

Portuguese Pseudoword Generator

Roshanak Hamidi, Ignacio Rubio Majano, Andreia Schurt Rauber

Contact: roshanak.hamidi@gmail.com, i.rubio-majano@student.uni-tuebingen.de,
asrauber@gmail.com

March 11, 2016

Abstract

This paper aims at reporting the method adopted to create a pseudoword generator for Portuguese. Pseudowords were generated taking into account the permissible graphemes that form syllable onsets, peaks and codas, respecting Portuguese phonotactics. In order to weight the frequency with which each grapheme should be used to generate syllables, the frequency of occurrence of its equivalent phoneme(s) was considered. Although the phoneme frequency was based on a European Portuguese corpus (P-PAL, Soares et al., 2010, 2014), the words generated are also suitable for other Portuguese varieties, such as Brazilian Portuguese.

Introduction

Many psycholinguistic studies make use of pseudowords in their experiments. Pseudowords, legal nonwords, nonwords, nonsense words or nonce words are nonexisting words whose structure obeys the phonotactic rules, i.e., the permissible phoneme sequences, of a given language. For example, the pseudoword *squartel* does not violate English phonotactic rules: [skw] and [t] are possible syllable onsets, [ɑ:] and [ø] are possible syllable peaks, and [ɹ] and [t] are possible codas. However, *squartel* is meaningless.

In psycholinguistics, the use of pseudowords as part of the stimuli of an experiment is very common in lexical decision tasks. This kind of task measures participants' reaction time to determine if a string of letters, generally shown on a computer screen or presented aurally, is a real word or a pseudoword. Reaction time is shown to be lower to recognize real words than pseudowords. Moreover, the accuracy of responses is also greater for real words (Warren, 2012, p. 140).

Pseudowords are also used in priming experiments. Priming is a kind of memory effect that has to do with the identification of words or objects and can be observed by means of an experimental procedure which consists of testing participants' reaction time to a target stimulus that is preceded by another stimulus (the prime), which might be somehow related to the target stimulus. The idea is to activate representations or associations in memory before responding to the target stimulus. One example is to show the priming word "grape" on a computer screen followed by the word "wine". Participants are expected to recognize "wine" faster if they see the priming word "grape" than the priming word "ocean" because "grape" and "wine" are more closely associated in memory than "ocean" and "wine". When pseudowords are used as primes, reaction times tend to be slower than when real words are shown.

Considering the importance of pseudowords for psycholinguistic studies, this paper aims at describing how a pseudoword generator was created in Python. For example, one multilingual pseudoword generator for Dutch, English, German, French, Spanish, Serbian and Basque, named Wuggy, was developed by Keuleers and Brysbaert (2010), but we could not find any tool for Portuguese. Therefore, the Portuguese Pseudoword Generator application we developed could be helpful for researchers who work with that language.

The main difference between Wuggy and our application is that Wuggy makes use of a list of syllabified words to determine what graphemes should be used. For example, if the user wants to generate a monosyllabic pseudoword, only the possible elements of monosyllabic real words will be used as candidates; if the user wants disyllabic pseudowords, only the structure of disyllabic real words will be considered, and so on. Moreover, Wuggy does not make use of phonemic representations, only graphemes. Considering that searching in the whole database of real words in order to generate pseudowords would be excessively costly, Wuggy users are asked to input a real word that will be used as a template to generate pseudowords. For instance, the English word *bridge* is segmented as br-i-dge. This way, Wuggy generates pseudowords by replacing the high frequency segments of the real word by other high frequency candidate segments, and the low frequency segments of the real word by other low frequency candidate segments. Differently from Wuggy, the Portuguese Pseudoword Generator disregards the frequency of graphemes in a given syllable position and solely makes use of (i) the frequency of occurrence of isolated phonemes and (ii) the transitional probabilities between the phonemes in the consonant clusters and diphthongs that are present in a Portuguese corpus (see Section 2.1) in order to generate pseudowords.

