interspeech 2008 incorporating SST 08

Automatic Word Stress Marking and Syllabification for Catalan TTS

Sílvia Rustullet, Daniela Braga, João Nogueira, Miguel Sales Dias

Microsoft Language Development Center, Portugal

i-sirust@microsoft.com, i-dbraga@microsoft.com, i-jonogu@microsoft.com, midias@microsoft.com

Abstract

Stress and syllabification are essential attributes for several components in text-to speech (TTS) systems. They are responsible for improving grapheme-to-phoneme conversion rules and for enhancing the synthetic intelligibility, since stress and syllable are key units in prosody prediction. This paper presents three linguistically rule-based automatic algorithms for Catalan text-to-speech conversion: a word stress marker, an orthographic syllabification algorithm and a phonological syllabification algorithm. The systems were implemented and tested. The results gave rise to the following word accuracy rates: 100% for the stress marker algorithm, 99.7% for the orthographic syllabification algorithm and 99.8% for the phonological syllabification algorithm.

Index Terms: Catalan text-to-speech, stress, orthographic and phonologic syllabification, prosody

1. Introduction

Stress and syllabification have major impact in TTS systems: on the one hand, in grapheme-to-phone(me) conversion, and on the other hand in the prosody prediction. Stress and syllabification are part of the phonetic information of lexica used as input in dictionary-based TTS systems. It was proved that stress and syllabification dramatically improves Hidden Markov Models-based Speech Synthesizers (HTS) offline training and synthetic voice naturalness [1]. There is a large tradition of advanced studies on the Phonology of Catalan [2], [3] and extensive reports on linguistic rule-based grapheme-to-phone converters [4], [5]. However, details either about the rules or about the algorithms architecture or even about the system performance results seem to be scarce, especially regarding the stress marking and syllabification in Catalan.

This paper is structured as following: in section 2, the automatic stress marking and syllabification algorithms for Catalan are presented; in section 3, the tests and results are discussed; in section 4, main conclusions are summarized and future work is foreseen.

2. Automatic Word Stress Marker and Syllabification Algorithms

2.1. Methodology

In this paper, we used the linguistic rule-based methodology proposed in [6] for Portuguese, since it was demonstrated that this method has better results than data-driven or statistical approaches. Like Portuguese, Catalan has a phonologically-based orthography. The rules were driven using a 1000 words lexicon and were tested with two types of corpora: a lexicon using a different set of 1000 words and a newspaper text, composed by 223 words. The algorithms were not designed to deal neither with foreign words, in spite of their phonetic adaptation to Catalan, nor with abbreviations, titles and acronyms, since the graphical patterns of these items are not consistent with the Catalan orthography and phonology. Compound words were also disregarded from the corpus.

2.2. Architecture and symbols definition

In Figure 1, the pipeline of the automatic stress and syllabification algorithms for Catalan is shown.

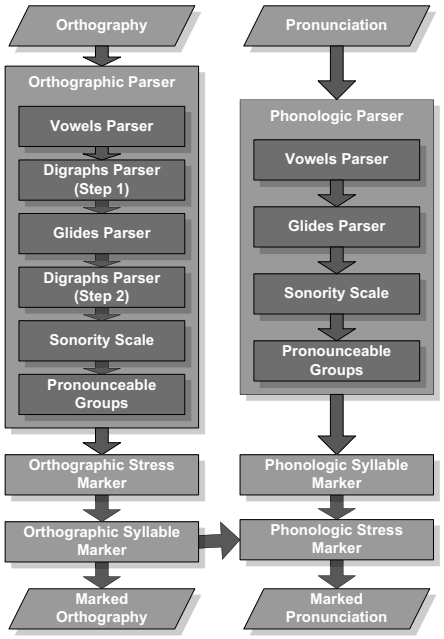

Figure 1: *Catalan automatic stress and syllabification algorithms architecture.*

The input of the orthographic-based stress and syllabification algorithm are ASCII characters (left column in Figure 1), whereas the input of the phonologic-based stress and syllabification algorithm is the phonologic alphabet (right column in Figure 1). Then, a parser module is activated with the purpose of tagging vowels, digraphs (only in the orthographic stress and syllable marker), glides, the level of sonority scale and the pronounceable groups. In the end, two outputs are expected: an orthographic output with stress and syllable boundaries and a phonologic one, only with syllable boundaries. There is a matching between orthographic and phonologic syllabification and the phonologic stress is extrapolated from the orthographic marker. Not only the stress marker but also both orthographic and phonologic syllabification algorithms work independently.

