Nativization of English words in Spanish using analogy

Tatyana Polyákova, Antonio Bonafonte

Universitat Politècnica de Catalunya, Barcelona, Spain
tatyana.polyakova@upc.edu, antonio.bonafonte@upc.edu

Abstract

Nowadays modern speech technologies need to be flexible and adaptable to any framework. Mass media globalization introduces the challenge of multilingualism into most popular speech applications such as text-to-speech synthesis and automatic speech recognition. Mixed-language texts vary in their nature and when processed, some essential characteristics ought to be considered. In Spain, the usage of English and other foreign origin words is growing as well as in other countries. The particularity of the peninsular Spanish is that there is a tendency to nativized foreign words pronunciation so that they fit in properly into Spanish phonetics. In this work our goal was to approach the nativization challenge by data-driven methods, since they are transferable to other languages and do not yield in performance. Training and test corpora for nativization were manually crafted and the experiments were carried out using pronunciation by analogy. The results obtained were encouraging and proved that even a small training corpus of 1000 words allows obtaining a higher level of intelligibility for English inclusions in Spanish utterances.

Index Terms: nativization, grapheme-to-phoneme conversion, phoneme-to-phoneme conversion, Spanish TTS, pronunciation by analogy

1. Introduction

Spain is a country of a remarkable linguistic patrimony, which is of course a cultural treasure but when it comes to speech technologies it represents an additional challenge. Speech technologies in the framework of their rapidly expanding usage must be adapted to the multilingual scope allowing a higher level of flexibility and answering the modern users' needs. Today in Spain it is a usual thing to hear proper names from all over the world. The text-to-speech synthesis finds many important applications on the emerging market of speech technologies. Voices capable of embracing more than one language are highly demanding in the era of mass media globalization. The TTS systems are used in telephone companies, hearing-aid products, and recently in speech-to-speech translation, a technology that is highly demanded due to the globalization of the world industry and mass media.

Every language receives a constant incoming flow of new words. In addition to the natural process of appearance of neologisms, by morphological or semantic word and word meanings creations, a lot of new words come to the current language from foreign languages. The are several ways that the words of foreign origin are incorporated into a receptor language.

Very few databases containing non-native pronunciation are available, while the nativization corpora is simply inexisting. This need for training data lead us to a creation a minimalistic nativization corpus described in Section 1.2. In order to have a synthesizer always up-to-date we need an ultimate automatic method for the derivation of the nativized pronunciation. The problem of foreign words, more particularly, of proper names of foreign origin was studied in [1]. The goal in [1] was to trancribe proper names of different origins correctly from the point of view of English phonetics. The nativization problem and different influencing factors were also described in [2] and [3]. Summarizing all possible influence factors and the difficulties encountered for the correct nativization of foreign words we are betting on an approach that can combine the knowledge of the orthographic and phonetic forms in the language of origin with pronunciation adaptation rules to the target language. The approach that we are willing to use here is called pronunciation by analogy, previously used by [4, 5] to solve the task of grapheme-to-phoneme conversion. The results obtained using the full inventory of eleven strategies available for choosing the best pronunciation were found to be the best in automatic grapheme-to-phoneme(g2p) conversion in comparison with n-gram and HMM based approaches [5]. We believe that the analogy between the nativized pronunciation and the original one can be inferred in an even more reliable and simpler way since the nativization of English words in a Spanish text is a easier task than the pronunciation of unknown English words and yet all human attempts to nativization are highly dependable on the analogy between known and unknown words. The final goal of the nativization is to be able to produce highly intelligible synthesized speech that would be well accepted by native speakers of Spanish with only some knowledge of English as well as by native English speakers with considerable Spanish background.