This paper is divided into three sections. In Section 1 we will describe the Portuguese syllable structure and the graphemes used to generate pseudowords. In Section 2 we will (i) report the procedures adopted to give weightings to the

graphemes used to form syllables; (ii) explain the workflow used to write the Portuguese Pseudoword Generator; and (3) present the graphical user interface (GUI) of the application. In Section 3 we will report the limitations of our application and suggestions for improvement.

1 The Portuguese syllable structure

The definition of a syllable is considerably controversial, but one possible explanation is to take the pulmonic airstream mechanism into account (Abercrombie, 1976). When one speaks, the airstream that comes from the lungs is not constant or homogeneous, but it is expelled as successive puffs of air. Each puff of air is the articulatory basis of a syllable.

Each syllable has two components: the onset and the rime. In Portuguese, the onset can be either a single consonant or consonant cluster with no more than two consonants. The rime has two elements: the peak and the coda. In Portuguese, only a single vowel or a diphthong can form a syllable peak, and the coda can consist of a single consonant or no more than two consonants (Cristófar-Silva, 2009, Mateus et al., 1990). In sum, the Portuguese syllable structure can be described as (C)(C)V(V)(C)(C), where C stands for consonant, V for vowel and the parentheses represent optional elements. Table 1.1 shows possible syllable structures in Portuguese with an example word, its phonetic transcription and its translation to English. In this paper, the phonetic/phonemic transcriptions will be made in European Portuguese. However, considering the last spelling reform, the results generated by our application are also valid for other Portuguese varieties.

Table 1.1 Examples of Portuguese syllable structures

Syllable structure	Example in Portuguese	English translation
V	<i>é</i> ['ɛ]	is
VC	<i>ar</i> ['ar]	air
VCC	<i>us</i> ['ũʃ]	some
CV	<i>tu</i> ['tu]	you
CVC	<i>lar</i> ['lar]	home
CVCC	<i>mons.</i> <i>tro</i> ['mõʃ.tru]	monster
CCV	<i>tri</i> ['tri]	tri
CCVC	<i>três</i> ['treʃ]	three
CCVCC	<i>trans</i> ['trãs]	trans
VV	<i>eu</i> ['ew]	I
CVV	<i>lei</i> ['lej]	law
CCVV	<i>grau</i> ['graw]	degree
CCVVC	<i>claus.</i> <i>tro</i> ['klawʃ.tru]	cloister

As regards the individual letters and digraphs used in Portuguese, Table 1.2 shows the possible phonemes in onset, peak and coda positions and their equivalent graphemes within example words.

Table 1.2 Possible Portuguese phonemes and examples of their corresponding graphemes in onset, peak and coda positions (English translations within parentheses)

