INTERSPEECH 2009 BRIGHTON

Automatic Syllabification for Danish Text-to-Speech Systems

Jeppe Beck [1], *Daniela Braga*[1], *João Nogueira*[2], *Miguel Sales Dias*[1], *Luis Coelho*[3]

[1] Microsoft Language Development Center, Portugal
[2] Faculty of Sciences of University of Lisbon, Portugal
[3] Polytechnic Institute of Oporto, Portugal

{v-jeppeb, i-dbraga, Miguel.Dias}@microsoft.com, jnogueira@lasige.di.fc.ul.pt, lcoelho@eu.ipp.pt

10.21437/Interspeech.2009-403

Abstract

In this paper, a rule-based automatic syllabifier for Danish is described using the Maximal Onset Principle. Prior success rates of rule-based methods applied to Portuguese and Catalan syllabification modules were on the basis of this work. The system was implemented and tested using a very small set of rules. The results gave rise to 96.9% and 98.7% of word accuracy rate, contrary to our initial expectations, being Danish a language with a complex syllabic structure and thus difficult to be rule-driven. Comparison with data-driven syllabification system using artificial neural networks showed a higher accuracy rate of the former system.

Index Terms: automatic syllabification, rule-based techniques, artificial neural networks, text-to-speech

1. Introduction

As Ladefoged (2001) [1] states, "Although nearly everybody can identify syllables, almost nobody can define them". The definition of syllable unit varies widely, as well as the agreement towards the number of syllables of a word. This concept seems also to be language related, since some languages are claimed to be syllabically less complex, as Italian, Finish and Spanish, whereas English has a more complex syllabic structure [2]. The syllable information (its phonological structure, distribution, boundaries, number of syllables of a word) plays an important role in speech technology, both in Text-to-Speech (TTS) [3] generation and in Speech Recognition (SR) modeling. The syllable information plays an important role in the following levels of speech technology: a) syllable boundaries are part of the phonetic information of lexicons used as input in dictionary-based TTS and SR language models and grammars; b) syllable information is a parameter used in HTS training, namely the number of syllables of a word and its correlation with stress improves Hidden Markov Models-based Speech Synthesizers (HTS) offline training and synthetic voice naturalness [4]; c) syllable information is important to feed data-driven methods of prosody generation for TTS systems [5]; d) the pronunciation of a given phoneme may rely on its localization in the syllable structure; e) duration models in TTS and SR systems are enhanced when using syllable information and its organization [6]. The automatic prediction of syllabification is a very useful tool for everyone who develops speech technology in a given language. However, there are two main challenges associated: the concept of syllable as a unit is not consensual and, in addition, syllabification process is language dependant, which requires specialized linguistic knowledge. There are numerous descriptions of automatic syllabification systems for several languages (English, Portuguese), but they seem to be scarce when we address to Danish automatic syllabification. In order to tackle automatic syllabification, there are two common approaches: rule-based techniques based on the phonologic regularities and restrictions of the languages [6], [7], [8] and data-driven methods [9], among which there are different techniques, such as look-up procedure using N-grams, exemplar-based generalization technique, and syllabification by analogy [10]. Finite State Transducers are another technique which allows the development of rule-based and data-driven syllabification systems [11]. Comparisons between these two approaches showed that rule-based approaches seem to be more effective for certain languages like Portuguese [8], [12], [6] whereas data-driven methods showed better performance for English and Italian [2], [9]. However, in our opinion, the success rate of either approach may not be explained by the complexity of the syllabic structure of the language, since Portuguese and Italian are possibly similar in syllabic structure. The main reason may rely on the amount of available linguistic resources for each language, mainly syllable-tagged dictionaries, which explains the investment in data-driven techniques that require less skilled linguistic expertise than rule-based systems. Although Danish belongs to the same language family of English, it has much less available linguistic resources, such as phonetic dictionaries, which makes data-driven approaches difficult to implement. There is a large tradition of advanced studies on the Phonology of Danish [13], [14] and reports on linguistic rule-based grapheme-to-phone converters [15]. However, details either about the rules or about the algorithms' architecture or even about the system's performance results seem to be scarce, especially regarding the automatic syllabification in Danish using a rule-based approach.