1.1. Information sources

We can obtain the information about correct pronunciation of foreign words from different sources. For instance, the *book of styles* [6], used by the television channels and radio stations can give us the first idea of how different foreign words should be adapted to the official language. The tendencies for the pronunciation of frequently used words are rather clearly defined, yet the degree of multilingualism for spoken programs is considerably inferior to that of written texts. Usually, during a news flash the only foreign words to appear are the proper names and already orthographically assimilated foreign words. Nonetheless, in order to synthesize high quality intelligible speech from multilingual texts it is necessary to be able to pronounce any new word that we would come across. The criteria used differ based on the frequency of usage of the word in the language and the target auditory. Unfortunately, only a small percentage of Spanish TV viewers are fluent in English according to the recent statistics.

1.2. Nativization corpus

The rule-based approaches to phonemization require significant amount of linguistic engineering, besides they are always language dependent and are not flexible at all. Data-driven approaches were proven to be more efficient than the ones based on the explicit linguistic modeling and they undoubtedly gain in adaptability [7]. The main idea of this work was to train a nativization model to convert English pronunciation to an acceptable pronunciation in Spanish. Two ways to do it were proposed: to train a nativization model using the information about the orthographic form and the nativized phonetic transcription and to use the original English pronunciation together with the nativized pronunciation for training. In order to apply data-driven techniques to nativization a need for training and test data raised. For usual grapheme-to-phoneme conversion tasks large pronunciation corpora of 100 thousands words and their corresponding pronunciations are available. Since we did not find any existing nativization database we chose to manually create a minimalistic corpus, that would not require expert linguistic knowledge. For our task it was necessary that the training corpus be orthographically balanced in order to infer nativized pronunciation by analogy, see Section 3. A greedy corpus-balancing tool was used for selecting words to be nativized from the available LC-STAR [8] dictionary of U.S. English with more than 50K entries previously used by the authors for grapheme-to-phoneme conversion experiments [5]. In order to have all possible letter bi-grams in the corpus we selected 1000 words. The original phonetic transcriptions of these words were manually nativized according to the criteria described by Llorente in the book of styles for one of the Spanish TV channels [6]. It is necessary to stress out that the phoneme inventory used for nativization was limited to the Spanish phoneset, defined in Sampa website [9], including two allophones [N] and [z]. The proportion of the rare words in the resulting corpora is noticeable, however, a few words of non-English origin were removed since their pronunciation did not obey English phonetic rules therefore their presence in the nativization corpus could have introduced additional ambiguity. The test data was manually collected from the available on-line free daily newspaper *www.20minutos.es*. Since a thousand words was selected for training, we decided that the test data should comprise 10% of the training corpus. None of the test words were present in the training dictionary. It was intended that the test words were frequently used and with simple meaning in order for the results to be unbiased by other factors.

2. Overview of a multilingual grapheme-to-phoneme system

In order to work with multilingual text it is necessary to know the language of each word. First of all, it is important to have a tool capable of efficiently determining the language of the paragraph of the mixed-language texts extracted form newspapers on-line forums, emails, scientific articles, technical support manuals, web pages and other sources where the language can vary from word to word. Knowing the source language, the synthesis quality can be improved considerably. Whenever the paragraph language is known, it is necessary to determine the language of each isolated word in the paragraph. It is important to improve the pronunciation of the foreign words, adapting their pronunciation to the language of the paragraph. In current work we assume that all foreign words are labeled and their source language is known. Some results on detection of the language of the paragraph and isolated words was presented in our previous work [10]. The diagram in Figure 1 shows our nativization system given that each foreign word had been assigned a special label, let us call it F_LANG; the default paragraph language therefore would be LANG or "target language." Our pronunciation module consists of system dictionaries in several languages and the corresponding language-specific grapheme-to-phoneme converters. The pronunciation is sought separately for every isolated word. The first step is to find out whether or not the word in question is present in the system dictionary of the target language. In that case, that pronunciation is validated. It is important to notice that if a foreign word is found in the target language dictionary it is already considered nativized, that is why there is no need to check the language before the first step. If the word is absent from the dictionary, the next step is to find out if its language is different from the target language (does it have a F_LANG label?). If no label was found the pronunciation is derived using automatic transcription system for target language. For the words identified as foreign the search continues in the dictionary of their language of origin or "source language". Before validating the pronunciation, in the case it is found in the dictionary, the nativization phoneme-to-phoneme converter is applied to the source language pronunciation. The output of the nativization module is the nativized "target" language-adapted pronunciation. In the last case, if the word is absent from the source language dictionary, its source pronunciation is derived using the automatic transcription system for the source language and afterwards the nativization is applied before validating the pronunciation.