	Phonemes	Orthography
Onset		
CC	/pr, br, tr, dr, kr, gr, fr, vr, pl, bl, tl, kl, gl, fl/	<i>pr</i> ato (plate), <i>bra</i> ço (arm), <i>tr</i> az (bring), <i>dro</i> ga (drug), <i>cr</i> er (believe), <i>gr</i> ave (serious), <i>fr</i> uta (fruit), <i>livro</i> (book), <i>pl</i> uma (feather), <i>pr</i> isma (prisma), <i>at</i> las (atlas), <i>cl</i> ima (climate), <i>gl</i> obo (globe), <i>fl</i> or (flower)
C	/p, t, k, b, d, g, f, v, s, z, ʃ, ʒ, m, n, ŋ, l, λ, r, ʁ/	<i>p</i> ato (duck), <i>t</i> ato (tact), <i>ca</i> sa (house), <i>qu</i> erer (want), <i>be</i> ge (beige), <i>da</i> ma (lady), <i>ga</i> to (cat), <i>fa</i> ca (knife), <i>va</i> so (vase), <i>sa</i> co (bag), <i>ze</i> bra (zebra), <i>cha</i> to (boring), <i>ja</i> to (jet), <i>ma</i> pa (map), <i>na</i> ta (cream), <i>nin</i> ho (nest), <i>la</i> ta (can), <i>fil</i> ho (son), <i>ca</i> ra (face), <i>car</i> ro (car)
Peak		
V	/i, e, ε, a, ɔ, o, u, ã, ê, ĩ, õ, ũ/ ((pre)tonic) /ɪ, ɐ, ʊ/ (posttonic)	<i>di</i> a (day), <i>mer</i> cado (market), <i>le</i> que (fan), <i>ca</i> ra (face), <i>co</i> po (glass), <i>co</i> mo (how), <i>cu</i> ra (cure), <i>sa</i> nta (saint), <i>se</i> mpre (always), <i>ci</i> nta (belt), <i>co</i> nta (bill), <i>um</i> (one) <i>po</i> n-te (bridge), <i>ca</i> sa (house), <i>co</i> mo (how)
VV	/jɐ, jʊ, wɐ, wə, wu, aj, ej, ɛj, ɔɪ, oj, uj, aw, ew, ɛw, iw, ẽj, ẽw, õj, ũj/	<i>sé</i> ria (serious), <i>ce</i> ná-rio (scenario), <i>á</i> gua (water), <i>tê</i> nue (thin), <i>vá</i> cuo (vacuum), <i>pa</i> i (father), <i>lei</i> (law), <i>pas</i> téis (cakes), <i>he</i> rói (hero), <i>boi</i> (ox), <i>fui</i> (I went), <i>mau</i> (bad), <i>seu</i> (your), <i>céu</i> (sky), <i>riu</i> (he laughed), <i>mãe</i> (mom), <i>mão</i> (hand), <i>põe</i> (put), <i>ruim</i> (bad)
Coda		
C	/ʃ, ʒ, r, m, n, l/	<i>ma</i> s (but), <i>mes</i> mo (even), <i>ma</i> r (sea), <i>tem</i> po (time), <i>ci</i> nta (belt), <i>me</i> l (honey)
CC	/Nʃ/	<i>trens</i> (trains)

After the brief description of the Portuguese syllable structure, fundamental for the development of the Portuguese Pseudoword Generator, the next section will describe the method used to develop our application.

2 Method

In this section, we will (i) describe the Portuguese corpora used to create the Portuguese Pseudoword Generator; (ii) explain how we computed the weightings for each grapheme that was used to form a pseudoword; (iii) show the workflow used to generate pseudowords in Python; and (iv) display the GUI of our application.

2.1 The Portuguese corpora

In order to generate pseudowords in Portuguese, we considered both its syllable structure and graphemes to make combinations of letters become pseudowords. The corpus used for this task was the Procura-PALavras Corpus (P-PAL, Soares et al., 2010, 2014, <http://p-pal.di.uminho.pt/about/project?lang=en>). P-PAL stands for “search words” and was designed to provide users with lexical frequency data compiled from eight European Portuguese corpora, totaling 227,770,752 words from texts of different genres (news (55.5%), literary (14%), technical-scientific and didactic (12%), miscellanea (6%) and oral (12.5%)) and modality (oral (12.5%) and

written (87.5%)). The variety of genres and modalities provides users with a considerable language diversity and great number of low-frequency entries.

After analyzing, treating and compiling the eight corpora that form P-PAL, there was a total of 207,614 forms (the natural occurrence of a word, such as the conjugated forms *plays* and *played*) and 52,404 lemmas (the abstract lexical item (or dictionary form) used to represent all forms, such as *play*). All the forms were phonemically transcribed manually. The phonemes of the P-PAL Corpus constitute the raw input to calculate the frequencies of individual phonemes and the transitional probabilities of consonant clusters and diphthongs, essential information to design the Portuguese Pseudoword Generator, as described in Section 2.2.

After generating strings of characters, the Portuguese Pseudoword Generator compared these strings to the list of real words compiled from two sources: (i) the European Portuguese corpus P-PAL, and (ii) the Brazilian Portuguese entries freely available at the Dicionário Criativo's (Creative Dictionary) homepage: www.dicionariocriativo.com.br. Since the two sources contained entries with some sort of punctuation (hyphens, parentheses or periods), these punctuations were first removed from the 207,614 forms from P-PAL and the 138,795 entries from Dicionário Criativo. Then, the two corpora were combined and all duplicates were deleted. Therefore, a total of 237,839 real words were used to check if the generated strings of characters were indeed pseudowords. In case a real word was detected, it was deleted and a new pseudoword was generated.