This paper is structured as following: section 2 presents the methodology followed in this work; in section 3, the automatic syllabification algorithms for Danish is presented; in section 4 and 5, the tests and results are shown and discussed; in section 6 we present a data-driven approach for Danish automatic syllabification using artificial neural networks (ANN) and its results using the same test set; in section 7, main conclusions are summarized and future work is foreseen.

2. Methodology

In this paper, we used the linguistic rule-based methodology proposed in [8] for Portuguese, and applied in [16] for Catalan, since it was demonstrated that this method has better results than data-driven approaches. The rules were driven using a 1000 words lexicon and were tested on different corpora: a lexicon using a different set of 1000 words, and on running text containing 1000 words from a Danish newspaper. The algorithm was not designed to deal neither with foreign words, in spite of their phonetic adaptation to Danish, nor with

Copyright © 2009 ISCA

abbreviations, titles and acronyms, since the graphical patterns of these items are not consistent with the Danish orthography and phonology. Furthermore, the algorithm was not specifically designed for handling compound words, but these were included in the test material, since compounding is a highly productive phenomenon in Danish.

3. Automatic Syllabification Algorithm

3.1. System Architecture

Figure 1 shows the pipeline for the automatic syllabification algorithm for Danish. Syllabification of foreign words, titles, abbreviations and acronyms was not considered at this point of our system development, since these words do not follow the standard syllabic structure of Danish. The input for the syllabification algorithm is a phonetic representation rendered in ASCII characters from a 100k Danish lexicon without syllable boundary information. The parser module is activated and it tags (in order) full vowels, valid three, two and one consonant clusters, weak vowels and weak vowel onset groups. Syllable boundaries are introduced according to the specifications of the algorithm. Expected output is an identical phonetic representation as input but including syllable boundary markers.

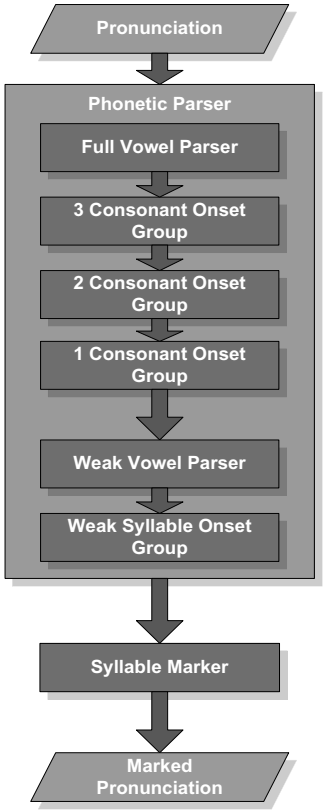


Figure 1: *Automatic syllabification algorithm architecture.*

3.2. Symbols Definition

The following symbols were used for the Danish syllabification algorithm: V_{weak}: [ə ɐ]; V_{full}: remaining vowel sounds, excluding [ə ɐ]; O_1: valid 1 consonant onset group; O_2: valid 2 consonant onset group; O_3: valid 3 consonant onset group; (L): leftmost vowel; (R): rightmost vowel; $(n): number of consonants between leftmost and rightmost vowel sound; $(1): first consonant sound to the left of the rightmost vowel sound; $(2): second consonant sound to the left of the rightmost vowel sound; $(3): third consonant sound to the left of the rightmost vowel sound; -: syllable marker; #: word boundary marker. The operational definition of valid syllable onsets and codas will be: consonant clusters which are possible in word initial and final position. Further details on the definition of valid onsets and codas can be found in [20].