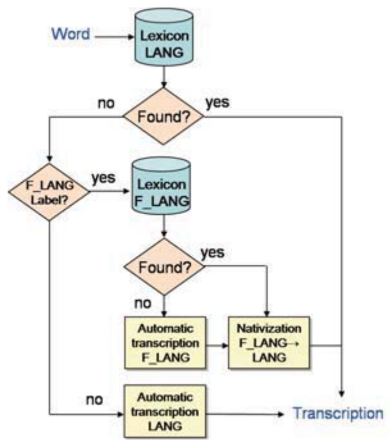

Figure 1: Multilingual g2p system.

3. Pronunciation by analogy

Data-driven approaches were proven to be more efficient than the ones based on the explicit linguistic modeling and they undoubtedly gain in adaptability [7]. For g2p conversion the best results were obtained using data-driven corpus-based methods. Pronunciation by analogy method previously used in [4, 5] was found to be the most efficient for grapheme-to-phoneme task. In this section we review the pronunciation by analogy algorithm. Our implementation is based on [4] with the new strategies introduced in [5]

3.1. Alignment

In order to apply most of the data-driven methods for pronunciation inferring, in the first place it is necessary to make sure that there is a one-to-one match between the orthographic and phonetic strings, or, in other words, each letter has to be aligned with a corresponding phonetic representation. Finding the correct alignment presents a challenge since the orthographic and phonetic representations of a word in English do not always have the same length. Due to its rather complex orthography, English words usually have more letters than sounds. In that case a null phone /_/ need to be inserted into the phoneme string, e.g, *thing* / T _ i N _ /, otherwise, if the number of phonemes is greater than that of letters, the phonemes corresponding to the same letter are joint together in one, e.g. *fox* /f A k_s/. The alignment used is based on EM algorithm, and it is similar to that described in [11]. However, the alignment is not always perfect and it can influence negatively on the results.

3.2. Algorithm description

After the training dictionary has been aligned, the matcher starts to search for common substrings between the input word and the rest of the dictionary entries. Every input word is then compared to all the words in the lexicon in order to find common "arcs". Let us call the substrings in the grapheme context "letter arcs" and the corresponding substrings in the phoneme context "phoneme arcs". All the possible letter arcs with the minimum length of 2 letters and the maximum length equal to the input word length are generated and then searched in the dictionary. For every letter arc from the input word, matching with the same letter arc in a dictionary word, the corresponding pronunciation or the phoneme arc is extracted. The frequency of appearance of each phoneme arc corresponding to the same letter arc is stored along with the starting position and length for each arc. As an example, let us say that the word `#top#` is absent from our dictionary; the list of all possible letter arcs for this word can be given as `#t, #to, #top, to, top, top#, op, op#, p#`. Now, let us suppose that in the lexicon we have the word `#topping#` with the pronunciation /# t A p _ I _ N #/, here the matcher finds the letter arcs `#t, #to, #top,` and `op`, with their corresponding phoneme arcs /# t/, /# t A/, /# t A p/, /A p/.

Each time that for the same letter arc we find the same phoneme arc; the frequency of the phoneme arc is incremented. The matching phoneme arcs are introduced into the pronunciation lattice that can be represented by nodes and connecting arcs. If an arc starts at a position i and ends at a position j, and if there is yet no arc starting or ending at position j, the nodes L_i and L_j are added to the graph. An arc is drawn between them. All the nodes are labeled with the corresponding "juncture" phoneme and its position in the word. The arcs are labeled with the remaining phonemes and their frequency of appearance. An example of the lattice construction for the word *top* using the arcs found in the word *topping* is illustrated in Figure 2. In this example all the arc frequencies are assumed to be equal to 1. These arcs and frequency counts are updated when the search continues through all the words of the dictionary. Each complete path through the lattice is called "pronunciation candidate". We considered only the shortest paths through the lattice [4]. If there is a unique shortest path, it is chosen as the best pronunciation and the algorithm stops. Usually there are several shortest paths through the lattice, and a decision function is necessary to choose the best pronunciation candidate among them.