After presenting the two Portuguese corpora used in this study, the next section will describe how weightings were calculated in order to generate pseudowords according to the frequency of occurrence of each phoneme in the P-PAL Corpus.

2.1 Weighting graphemes

When forming pseudowords, one of our concerns was to first determine the weight each individual letter/digraph (or grapheme) should have. Thus, we took into account the information available in the P-PAL corpus and computed (i) the frequency of occurrence of each individual phoneme; and (ii) the transitional probabilities between the phonemes of a consonant cluster (e.g., the probability of /p/ to be followed by /r/), as well as the transitional probabilities between a consonant sound in word-final position and silence (e.g., the probability of /l/ to be followed by #, which is the word boundary symbol in the corpus). Moreover, we considered the mappings of phonemes and letters. For example, the letter <a> can be phonemically transcribed as /a/ if in stressed position and as /e/ if in unstressed position. In these cases, the frequency of occurrence of the two isolated phonemes was summed.

As regards the calculations, the frequency of occurrence of each phoneme was obtained by counting the number of types of a given token and dividing it by the total number of tokens. The weight of each grapheme was calculated based on their phonemic representation. Since the least frequent grapheme should be considered as a candidate to form a pseudoword at least once, the starting point was its lowest frequency of occurrence (e.g., <nh> (/ɲ/) = 0.002, thus this value determines the

minimum weight, or weight = 1). For example: the frequency of occurrence of <i> in the corpus is 0.0669 and its weight is 33 ($0.0669/0.002 = 33$). Therefore, the grapheme <i> should be considered a possible candidate to form a syllable 33 times, since it occurs 33 times more often than <nh> in the corpus. Tables 2.1 and 2.2 show the frequency of occurrence and weightings for all graphemes. The phonemic representations were added so as to inform the origin of the calculations, i.e., if a grapheme represents one or two phonemes.

In the cases when one grapheme represented two phonemes, the sum of the weights of the two phonemes was considered. For example: <a> sounds like /a/ in stressed syllables and like /e/ in unstressed syllables. Thus, the weighting of /a/ was summed to the weighting of /e/ ($20 + 47$), thus resulting in the grapheme <a> being present in the array 67 times.

In the case of the 14 consonant clusters in onset (C+/r/ and C+/l/) and the diphthongs, first the transitional probabilities between the two elements of the cluster or diphthong were measured. Since the weightings of the consonant clusters were over twice as much as that of the most frequent single grapheme (<a> = 67), their weightings were normalized considering the weighting of <a>. For example, the value for the transitional probability of <pr> (0.2225) was divided by 67 (<a>), resulting in 0.003321. The latter value was then divided by the lowest value for clusters (i.e., <tl> = 0.0006), thus resulting in weight 6. The same procedure was adopted to calculate the weightings of diphthongs. Because the clusters <vr>, <gl>, <dr> and <tl> resulted in weightings smaller than 1, their weight was considered 1.

Table 2.1 Frequency of occurrence and weights of graphemes in coda-medial (word internal) and coda-final (word final) positions

Coda medial	Frequency	Weight	Coda final	Frequency	Weight
s (/ʃ, ʒ/)	0,0644	32	s (/ʃ/)	0,7427	5
r (/r/)	0,0609	30	r (/r/)	0,1821	1
m (/ã, ê, ĩ, õ, ũ/)*	0,0601	15	m (/ã, ê, ĩ, õ, ũ/)*	0,8540	3
n (/ã, ê, ĩ, õ, ũ/)*+	0,0601	15	n (/ã, ê, ĩ, õ, ũ/)*+		2
l (/ʎ/)	0,0242	12	ns (/Nʃ/)+		2
b (/b/)	0,0087	4	l (/ʎ/)	0,1646	1
d (/d/)	0,0625	31			
g (/g/)	0,0087	4			
p (/p/)	0,0298	15			
t (/t/)	0,0506	25			
c (/k/)	0,0372	19			

* The frequencies for <m> and <n> were divided by 2 because the phonemic transcription is the same to indicate these two graphemes.