3.3. Rules and Algorithm

The syllabification algorithm reads input from right-to-left. Since any syllable requires a vowel and since onset is filled up before coda, which may or may not be present, the algorithm takes the rightmost coda for granted and looks for the beginning of the onset according to the Maximal Onset Principle [17], [18], [19], modified so as to not create invalid consonant clusters in onset or coda [19], [20]. However, weak syllables deviate from this principle and must be treated separately [21]. The algorithm presented here does not check the validity of consonant clusters in coda. Valid consonant clusters constitute a finite set of combinations based on the phonotactics of Danish [14]. However, since the input for this algorithm is a broad phonetic representation rather than a phonological representation a mapping of consonant clusters is required (O_3: [sbl sbr sbj sdr sdj sgl sgʁ sgj sgv]; O_2:[sb sd sg sf sv sm sn sl pn pl pʁ pj tʁ tj kn kl kʁ kj bl bʁ bj gn gl gʁ gj fn fl fʁ fj kv dv dʁ dj vʁ] ; O_1: [p t k b d g s f ɕ m n ʁ l h v j]; V_{full}: [i iː ɪ e eː ɛ ɛː a æː ɑ ɑː y yː ø øː œ œː u uː ʊ o oː ɔ ɔː ʌ ɒː]; V_{weak}: [ə ɐ]). Broad IPA transcriptions are used throughout this paper (cf. [13] for details). Applying phonologic syllable boundaries on phonetic input makes the algorithm more robust when dealing with syllabification compounds in the lexicon since consonant gradation and several other phonological processes (for details, cf. [21]) that can be based on the phonological syllable are already present in the phonetic representation. Tables 1 and 2 present the rules for automatic syllabification of Danish.

Table 1. *Syllabification rules for full vowels.*

#	Rules	Example
0	Two consecutive identical sounds → $(n) - $(n-1)	[ˈba**l-l**ɑjˀ] [ˈkɑm**b-b**oɐ̯ˀ]
1	IF $(n) ≥3 and $(3) + $(2) + $(1) = O_3 → - $(3) $(2) $(1) +(R) ELSE go to rule 2	[ˈan**-sgʁi**:ˀ] [ˈɑfd-ən**-sdja**ɐ̯n-ə]
2	IF $(n) ≥3 and $(2) + $(1) = O_2 → $(3) - $(2) $(1) +(R) ELSE go to rule 3	[ˈɛ**n-dʁɛ**ŋ] [e**n-fla-**ɛˈoːˀn]
3	IF $(n) ≥3 and $(1) = O_1 → $(3) $(2) - $(1) +(R)	[ˈfʁɑ**nsg-m**anˀ] [ˈfɛ**sd-li**]
4	IF $(n)=2 and $(2) + $(1) = O_2 → +(L) - $(2) $(1) +(R) ELSE go to rule 5	[fʌ**-**ˈ**dʁɔ**g-ən] [**i-**ˈ**bla**nˀd]
5	IF $(n)=2 and $(1) = O_1 → +(L) $(2) - $(1) +(R)	[ˈ**tɑŋ-go**] [ˈ**hɛl-ga**]
6	IF $(n)=1 and $(1) = O_1 → +(L) - $(1) +(R) ELSE go to rule 7	[ˈn**o**ː**-ta**] [ˈ**ko**ː**-da**]
7	IF $(n)=1 and $(1) = ¬$O_1$ → +(L) $(1) - +(R)	[p**iɐ̯-u-**ˈɛd] [ˈm**oð-a**n-gʁæːʔb]
8	IF $(n)=0 → +(L) - +(R)	[i-ni-**ti-a-**ˈtiwˀ]

		[n**y**-ˈ**ɑ**ŋs-ə]