The following symbols were used in the Catalan stress and syllabification algorithms: **V**: phonological vowel; **VG**: vocalic grapheme; **D**: digraph; **G**: glide; **C:** consonant (includes digraphs); **#:** end of word; **SC(n)**: sonority scale

number n; **PG**: pronounceable group; **^(1):** word last grapheme; **^(2):** word penultimate grapheme; **^(n):** any position of a grapheme in the word; **+(1):** word last vowel; **+(2):** word penultimate vowel; **+(n):** any position of a vowel in the word; **T=+( ):** position of the stressed vowel; **T=+(1):** stressed vowel is the last vowel; **/:** except; **→:** then; **+(L)**: vowel on the left of a non-vowel group; **+(R):** vowel on the right of a non-vowel group; **$(1):** last non-vowel in the group of non-vowels; **$(2):** penultimate non-vowel in the group of non-vowels; **$(n):** non-vowel in any place in the group between 2 vowels; **$(n)=n:** distance of non-vowels between vowels; **-**: syllable boundary.

2.3. Orthographic parser

The pre-processing module of the orthographic stress and syllabification algorithm is the orthographic parser, whose goal is to classify graphemes into vowels, digraphs and glides. The first stage of this module is the tagging of the vocalic graphemes (VG: <a,e,i,o,u,à,é,è,í,ï,ò,ó,ú,ü>). A group of vocalic graphemes is a set of adjacent graphemes in which all elements belong to VG's category. Then, the system identifies the digraphs containing VGs using the rules shown in Table 1.

Table 1. *Rules for digraphs parser (step 1).*

#	Rule	Example
1	…<qu, gu> <ix>… → <qu, gu> = D	guix, quixot
2	…VG + <ix>… → <ix> = D	reixa, faixa
3	…VG + <ig> # → <ig> = D	raig, faig
4	…<qu, gu> < e,è,é,i,í >… → <qu, gu> = D	quincalla, guerra

The following step is to identify the glides among the VG's set. Only the graphemes <i,u,ü> will be considered glides. The identification of the glides is absolutely necessary to the orthographic syllabification algorithm. This problem does not occur in the phonologic syllabification, as there are different phonologic transcriptions to represent glides and vowels. Some normative full vowels have been considered glides to account for oral speech [7], [8]. The glides' parser will execute the following rules, depending on the number of successive VGs:

1. If there is a 5 vowels group, only first two VGs are considered; the remaining 3 will be considered in the next iteration (e.g. < rad**iooie**nt>);

2. If there is a 4 vowels group
2.1 and if the forth is <ï,ü> → the first three VGs will be processed by the next iteration (e.g. < an**iui**ïta>);
2.2. otherwise, only consider the first two VGs ; the last two will be processed by the next iteration (e.g. <o**ua**ire>);

3. If there is a 3 vowels group
3.1. and the second one is <i, u> → <i, u> are glides (e.g. <que**i**a, je**i**a>);
3.2. and the first and the third are <i, u, ü>
3.2.1. and if the word starts by an one of the prefixes listed in the exception prefix list 1 (<bio->) that includes the first VG → none of them is a glide (<b**ioi**ntel·ligència>);
3.2.2. and if the group is preceded by <g,q> → the first and the third are glides (<g**ua**itar>)
3.2.3. otherwise the third VG is a glide (<atia**i**re>);
3.3. and if the third VG is <i, u, ü> → it is a glide (e.g. < centree**u**ropeu>);
3.4. otherwise none of the three is a glide (e.g. <g**eoe**lement>);