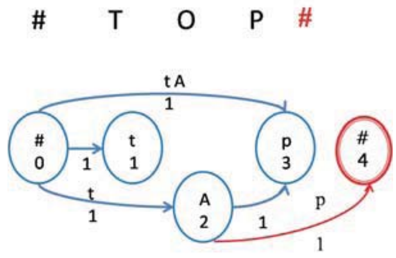

Figure 2: Pronunciation lattice for the word *top* using the arcs extracted form the word *topping*.

Each candidate can be represented as C_j={F_j, D_j, P_j}, where $F_j = \{f_1,\ldots,f_n\}$ are the phoneme arc frequencies along the j^{th} path, $D_j = \{d_1,\ldots,d_n\}$ are the arc lengths and $P_j = \{p_1,\ldots,p_k\}$ are the phonemes comprising the pronunciation candidate, being k the pronunciation length. Marchand and Damper in 2000 [4] proposed to use 5 scoring strategies in order to choose the best pronunciation. Also two ways of strategy combination were introduced. Each strategy gives us a score for each candidate and based on its score each candidate is assigned a rank. According to the rank, each candidate is awarded points. If a strategy gives the same score for several candidates, they are given the same rank and the same number of points.There are two ways of combining strategies to determine the winner candidate; the first one is the sum rule, which chooses the candidate that has the largest value of the sum of points for all of the included strategies. The product rule chooses the candidate with the largest value of product of the points awarded by each of the included strategies. For the best accuracy reported for the NETtalk dictionary was 65.5% words correct and 92.4% phonemes, using all five strategies [4], which is better than using any one the mentioned strategies alone. The sum and the product rules of strategy combination seemed to give similar results.

In our previous work [5] we proposed 6 additional strategies for choosing the best candidate which in combination with the others outperformed the original ones. The scoring strategies are based on the following parameters, frequency of appearance of a given phoneme arc in the dictionary, its length and the actual phonemes which constitute the candidate. Different strategies work with different aspects of analogy. High arc frequency is considered to be a major advantage over the low arc frequency. The frequency of suffixes and prefixes are prioritized by different strategies. The more common phonemes the candidate shares with the others the higher will be its final score. If a candidate has exactly the same pronunciation as the other one both of them are prioritized. These measures are used separately or combined across the strategies.

3.3. The strategies

All the strategies previously used in grapheme-to-phoneme conversion are described below [4, 5].

1. Maximum arc frequency product (PF)
2. Minimum standard deviation of arc lengths ($SDPS$)
3. Highest same pronunciation frequency (FSP)
4. Minimum number of different symbols (NDS)
5. Weakest arc frequency (WL)