Table 2. *Syllabification rules for weak vowels.*

#	Rules	Example
9	IF +(R) = V_{weak} and $(1)= ʁ → - $(1) +(R) ELSE - +(R)	[ˈno:**ð-ə**] [ˈko:**ð-ə**]

4. Tests and Results

One test was carried out in order to assess the performance of the proposed algorithm. Errors are defined as word forms that contain invalid consonant onsets or codas according to the definition above. Test 1 corpus set consisted of 1000 randomly selected word forms from a phonetically annotated lexicon of 100k words in Danish, excluding abbreviations, acronyms. We started by eliminating the syllable boundary tags out of the phonetic transcription of the words. However, since compounds are not marked as such in the lexicon, these were included in the random selection. Table 3 below presents the results for test 1. Overall accuracy of the syllable breaker was 96.9%. Compounds were responsible for 0.9% plus 1.2% of the errors and the syllabification rule for weak syllables was responsible for 1.0%. Test 2 corpus set consisted of 1000 words (604 unique words) from running text extracted from a Danish newspaper article. This second group of words was firstly phonetically transcribed in order to serve as input of the rule-based syllabifier. Both the results of Test 1 and Test 2 were manually checked by a Danish Linguist. Table 4 below presents the results for test 2. Overall accuracy of the syllable breaker was 98.7%. Compounds were responsible for 0.9% of the errors and the syllabification rule for weak syllables was responsible for 0.3%.

Table 3. *Results in % of the rule-based syllabifier using Test 1 corpus set (1k words randomly selected from a phonetic lexicon).*

Error type	Test 1 (WER%)
compound	0.9
compound with epenthetic “s”	1.2
weak syllable	1.0
total	3.1

Table 4. *Results in % of the rule-based syllabifier method using Test 2 corpus set (1k words selected from a newspaper article).*

Error type	Test 2 (WER%)
compound	0.1
compound with epenthetic “s”	0.9
weak syllable	0.3
total	1.3

5. Discussion

As expected, the Maximal Onset Principle conflicts with the general rule that a word boundary always is a syllable boundary as well [13]. This applies to Danish compounds without epenthetic “s” or “e”. The impact of this in the test material is fairly small since only 12 compounds formed valid onsets across word boundaries. An example is [ˈva-nɛ:-ʁ] vanære (translation: dishonour). Here the algorithm breaks to the left of the consonant since [n] is a valid 1 consonant onset. However, the word boundary is on the right of the consonant van#ære, and this is the syllable boundary as well. In compounds with epenthetic “s”, the algorithm breaks to the left if the epenthetic “s” plus following consonants form a valid onset, and to the right if not. For instance, in the compound notesbog [ˈno:d-ə-sbɔwʔ] (translation: *note book*) epethentic “s” is syllabified to the right because [sb] is a valid onset, whereas in the compound verdenshandel [ˈvaɐ̯d-əns-hanʔ-əl] (translation: world trade) it is syllabified to the left since [sh] is not a valid onset. The only satisfactory resolution to these irregular syllabification results for compounds is to parse input text through a compound word breaker inserting word boundaries prior to syllabification.The simple algorithm for syllabification of weak vowels presents some problems in $V_{full}C_1C_2V_{weak}$ structures, since the rule states that unless C_2 = [ʁ] syllabification occurs directly to the left of the weak vowel, resulting in an invalid coda in the previous syllable e.g. yngler [ˈøŋl-ɐ] (translation: breed (vb.)). However, parsing syllable codas in a phonetic transcription is time consuming and, in this case, yields limited results in comparison to the effort involved, and this is the only satisfactory solution to resolve this irregularity. Nevertheless, it is possible to expand the rules for syllabification of weak vowels according to a set of ordered principles as shown in [13], [20]. Another approach could be to treat weak vowels exactly like full vowels, thus reducing the number of invalid codas in $V_{full}C_1C_2V_{weak}$ structures. However, the approach outlined here has the distinct advantage that schwa assimilation becomes highly predictable since schwa will never assimilate to a sonorant in syllable onset [22].