4. If there is a 2 vowels group
4.1. and if the first one is <ü> → the first one is a glide (e.g. < paraig**ü**es>);
4.2. and if the second vowel is <ú,í, ï,ü > → none is a glide (e.g. <pa**ï**sos>);
4.3. and if the group is <iu>
4.3.1. and if the glides are in the beginning of the word → the first is a glide (e.g. <**i**ugoslau>);
4.3.2 and if the vowels are followed by <-m> ending the word, the first is a glide (e.g. <solàr**i**um>);
4.3.3. and if <iu> belongs to one of the prefixes in the exception prefixes list 2 (<poliu-, biun-, trium->)→ none is a glide (e.g. <pol**iu**retà>);
4.3.4. otherwise the second vowel is a glide (e.g. (e.g. <r**iu**re>);
4.4. and if the group is <ui>
4.4.1. and if it is followed by one of the exception suffixes 1 (<-r, -r-se, -sme, -sta>)→ none is a glide (e.g. <prod**ui**r-se>);
4.4.2. otherwise the second vowel is a glide (e.g. <cu**i**na>);
4.5. and if the second vowel is <i, u>
4.5.1. and if the two vowels are followed by one of the exception suffixes 2 (<-sme, -sta, -r, -r-se, -t, -nt, -ré, -ràs, -rà, -rem, -reu, -ran, -ria, -ries, -ríem, -ríeu, -rien>)→ none is a glide (e.g. <eg**oi**sme>);
4.5.2. and if the two VGs are preceded by one of the exception prefixes 3 (<pre-, re-, sobre-, anglo-, franco-, auto->) → none is a glide (e.g.<pr**eu**niversitari>);
4.5.3. otherwise, the second VG is a glide (e.g. <mo**u**re>);
4.6. and if the first VG is <u> and it is preceded by <k, g, q> → the first vowel is a glide (e.g. <g**u**apo>);
4.7. and if the first VG is <i, u>
4.7.1. and if the first VG is a word's first grapheme → the first vowel is a glide (e.g.<**i**ode>);
4.7.2. and if the second VG carries a graphical accent and is at the end of the word
4.7.2.1. and if there are more VGs in the word → the first VG is a glide (e.g. <ital**i**à>);
4.7.2.2. otherwise none is a glide (e.g. <N**i**àgara>);
4.7.3. and if there is a previous vowel that carries a graphical accent → the first VG is a glide (e.g. <eminènc**i**a>);
4.7.4. otherwise none is a glide (e.g. <r**i**erol>);
4.8. otherwise none is a glide (e.g. <**ae**ronàutica>).

Having the glides and the vowels clearly tagged, the system runs the rules presented in Table 2, in order to identify the remaining digraphs which do not contain VGs and which cannot run with the digraphs listed in Table 1, because they need a phonological vowel preceding them.

Table 2. *Rules for digraphs parser (step 2).*

#	Rule	Example
1	…<-rr-, -ss-, -l•l-, -ll-, -ny-, -sc-,wh>… = D	carro, bossa, estany
2	…<ch>…= D	Antich
3	…V + <ng, nc> #→ <ng, nc> =D	fong, banc
4	…VG + <tj, tg, gg, dj, tx, >…→ <tj, tg, gg, dj, tx> = D	sutge, adjectiu, metxa
5	…V + <ng> + C\ <r,l>…→ <ng> = D	sangtraït

Other important elements to the syllabification rules are the sonority scale (SC) and the pronounceable groups (PG). The sonority scale is a ranking of speech sounds by resonance or amplitude [8]. The sonority scale for Catalan has 6 levels and presents the following graphic patterns: SC(6) = V; SC(5) = G; SC(4) = C (<r,l>) and D (<rr,ll, l•l>); SC(3) = C (<m,n>) and D ><ny, nc>); SC(2) = C(<f,s,z,ç,x,j>), D (<ss>) and C(<g>) + V <e,è,é,i,í>; SC(1) = C (<p,b,v,t,d,c, g, q>) and D (<qu, gu>). Pronounceable groups [9] are those consonant clusters which are allowed in an onset context and fulfill the requirements of the sonority scale. The pronounceable groups in Catalan are the following graphic sequences: <pr,pl,br,bl,vl,tr, dr, cl,cr,gl,gr,fl,fr>.