6. Weighted arc product frequency (WPF) Similar to 1^{st} strategy described in [4], where for each arc, the corresponding arc frequencies are multiplied $PF(C_j) = \prod_{i=1}^{n} f_i$, being n the candidate length, or the number of arcs that comprise the candidate. Rank 1 is given to the candidate scoring the maximum $PF()$. The difference is that in this strategy for each phoneme arc, A_k the frequency of its appearance is divided by k, the number of different phoneme arcs found in the dictionary for the corresponding letter arc, L_j. For example if our unknown word, is `#infinity#` and if in the pronunciation lattice we have a path that starts with a letter arc, L_1= `#in` and a corresponding phoneme arc A_1=/# @ N/, whose frequency is equal to 12, in order to obtain weighted arc frequency, we have to divide 12 by the number of different phoneme arcs available in the dictionary for the letter arc `#in`.
7. Strongest first arc (SF) This strategy aims at capturing the analogy in prefixes. The candidate with the highest frequency score for the first arc is given rank 1.
8. Strongest last arc (SL) This strategy is analogous to the previous one but for the suffixes. The candidate with the highest frequency score for the last arc is given rank 1.
9. Strongest longest arc (SLN) The candidate who has at the same time the longest and the most frequent arc is given rank 1. First the longest arc is chosen and if there is a tie the next step is to choose the most frequent one. The candidate that have the longest and arcs seem to be more reliable, and of course, the more frequent the arc is the stronger is the analogy.
10. Same symbols multiplied by arc frequency ($SSPF$) The 10^{th} strategy is similar to the fourth one (NDS). NDS gives preference to the candidates whose phonemes appear in the majority of other candidates. $NDS(C_j) = \sum_{i=1}^{l} \sum_{k=1}^{N} \delta(P_{j,i}, P_{k,i})$ being l the number of phonemes in a pronunciation, δ the Kroneker delta, equal to 1 if $P_j \neq P_k$ and 0 otherwise, and N the number of candidates. In our strategy when counting the common phonemes, we also take into consideration the phoneme arc frequencies. For every candidate the pronunciation is. If a candidate has a common phoneme with other candidates, we give it a higher score, depending also on the number of times the phoneme arc containing that phoneme appears in the dictionary $SSPF(C_j) = \sum_{i=1}^{l} \sum_{k=1}^{N} (1 - \delta(P_{j,i}, P_{k,i})) * f_{arc(i)}$
11. Frequency product, same pronunciation ($PFSP$) This strategy is a combination of 1^{st} and 3^{rd} strategies in [4]. The 3^{rd} strategy gives the privilege to the candidates sharing the same pronunciation with the others, rank 1 is given to the candidate scoring the maximum $FSP()$. $FSP(C_j) = \text{cand}\{P_j \mid P_i = P_k\}, j \neq k$ and $\in [1, N]$ In eleventh strategy all the candidates that share the same pronunciation obtain the same score equal to the combination of the scores assigned to each one of the candidates by the 1^{st} strategy $PFSP(C_j) = \sum_{\forall k, P_k = P_j} \sqrt[n]{PF(C_k)}$.

The pronunciation by analogy algorithm was previously applied to grapheme-to-phoneme conversion [12, 4, 5]. In this work it was extended to the nativization task.

4. Experimental results

The experimental results are given below for each method.

4.1. Previous results: nativization tables

In our previous work [10] we developed a nativization system based on nativization tables (Ntab). Pronunciations were derived according to the scheme shown in Figure 1. The nativization was carried out in a phoneme-to-phoneme manner, using nativization tables for source→target phoneme transformations. The source language was U.S. English and the target language was Spanish. Therefore all English phonemes were mapped to the closest Spanish ones. The nativization tables were able to convert 73.88% phonemes and 23.81% words correct. These results are given for the same 100 word test corpus described in 1.2. However, these results are much better than those obtained without using nativization, applying the Spanish g2p to derive the pronunciation of English words, Spanish g2p scored only 61.16% correct in phoneme ans 8.57% on word nativization on a 100 word test corpus. The only words that this kind of system can “nativize” correctly are those that are pronounced very closely to their orthography, for example *bed to* /b e D/ or *car to* /k a r/.

4.2. Grapheme-to-phoneme nativization by analogy (g2p_nat)

The first hypothesis to be tested was prediction of nativized pronunciation by analogy in the orthographic context. Out of eleven strategies available in the PbA for choosing the best pronunciation candidate it was necessary to determine the best strategy combination for our data. An n-fold cross evaluation was carried out on the training dataset, leaving out each word at a time and using the remaining words for pronunciation lattice construction described in Section 3. All possible strategy combinations were considered and compared. For grapheme-to-phoneme nativization (g2p_nat) the resulting best strategy combination was the following: 10001001011 (1 means that the strategy corresponding to that position was included and 0 means it was left out). The best results obtained on training data equaled to 85.73% in phoneme and 45.63% in word accuracy. After having determined the best combination a series of experiments were carried out on the test dataset of 100 English common names. The results obtained are 84.17% phonemes correct and 43.81% words correct. These results do not include stress prediction. If when considering each strategy individually the best results are obtained for the sixth strategy based on the weighted arc product frequency. The highest score is given to the candidates that consist of the most frequent arcs with less pronunciation variability, or in other words, those letter arcs form which less different phoneme arcs were found. The lowest scoring strategy is the one based on the number of same symbols shared by the candidates multiplied by arc frequencies. More results can be found in Table 1

