D, Rocha P (2001) Evaluating CETEMPúblico, AFree Resource for Portuguese”. In: Proceedings of the 39th annual meeting of the association for computational linguistics, Toulouse, France, pp 442–449
27. SpeechDAT (1998) Portuguese SpeechDat(II) FDB-4000, European Language Resources Association. http://www.elda.org/catalogue/en/speech/S0092.html
28. Taylor P (2005) Hidden markov models for grapheme to phoneme conversion. In: Proceedings of INTERSPEECH, Lisbon, Portugal, pp 1973–1976
29. Teixeira JP (2004) A prosody model to TTS systems. PhD Thesis, Faculdade de Engenharia da Universidade do Porto
30. Wells JC (1997) SAMPA computer readable phonetic alphabet. In: Gibbon D, Moore R, Winski R (eds) Handbook of standards and resources for spoken language systems, Part IV. Berlin, Mouton de Gruyter
31. Witten I, Bell T (1991) The zero-frequency problem: estimating the probabilities of novel events in adaptive text compression. IEEE Trans Inf Theory 37(4):1085–1094
32. Ribeiro R, Oliveira LC, Trancoso I (2003) Using morphossyntactic information in TTS systems: comparing strategies for European Portuguese. In: PROPOR’2003—6th workshop on computational processing of the Portuguese Language. Springer, Heidelberg, pp 143–150
33. Ribeiro, R, Oliveira, LC, Trancoso I (2002) Morphossyntactic Disambiguation for TTS Systems. In: Proceedings of the 3rd international conference on language resources and evaluation, vol V. pp 1427–1431 (ELRA)
34. Braga D, Marques MA (2007) Desambiguação homógrafos para Sistemas de conversão Texto-Fala em Português”, Diacrítica, 21.1 (Série Ciências da Linguagem) Braga: CEHUM/Universidade do Minho, pp 25–50
35. Seara I, Kafka S, Klein S, Seara R (2001) “Considerações sobre os problemas de alternância vocálica das formas verbais do Português falado no Brasil para aplicação em um sistema de conversão Texto-Fala”, SBrT 2001—XIX. Simpósio Brasileiro de Telecomunicações, Fortaleza, Brazil
36. Seara I, Kafka S, Klein S, Seara R (2002) Alternância vocálica das formas verbais e nominais do Português Brasileiro para aplicação em conversão Texto-Fala. Revista da Sociedade Brasileira de Telecomunicações 17(1):79–85
37. Barbosa F, Ferrari L, Resende F Jr (2003) A methodology to analyze homographs for a Brazilian Portuguese TTS system. In:

PROPOR’2003— 6th workshop on computational processing of the Portuguese Language. Springer, Heidelberg

38. Ferrari L, Barbosa F, Resende F Jr (2003) Construções gramaticais e sistemas de conversão texto-fala: o caso dos homógrafos. In: Proceedings of the international conference on cognitive linguistics, Braga

39. Silva D, Braga D, Resende F Jr (2009) Conjunto de Regras para Desambiguação de Homógrafos Heterófonos no Português Brasileiro. In: XXVII Simpósio Brasileiro de Telecomunicações — SBrT 2009, September 29–October 2, Blumenau, Santa Catarina, Brazil, vol 1. pp 1–6