2.4. Automatic stress marker algorithm

The proposed word stress marker for Catalan is composed by 8 rules (cf. Table 3), and is based on the analysis of context around the vowels of each word. After the text is separated into sentences and the sentences are split into words, the system starts by verifying whether there is a non-stressed word in the sentence. According to literature [10], non stressed words are monosyllabic high frequent function words which receive no stress mark (e.g. <el, la, els, les, na, un, uns, a, amb, de, en, per, al, del, pel, als, dels, pels, mon, ma, mos, mes, i, ni, si, que, me, nos, te, vos, se, lo, la, li, los, les, ho, hi, ne, em, ens, et, es, el, els>). If the word is not in the non-stressed list, the system runs the orthographic parser described in section 2.2. This parser identifies the position of the VG (vocalic grapheme) in the word, allowing the analysis of the graphical context around them. Adverbs in <–ment> are considered to have only one stress mark [11]. The suffix <–ment> will carry that stress.

Table 3. *Rules for Catalan stress marker.*

#	Rule	Exemple
1	List of non-stressed words	el, la, els, les, na
2	If the word has only one vowel → T= V	pols, lloc
3	If …<–ment># →T = +(1)	càndidament, gràcilment
4	If there is an orthographic accent <´>, <`>, the accented VG is the stressed one	història, política
5	If ^(1) = V → T = +(2)	casa, arbre
6	If ^(1) = <s> and ^(1) = V → T = +(2)	carros, botes
7	If ^(1) = <n> and ^(1) = V(<e, i>) → T = +(2)	oxigen, vermin
8	If none of the above rules occur → T = +(1)	cavall, rebost

2.5. Orthographic syllabification

Both orthographic and phonological syllabification algorithms follow a right to left approach. Since any syllable requires a vowel and since onsets are filled before codas [12], which may not be present, the algorithm takes for granted the last coda (if any) and looks for the beginning of the onset according to the sonority requirements and the phonotactics of Catalan. In Table 4, the set of rules for the Catalan orthographic syllabification is presented.

Table 4. *Rules for Catalan orthographic syllabification.*

#	Rule	Example
1	If $(n)=0 →+(L) – +(R)	ll**u-ir**, alv**e-o**lar
2	If $(n)=1 → +(L) - $(1)	b**o-t**a, cat**i-f**a
3	If $(n)=2 and if $(2) + $(1) = D → +(L) - $(2)	m**a-ll**a, c**a-ny**a
4	If $(n)= 2 and if $(2) = G → $(2) - $(1)	ga**i-r**e, mo**u-r**e
5	If $(n)= 2 and if $(1) = G → +(L) - $(2)	à–**ri**a, và-**lu**a
6	If $(n)= 2 and if $(2) + $(1) = PG → +(L) - $(2)	sa-**br**e, pe-**br**e
7	If rules 3,4,5,6 are false → $(2) - $(1)	mos-**c**a, conde**m-n**a
8	If $(n)=3 and if $(3) + $(2) = D → $(2) - $(1).	co**ll-p**elat
9	If $(n)=3 and if $(2) + $(1) = D → $(3) - $(2).	e**n-ll**oc
10	If $(n)=3 and if $(1) = G → $(3) - $(2).	so**m-ni**ar, calu**m-ni**ar
11	If $(n)=3 and if $(2) and $(1) = SC → $(2) - $(1).	co**mp-t**ar
12	If $(n)=3 and if $(2) = C(<s>) → $(2) - $(1).	Po**ts-d**am
13	If $(n)=3 and if $(2) + $(1) = PG → $(3) - $(2)	po**r-pr**a, a**r-br**e
14	If $(n)=4 and if $(2) = C**(<s>)** → $(2) - $(1).	e**rns-t**ita
15	If $(n)=4 and if $(3) +$(2)= PG and if $(1)= G → $(4) - $(3)	indú**s-tri**a
16	If rules 14,15 are false → $(3) - $(2)	u**ll-bl**au