4.3. Phoneme-to-phoneme nativization by analogy(ph2ph_nat)

It makes a lot of sense to perform grapheme-to-phoneme nativization, in fact, most of the Spanish listeners are only familiar with the orthographic form of English words; however, if there is a phonetic transcription available in the source language, finding automatic correspondences between source and target (nativized) phonemes is a more consistent task than in the case of letters, being g2p conversion already a difficult task for

Table 1: Single strategy results for g2p_nat and best strategy combination

strategy mask	ph. acc.	word. acc.
10000000000	84.49	43.81
01000000000	81.76	37.14
00100000000	83.25	37.14
00010000000	83.07	38.10
00001000000	82.53	36.19
00000100000	84.32	44.76
00000010000	82.04	38.10
00000001000	83.65	39.05
00000000100	82.87	40.95
00000000010	82.21	35.24
00000000001	84.32	42.86
10001001011	84.17	43.81

English. An important advantage at this point is the fact that stress markers can be directly copied from the source language pronunciation while length can be inferred using a simple phone mapping table. For phoneme-to-phoneme nativization experiments the PbA was modified in order to receive phoneme input. The best strategy combination (11011000010) as in the g2p_nat case was determined performing n-fold evaluation of all possible strategy combinations. The best n-fold results obtained on the training data were 91.81% for phoneme and 61.29% for words. The results obtained on 100 word test set of common names are 91.61% phonemes and 63.81% words correct. These results show that p2p_nat nativization outperforms g2p_nat nativization by 22% in word accuracy terms. Permorming single strategy experiments for phoneme-to-phoneme nativization we can also observe that the best scoring strategy is the sixth one, while the worst places is tied between the tenth and the second one. The second strategy prioritized those candidates with minimum arc length standard deviation. For more results see Table 2.

Table 2: Single strategy results for p2p_nat and best strategy combination

strategy mask	ph. acc.	word. acc.
10000000000	90.91	63.81
01000000000	89.34	57.14
00100000000	89.69	59.05
00010000000	89.69	59.05
00001000000	90.03	62.86
00000100000	91.08	64.76
00000010000	90.38	61.90
00000001000	90.03	60.95
00000000100	89.34	57.14
00000000010	89.69	59.05
00000000001	90.91	63.81
11011000010	91.61	63.81

4.4. Discussion

When it comes to such a specific task as nativization an objective evaluation is insufficient to determine the validity of the results. Test results obtained with PbA using grapheme-to-phoneme, phoneme-to-phoneme nativization were compared and exaustively evaluated by the authors. Three types of errors were determined. *Severe errors:* the word is either unrecognizable and/or can be confused with another one. *Medium errors*: vowel confusion cases between such vowels as (a/e) (e/i), (o/a). Vowel insertions and deletions together with similar consonant confusions (k/G, t/d, etc.) that do not affect the intelligibility of the words were considered to be *Light errors*. The results obtained using grapheme-to-phoneme conversion by analogy on a test corpus of 100 common names give us 22 "severe" errors affecting the intelligibility while for for the same test corpus but using phoneme-to-phoneme conversion by analogy only 10 "severe" errors were found. An example of "severe error" is the pronunciation of the word *agency* nativized to /a G e n s a j/ or *general* to /D j n e r a l/. We consider "medium" the following nativization error for the word *agency* /e j tS u n s i/. And a "light error" would be pronouncing the word *ball* with an /a/ as in /b a l/. In Spain there is a tendency to pronounce well known frequently used words according to the British pronunciation rules; the word *beautiful* nativized to /b j u D i f u l/ is another example of a "light error". Our experiments were carried out using the pronunciation of isolated words, making no pronunciation adjustments at word boundaries at this point.