6. Automatic Danish syllabification using artificial neural networks

Since we could not find results for automatic Danish syllabification problem using other approaches we decided to develop a data driven system in order to compare the rule-based system described in section 3. Our system (initially developed and tuned for European Portuguese but easily adapted) consists of an artificial neural network (ANN) with a 4 layer feed forward structure inspired in [23] but with extended class tag information. The input receives a phoneme within a k length context ($2k$+1 phonemes) and the ANN outputs a probability value for the onset, nucleus and coda (ONC) classes. Our phoneme inventory, with length n, was used to create a n-dimensional orthogonal continuous space that provides a suitable representation base for an appropriate network training. In this space each phoneme is represented by a vector where all the components are 0 except the one related with his own axis which will be 1. The ONC output classes were considered since we are using a phonetic base; however they do not provide any position information inside de syllable. To overcome this limitation we added to each tag an extra symbol resulting in Ox, Nx and Cx with 1<x<4. This gives a total of 12 neurons on then ANN output layer. Network training was made off-line using a steepest descent criteria and a standard back-propagation algorithm. Each training example was composed of 1 phoneme with his related neighborhood (we used a 2 phoneme context) and the ONC classification with position information. The output is normalized according to a maximum entropy related criteria:

$$p_i = \frac{e^{q_i}}{\sum_{j=1}^{N} e^{q_j}} \quad (1)$$

During testing the desired output class tag C_{ONCx} is obtained by finding the higher posterior probability as:

$$C_{ONCx} = \arg\max_{C_m} P(C_{ONCx} \mid p_{i-k}, \ldots, p_{i+k}) \quad (2)$$

Where the posterior probabilities are obtained using the ANN classifier.

We trained the ANN system with a lexicon of 100k Danish words containing phonetic transcriptions and syllable boundaries. This lexicon was developed internally and is not publically available. From the 100k words, we firstly extracted the 1000 words randomly selected to conduct Test 1 mentioned in section 4. We didn't consider these 1000 words in the ANN training set. We used the same test sets used to evaluate our rule-based system in order to assess the ANN syllable tagger, namely the same 1000 words randomly selected from the Danish lexicon (with phonetic transcriptions and syllable boundaries) and the 1000 words extracted from a Danish newspaper article. The word accuracy is lower when compared with the rule-base system described above, as depicted in Table 5:

Table 5. *Results of the ANN based syllabifier in incorrect positioned syllable boundaries (% Syl. Mism.) and % of incorrect syllabified words (% WER).*

Corpus set	% Syl. Mis.	% WER
Test 1	6.2	5.9
Test 2	6.0	5.5

7. Conclusions

In this paper, an automatic syllabification for Danish was presented, using a very small set of rules (10 in total). The successful results of 96.9% and 98.7% of word accuracy rate seem to prove that syllabification can be rule-driven in languages like Danish which present more syllabic complexity than Portuguese and Catalan. Rule-based approaches have proven to be computationally lighter than data-driven methods and equally or more efficient. The same corpus set was tested using an ANN based syllabification system adapted to Danish and the results showed a lower accuracy rate: 94.1% and 94.5% respectively for each test set.

An immediate application of the here described automatic syllabification for Danish was the assignment of syllable boundaries to large corpora and a Danish lexicon of 360k words, enabling this information to be used as input for TTS prosody model training, HTS offline training and acoustic models training for SR. Next steps should include developing a stable compound word handling module. This should consist of two sequentially ordered sub-modules. The first one should be a compound word breaker with an orthographical input that looks up the individual lemmas in a pronunciation lexicon. The second module should reassemble, insert word boundary markers, and modify the phonetic transcription according to the rules for pronunciation of Danish compounds. The syllabification algorithm could then be run successfully on the phonetic transcriptions of the reassembled compounds. A similar approach is envisaged for a morphology handling module. Future work will address the syllabification assignment of foreign words.

8. References

[1] Ladefoged, P. A Course in Phonetics. Fourth Edition. Heinle & Heinle. Thomson Learning. Boaston, USA. 2001.