2.6. Phonological syllabification

Table 5. *Rules for Catalan phonologic syllabification.*

#	Rule	Examples
1	Two consecutive identical sounds → $(n) - $(n-1)	/p O **b – b** l @/
2	If $(n)=0 →+(L) - +(R)	/f @ r **@ – o**/
3	If $(n)=1 → +(L) - $(1)	/@ s p **O – z** @/
4	If $(n)= 2 and if $(2) = G → $(2) - $(1)	/a **j - r** @/
5	If $(n)= 2 and if $(1) = G → +(L) - $(2)	/@ m p l **a - r j** @/
6	If $(n)= 2 and if $(2) + $(1) = PG → +(L) - $(2)	/l **e - p r** @/
7	If 4-6 equals FALSE → $(2) - $(1)	/rr a **m – p** @/
8	If $(n)=3 and if $(1) = G → $(3) - $(2)	/k u m @ **rr - s j** a l/
9	If $(n)=3 and if $(2) = C(/s,z/) → $(2) - $(1)	/s u p @ **rr s - t** i s j o/
10	If $(n)=3 and if $(2) + $(1) = PG → $(3) - $(2)	/p u **l - k r** @/
11	If $(n)=4 and if $(3) = C(/s,z/) → $(3) - $(2)	/m i **k s – t j** o/
12	If $(n)=4 and if $(2) + $(1) = PG → $(3) - $(2)	/i **n s - t r** u k t i w/

For the phonologic syllabification, a similar pipeline is required, bearing in mind that the inputs are phonologic transcriptions [13]. Therefore, vowels and glides' parsers, sonority scale sequences and pronounceable groups are based in a list of phonologic symbols (V: /a, @, E, e, i, O, o, u/; G: /j,w/; SC(6) = V; SC(5) = G; SC(4) = C(/r,R,l,L,B,D,G/); SC(3) = C(/m,n,J,N/); SC(2) = C(f,s,z,S,Z/), SC(1) = C (/p,b,t,d,k,g,tS,dZ/); PG : /pr,pl,br,bl,kl,kr,gl,gr,tr,dr,fl,fr/). Catalan SAMPA was used in the phonologic transcriptions, and is available in [14]. In Table 5, the rules for the Catalan phonologic syllabification are shown.

3. Results and Discussion

Two tests were carried out in order to assess the performance of the proposed algorithms: test 1 used 1000 words randomly selected from a Catalan phonologically transcribed lexicon and test 2 used a 223 words newspaper text. This lexicon does not include foreign words, abbreviations, or acronyms. The purpose of the second test was to use the algorithms with real Catalan texts. The phonologic syllable marker was not tested with the second corpus, because no phonologic transcriptions were available. Finally, we compared the outputs between both orthographic and phonologic syllable markers and analyzed the mismatches. Table 6 shows the results of the conducted tests. The word accuracy rate in test 1 was 100% for the stress marker, 99.7% for the orthographic syllable marker and 99.8% for the phonologic stress marker. Automatic stress marker failed in test 2 because ambiguations between monosyllabic weak words and monosyllabic stressed words were not considered. Regarding the orthographic syllable marker, the errors occurred in 2 compound words where the hyphen should be considered as a syllable boundary and one word where a digraph was not recognized.

Table 6. *Automatic stress marking and syllabification results for Catalan.*

Orthographic stress marker		
Error type	Test 1 (WER %)	Test 2 (WER %)
ambiguation	0.0	0.9
Total	**0.0**	**0.9**
Orthographic syllable marker		
Error type	Test 1 (WER %)	Test 2 (WER %)
hyphen	0.2	0.0
digraph	0.1	0.0
Total	**0.3**	**0.0**
Phonologic syllable marker		
Error type	Test 1 (WER %)	
PG + G	0.2	
Total	**0.2**	
Mismatch between orthographical and phonological syllable marker		
Error type	Test 1 (WER %)	
non recognized hiatus	2.9	
non recognized glide	1.6	
non recognized glide	1.7	
exception	1.1	
Total	**6.3**	

Regarding the phonologic syllable marker, the algorithm did not recognize the pattern pronounceable group followed by glide. This mistake could be avoided including rule 13: If $(n)= 3, if $(3) and $(2) = PG and if $(1) = G → $(L) - $(3). Since the phonological stress was extrapolated from the orthographical stress and to make that extrapolation the same number of syllables was required some mismatches arise. They were mainly caused because of the glide algorithm which did not recognize some medial glides belonging to a raising diphthong. Other possible mistakes were some possible mistakes involving glides in the prefix-root boundary: a hiatus had to be maintained. Exception also referred to a hiatus maintenance although its position was advised to be glide. Possible solutions to those mismatches would be enlarging the prefixes list, considering VG{i} glides when followed by V and r, t, or l in final position or when preceded by a sibilant sound in medial position. These results are similar to the WERs reported in [6] for Portuguese automatic stress marking and syllabification.