5. Conclusions

In this paper we proposed to use pronunciation by analogy for the nativization of English words in Spanish language. The best results were achieved using phoneme-to-phoneme nativization based on the analogy in the phoneme context. The nativization results obtained using analogy only in the letter context were rather poor, due to deep orthography of English language, even though we believe that the analogy-based nativization results are better than g2p results could have been for the same minimalistic training corpus containing only 1000 words. It is worth mentioning that even in the case of grapheme-to-phoneme nativization the results show very significant improvements in comparison to those obtained by direct phoneme-to-phoneme table-based mapping. Nativized pronunciations are more tolerant to the vowel and consonant substitutions, previously referred to as *light errors*. Even though the test corpus that consisted of 100 hundred frequently used common English names can be considered somewhat tiny, for both g2p_nat and p2p_nat methods n-fold evaluation was performed on the training coprus of 1000 rather unfrequent common English names (selected by the greedy corpus balancing tool) and the results obtained were quite similar to those obtained on the test data. The results are statisticaly significant at the level $p = 0.05$. There is no gold standard for nativization, and some exceptions coming from even sometimes incorrect but very frequent adapted pronunciations make their contribution to the difficulty of the problem, however these exceptions are created by humans and obey the analogy both in letter and phoneme contexts. Simple mapping rules were proven to be insufficient for the task.

6. Acknowledgements

This work has been partially funded by the Spanish Government under grant TEC2009-14094-C04-01 (BUCEADOR project)

7. References

[1] A. Llitjós and A. Black, "Knowledge of language origin improves pronunciation accuracy of proper names," vol. Proceedings of Eurospeech. Citeseer, 2001, pp. 1919–2002.

[2] I. Trancoso, "Issues in the pronunciation of proper names: the

experience of the Onomastica project," in *Project. Proceedings Workshop on Integration of Language and Speech*, 1995.

[3] I. Trancoso, C. Viana, I. Mascarenhas, and C. Teixeira, "On deriving rules for nativised pronunciation in navigation queries." Citeseer, 1999.

[4] Y. Marchand and R. Damper, "A multistrategy approach to improving pronunciation by analogy," *Computational Linguistics*, vol. 26, no. 2, pp. 195–219, 2000.

[5] T. Polyakova and A. Bonafonte, "New strategies for pronunciation by analogy," in *Proceedings of the 2009 IEEE International Conference on Acoustics, Speech and Signal Processing-Volume 00*. IEEE Computer Society, 2009, pp. 4261–4264.

[6] D. S. L. Llorente J.M, "Libro de estilo de canal sur tv y canal 2 andalucía."

[7] A. van den Bosch and W. Daelemans, "Data-oriented methods for grapheme-to-phoneme conversion," in *Proceedings of the sixth conference on European chapter of the Association for Computational Linguistics*. Morristown, NJ, USA: Association for Computational Linguistics, 1993, pp. 45–53.

[8] [Online]. Available: www.lc-star.org

[9] [Online]. Available: http://www.phon.ucl.ac.uk/home/sampa/

[10] T. Polyakova and A. Bonafonte, "Transcripción fonética en un entorno plurilingue," in *V Jornadas en Tecnología del Habla , VJTH*, 2008.

[11] R. Damper, Y. Marchand, J. Marseters, and A. Bazin, "Aligning Letters and Phonemes for Speech Synthesis," in *Fifth ISCA Workshop on Speech Synthesis*. ISCA, 2004.

[12] R. Damper and J. Eastmond, "Pronouncing text by analogy," in *Proceedings of the 16th conference on Computational linguistics-Volume 1*. Association for Computational Linguistics Morristown, NJ, USA, 1996, pp. 268–273.