[2] Adsett, C. R. and Marchand, Y. "Are Rule-based Syllabification Methods Adequate for Languages with Low Syllabic Complexity? The Case of Italian", 6th International Speech Communication Association (ISCA) Workshop on Speech Synthesis. Bonn, Germany. 2007.

[3] Sproat, R. (ed.). Multilingual Text-to-Speech Synthesis: the Bell Labs Approach. Kluwer. 1998.

[4] 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.

[5] Teixeira, J. P. A Prosody Model to TTS systems. PhD Thesis, University of Oporto, Portugal. 2004.

[6] Oliveira, C. Do Grafema ao Gesto: Contributos Linguísticos para um Sistema de Síntese de Base Articulatória. PhD Thesis. University of Aveiro, Portugal. 2009.

[7] Weerasinghe, R., Wasala, A. and Gamage, K.." A rule based syllabification algorithm for Sinhala", In Robert Dale, Kam Fai Wong, Jian Su & Oi Yee Kwong (eds.), Natural Language Processing - IJCNLP, 438–449. Berlin / Heidelberg: Springer. 2005.

[8] Braga, D. 2008. Natural Language Processing Algorithms for TTS systems in Portuguese. PhD Thesis. A Coruña University, Spain. (in Portuguese).

[9] Marchand, Y., Adsett, C. R. and Damper, R. I. "Evaluation of automatic syllabification algorithms for English". 6th International Speech Communication Association (ISCA) Workshop on Speech Syn-thesis. Bonn, Germany. 2007.

[10] Marchand, Y. and Damper, R. "Can syllabification improve pronunciation by analogy of English?", Natural Language Engineering 13(1), 1–24. 2007.

[11] G. Bouma, "Finite states methods for hyphenation," Journal of Natural Language Engineering, no. 1, Cambridge University Press. 2002.

[12] Oliveira, C.; Moutinho, L.; Teixeira, A. "On European Portuguese Automatic Syllabification", in. González González, et al. (coords), III Congreso Internacional de Fonética Experimental, Santiago de Compostela: Xunta de Galicia: 461-473. 2007.

[13] Basbøll, H. The Phonology of Danish. Oxford University Press. New York. USA. 2005.

[14] Grønnum, N. 2005. Fonetik og Fonologi – Almen og Dansk. Akademisk Forlag. København.

[15] Henrichsen, P. J. and Skadhauge, P. R. DanPO- a transcription-based dictionary for Danish speech technology. Proceedings of NODALIDA-2005. May 21-21, Joensuu, Finland. 2005.

[16] Rustullet, S.; Braga, D.; Nogueira, J.; Dias, M. "Automatic Word Stress Marking and Syllabification for Catalan TTS", Proceedings of Interspeech 2008, 22-26 September, Brisbane, Australia. 2008.

[17] Kahn, D. Syllable-based generalizations in English phonology. Doctoral dissertation, MIT. 1976.

[18] Selkirk, E. O. "English Compounding and the Theory of Word-structure", M. Moortgat, H. Van der Hulst & T. Hoestra (eds.) The Scope of Lexical Rules. Foris. Dordrecht. 1981.

[19] Grønnum, N. Rødgrød med fløde. Akademisk Forlag. København. 2007

[20] Grønnum, N. "Syllables at multiple levels of representation in Danish", Journées d'études Linguistiques. Nantes, France. 25-27 mars: 24-29. 1999.

[21] Basbøll, H. "The phonological syllable with special reference to Danish". ARIPUC 8, Univ. of Copenhagen: 39-128. 1974.

[22] Schachtenhausen R. Realisering af schwa i spontan dansk tale. M.S. Thesis. University of Copenhagen (unpubl.). 2007.

[23] Jilei, T., "Data-driven approaches for automatic detection of syllable boundaries", 8th International conference on Spoken language processing, October 4 - 8, 2004 Jeju Island, Korea 2004.