4. Conclusions

In this paper, an orthographic and phonologic stress and syllable automatic markers for Catalan TTS were presented. The successful results of nearly 100% of word accuracy rates for each algorithm seem to prove that stress and syllabification can be rule-driven in languages which have a phonologic-based orthography. Future work will address the stress and syllabification assignment of foreign and compound words. A comparison between linguistic rule-based and stochastic methods is envisaged.

5. Acknowledgements

We would like to thank Professors Teresa Cabré and Pilar Prieto, from the Universitat Autònoma de Barcelona, for sending us important references concerning the Phonology of Catalan.

6. References

[1] Maia, R., Speech Synthesis and Phonetic Vocoding for Brazilian Portuguese Based on Parameter Generation form Hidden Markov Models. PhD thesis. Department of Computer Science and Engineering, Nagoya Institute of Technology, Nagoya, Japan, 2006.

[2] Cabré, T. and Prieto, P., *Diftongs creixents versus hiats situació del català dins la Romània*, in : http://seneca.uab.es/ggt/reports.htm (04/04/2008)

[3] Wheeler, M., The Phonology of Catalan. Blackwell, Oxford, 2005.

[4] Llisterri, J., Machuca, M,- Madrigal, N., Mancini, F, Massimino, P., Mota, C., Riera, M., Ríos, A. "Aspectos lingüísticos en el diseño de un conversor de texto en habla en castellano y en catalán: El sistema LoquendoTTS®", VI Congreso de Lingüística General. Universidad de Santiago de Compostela, Santiago: 521-522, 2004.

[5] Pachès, P., Riera, M., Perea, P., Febrer, A., Estruch, M., Garrido, J.M., Machuca, Ríos, A., Llisterri, J., Esquera, I., Hernando, J., Padrell, J., Nadeu, C. "SEGRE: An Automatic Tool for Grapheme-to-Allophone Transcription in Catalan", Proceedings of the Workshop on Developing Language Resources for Minority Languages: Reusability and Strategic Priorities (LREC-2000 Second International Conference on Language Resources and Evaluation): 52-61, 2000.

[6] Braga, D., Algoritmos de Processamento da Linguagem Natural para Sistemas de Conversão Texto-Fala em Português. PhD Thesis. Universidade da Coruña, España, 2008.

[7] Cabré, T. and Prieto, P. ''Prosodic and analogical effects in lexical glide formation in Catalan'' Probus, 16 : 113-150, 2004.

[8] Bonet, E. and Lloret, R., Fonologia catalana, Barcelona, Ariel, 1998.

[9] Institut d'Estudis Catalans. Gramàtica de la llengua catalana, in http://www.iecat.net/institucio/seccions/Filologica/gramatica/ (04/04/2008)

[10] Bruguera, J., Diccionari ortogràfic i de pronúncia, Barcelona, Enciclopèdia Catalana, 1990.

[11] Conejero, D. and Moreno, A., LC-STAR CATALAN as spoken in Spain, Universitat Politècnica de Catalunya Publications, Barcelona, 2005.

[12] Julià-Muné, J., Fonètica aplicada catalana. Ariel, Barcelona, 2005.

[13] Institut d'Estudis Catalans, Aplicació al català dels principis de transcripció de l'Associació Fonètica Internacional. Edició a cura de Joaquim Rafel i Fontanals. Institut d'Estudis Catalans, Barcelona, 1999.

[14] Llisterri, J. *A proposal for Catalan SAMPA*, in http://liceu.uab.es/~joaquim/language_resources/SAMPA_Catalan.html (23/06/08).