+ <n> and <ns> occur graphemically less frequently than <m> word finally. Since we did not have the exact number of occurrences, we used <m> as reference and considered these two graphemes as slightly less frequent than <m>.

Table 2.2 Frequency of occurrence and weights of graphemes in onset and peak positions

Onset	Frequency	Weight	Peak	Frequency	Weight
pr (/pr/)	0,2225	6	a(/a, ɐ/)	0,1346	67
br (/br/)	0,2044	5	e(/e, ə/)	0,0792	40
tr (/tr/)	0,1261	3	i (/i/)	0,0669	33
dr (/dr/)	0,0044	1	o (/o, u/)	0,0964	48
cr (/kr/)	0,0381	1	u (/u/)	0,0831	42
gr (/gr/)	0,1802	4	an (/ã/)	0,0259	13
fr (/fr/)	0,0613	2	en (/ẽ/)	0,0132	7
vr (/vr/)	0,0127	1	in (/ĩ/)	0,0053	3
pl (/pl/)	0,0308	1	on (/õ/)	0,0108	5
bl (/bl/)	0,0842	2	un (/ũ/)	0,0049	2
tl (/tl/)	0,0006*	1	eu (/ew/)	0,1541	23
cl (/kl/)	0,0217	1	au (/aw/)	0,0963	14
gl (/gl/)	0,0113	1	ia (/je/)	0,1253	19
fl (/fl/)	0,0204	1	io (/ju/)	0,1247	19
p (/p/)	0,0298	15	ua (/we/)	0,0281	4
t (/t/)	0,0506	25	ai (/aj/)	0,0714	11
c (/s, k/)	0,0758	38	ei (/ej/)	0,0502	7
b (/b/)	0,0087	4	ui (/uj/)	0,0036	1
d (/d/)	0,0625	31	iu (/iw/)	0,0094	1
g (/g/)	0,0087	4	ie (/jə/)	0,0001*	1
f (/f/)	0,0105	5	uim(/wĩ/)	0,0015	1
v (/v/)	0,0138	7			
s (/s, ʃ/)	0,0958	48			
ss (/s/)	0,0386	19			
z (/z/)	0,0096	5			
ch (/ʃ/)	0,0572	29			
j (/ʒ/)	0,0305	15			
m (/m/)	0,0276	14			
n (/n/)	0,0217	11			
nh (/ɲ/)	0,002*	1			
l (/l/)	0,0242	12			
lh (/ʎ/)	0,0023	1			
r (/r, ʀ/)	0,0679	34			
rr (/ʀ /)	0,007	4			

* The numbers in bold were the minimum values used to calculate the weights of consonant clusters, individual phonemes and diphthongs.

After reporting the number of times each grapheme should be considered as a possible candidate when generating pseudowords, the next section will describe how the code was written in Python.

2.2 Generating pseudowords in Python

In this section, we will present the workflow used to generate Portuguese pseudowords in Python:

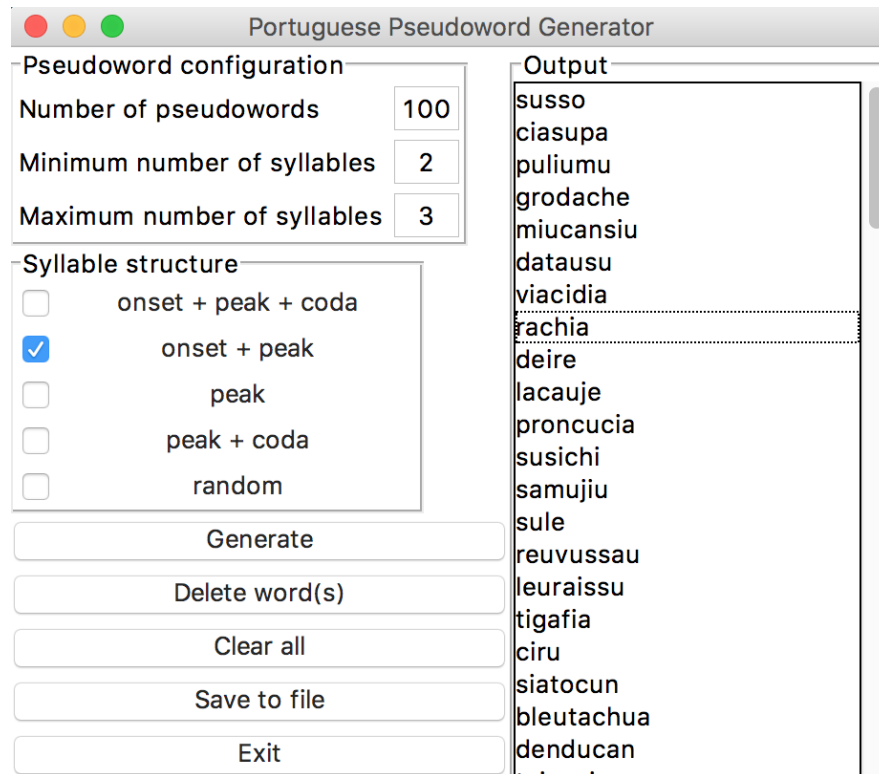
1. ask the user to input the number of pseudowords to be generated, as well as the minimum and maximum number of syllables each pseudoword should have;
2. ask the user to select the syllable structure among five options: (i) onset+peak+coda; (ii) onset+peak; (iii) peak; (iv) peak+coda; (v) random;
3. according to the user's input, activate the corresponding functions among the following: (i) generateOnsetComp, (ii) generatePeakComp, (iii) generateCodaMidComp, and (iv) generateCodaComp. Functions (i) and (ii) consider all graphemes allowed in onset and peak positions, respectively; Function (iii) considers only the graphemes which are possible in coda position word internally, while Function (iv) considers only the graphemes which are possible in coda position word finally. These functions concatenate graphemes that will constitute each syllable type by randomly creating an integer and checking what grapheme this integer corresponds to, always respecting the grapheme weight distribution (Tables 2.1 and 2.2). For example, the generatePeakComp Function considers integers that range between 1 and 362. If the integer 77 is generated, the grapheme <e> will be considered, since it falls within the 68-108 range. This range is determined by the weight each grapheme has according to the frequency of occurrence of its equivalent phoneme(s) in the P-PAL Corpus;
4. generate syllable structures according to the user's input (options (i) through (v) of Item 2);
5. add an exception when the structure to be generated is either onset+peak+coda or onset+peak: in these cases, the graphemes <nh>, <lh>, <vr>, <tl>, <ss>, <rr> can only occur word medially, never word initially;
6. check if the generated pseudowords are real words by comparing them with the list of 237,839 Portuguese words compiled from the P-PAL Corpus and the Dicionário Criativo entries. If the generated string of letters is a real word, delete it and keep generating a new string until it is a pseudoword;
7. print the generated pseudowords in the GUI listbox;
8. add warning messages in case the user (i) inputs a value for the minimum number of syllables that is smaller than the maximum number of syllables; (ii) inputs no value at all to configure the pseudowords or forgets to inform one of the requested values; (iii) selects none of the syllable structure types.

Once the pseudowords are generated and printed in the GUI, the user has some options to work with them. These options will be presented in the next section.

2.3 The graphical user interface (GUI)

The GUI of the Portuguese Pseudoword Generator was designed to be self-explanatory and easy to use. Its layout is shown in Figure 2.1.

Figure 2.1 GUI of the Portuguese Pseudoword Generator



As already explained in Section 2.2, the first pieces of information the user must input are (i) the number of pseudowords to be generated; (ii) the minimum and maximum number of syllables the pseudowords must have; and (iii) their syllable structure. Once these parameters have been selected, the user must click on the *Generate* button so as to see the generated pseudowords printed in the listbox.

As regards the printed pseudowords, the user can select a single entry or multiple entries and delete them by using the *Delete* button or s/he can delete all entries at once by clicking on the *Clear all* button. The user can also save the list of pseudowords to a .txt file by clicking on the *Save* button. The .txt file will be saved in the directory where the application is saved and the file name will be “Words_[number of pseudowords]-[minimum number of syllables]-[maximum number of syllables]”. Thus, if a user wants to generate, for example, 100 pseudowords, with the number of syllables ranging from one to three, the file name will be “Words_100-1-3”.

If the user wants to generate more pseudowords, s/he can click on the *Generate* button again. The next batch of pseudowords will be printed below the content already available in the listbox in case the user does not click on the *Clear all* button before generating a new batch. To terminate the program, the user can click on the *Exit* button.

3 Final remarks

In this paper, we reported the procedures adopted to design a pseudoword generator specific for Portuguese. We first described the Portuguese syllable structure, then reported the frequency of occurrence of graphemes in the P-PAL Corpus, and after that described the workflow used to generate the application in Python. Finally, we presented the GUI and some simple operations that can be done with the pseudowords once they have been generated and printed in the listbox.

One limitation of the Portuguese Pseudoword Generator is that it does not take into account the frequency of occurrence of a grapheme in a given syllable position, i.e., it only considers the frequency of occurrence of the Portuguese phonemes in the P-PAL corpus and maps phonemes to their respective graphemes. Since the frequency of occurrence of graphemes in the P-PAL corpus was not controlled, no graphemes containing diacritics (e.g., ã, á, à, â) were included as possible candidates to form pseudowords. Considering that our tool takes syllable position into account, a future version could add restrictions so as to encompass letters containing diacritics. Moreover, transitional probabilities between phonemes are computed only for consonant clusters or diphthongs, not for combinations of individual phonemes.

Despite its limitations, the Portuguese Pseudoword Generator can already be considered a valuable and reliable tool for generating a large number of pseudowords that take into account both Portuguese phonotactics and the frequency of occurrence of phonemes in this language.

Wir versichern, dass wir die Arbeit ohne Benutzung anderer als der angegebenen Quellen angefertigt haben und dass die Arbeit in gleicher oder ähnlicher Form noch keiner anderen Prüfungsbehörde vorgelegen hat und von dieser als Teil einer Prüfungsleistung angenommen wurde. Alle Stellen und Personen, welche uns bei der Vorbereitung und Anfertigung der Abhandlung unterstützten, wurden genannt und Ausführungen, die wörtlich oder sinngemäss uebernommen wurden, sind als solche gekennzeichnet.

Tübingen, den 11. März 2016

Roshanak Hamidi, Ignacio Rubio Majano, Andreia Schurt Rauber

References

Abercrombie, D. (1976). *Elements of general phonetics*. Edinburgh University Press, Edinburgh.

Cristófar-Silva, T. (2009). *Fonética e fonologia do português: Roteiro de estudos e guia de exercícios [Portuguese phonetics and phonology: Study guide and exercises]*. Editora Contexto, São Paulo.

Dicionário Criativo [Creative Dictionary], retrieved 7 February 2016, from <http://dicionariocriativo.com.br/>

Keuleers, E., & Brysbaert, M. (2010). Wuggy: A multilingual pseudoword generator. *Behavior Research Methods*, 42(3), 627-633.

Mateus, M. H. M., Andrade, A., Viana, M. C., Villalva, A. (1990). *Fonética, fonologia e morfologia do português [Portuguese phonetics, phonology and morphology]*. Universidade Aberta, Lisboa.

Soares, A. P., Comesaña, M., Iriarte, A., Almeida, J. J., Simões, A., Costa, A., França, P., & Machado, J. (2010). P-PAL: Uma base lexical com índices psicolinguísticos do Português Europeu [P-PAL: A European Portuguese lexical database]. *Linguamática*, 2(3), 67-72.

Soares, A. P., Iriarte, A., Almeida, J. J., Simões, A., Costa, A., França, P., Machado, J., & Comesaña, M. (2014). Procura-PALavras (P-PAL): Uma nova medida de frequência lexical do Português Europeu contemporâneo [Procura-PALavras (P-PAL): A new measure of word frequency for contemporary European Portuguese]. *Psicologia: Reflexão e Crítica*, 27(1), 1-14.

Warren, P. (2012). *Introducing psycholinguistics*. Cambridge University Press.