Morphossyntactic Disambiguation for TTS Systems

Ricardo Ribeiro*, **Luís Oliveira**†, **Isabel Trancoso**†

*INESC-ID Lisboa/ISCTE, †INESC-ID Lisboa/IST
Spoken Language Systems Lab
R. Alves Redol, 9
1000-029 LISBON, Portugal
{Ricardo.Ribeiro, Luis.Oliveira, Isabel.Trancoso}@inesc-id.pt

Abstract

The purpose of this paper is to present the development of a morphossyntactic disambiguation system (or part-of-speech tagging system) which is intended to be used as a component of a Text-to-Speech (TTS) system for European Portuguese. In the development of the tagger, we compared two approaches: a probabilistic-based approach and a hybrid approach. Besides comparing these two approaches, this paper considers the effects of the different classes of errors on the performance of the complete TTS system.

1. Introduction

The first stage of a Text-to-Speech system is a Text Analysis module, whose purpose is to generate tagged text that will be submitted to the Phonetic Analysis module. Then the next module is the one responsible for the Prosodic Analysis. Pitch and duration information are attached in this phase and the controls for the Speech Synthesis module are generated. The Speech Synthesis module then renders the appropriate voice sound.

The focus of this work is on the first module, Text Analysis (TAM), aiming to extract from the input text the maximum amount of information that may help the task of the remaining modules. This covers a wide range of possibilities that can go from the simple conversion of non orthographic items to more complex syntactic and semantic analysis. There are three basic phases in the TAM module: document structure detection; text normalization and linguistic analysis. The one that concerns us in this paper is the inclusion of a part-of-speech (POS) Tagger in the linguistic analysis.

The next section describes the motivation for using morphossyntactic disambiguation in general and in the context of TTS systems in particular. Section 3 is devoted to the description of the two approaches we have developed: a probabilistic-based approach and a hybrid approach. Section 4 describes the *corpus* and the tagset we have used for training and testing these approaches, and the lexicons involved. Before concluding, we compare the experimental results obtained, considering the effects of the different classes of errors on the performance of the complete TTS system.

2. Importance of morphossyntactic disambiguation for TTS

According to (Jurafsky and Martin, 2000) “the significance of the part-of-speech is that it gives a significant amount of information about the word and its neighbors”. This significant amount of information allow us, for example, to predict which words or word-types can occur in the neighborhood of a given word. That kind of information may be useful in the language models used for speech recognition. In the same way, knowing the part-of-speech of a word can help an information retrieval system to select special words or word-types, such as nouns, from documents.

In TTS systems, POS taggers may also play an important role. In Portuguese, as in other languages, the pronunciation of a word can depend on the word class (or part-of-speech, lexical tag, morphossyntatic class, etc.). For example, the word “almoço” is pronounced “almoço” (closed “o”) if used as a noun, and pronounced “alMOço” (opened “o”) if used as a verb. The same happens with the word “object” in English. “OBject” if used as a noun and “obJECT if used as a verb. Thus, knowing the part-of-speech may help the system produce correct pronunciations for some homograph words. Furthermore, it may also help identifying special classes of vocabulary for which specific pronunciation rules are needed.

On the other hand, part-of-speech information may also contribute to prosodic phrasing and accentuation. Usually, words are spoken continuously until some linguistic phenomena introduces a discontinuity that can be of various forms. Although it is commonly agreed that prosodic structures are not fully congruent with syntactic structures, morphossyntactic information can help to predict where these discontinuities can occur and of what type they can be (Viana et al., 2001). In terms of accentuation, a very basic method to decide if a word is accentable or not may be based on the part-of-speech category of that word, accenting “all and only the content words” (Huang et al., 2001). The content words belong to major open-class categories such as noun, verb, adjective, adverb, and certain closed-class words such as negatives and some quantifiers.

3. Probabilistic and hybrid approaches

In the development of the tagger, we compared two approaches: a probabilistic-based approach and a hybrid approach. The first one was aimed at integration within the Portuguese version of the Festival system. Festival is a modular freely available TTS system developed at the University of Edinburgh (Black et al., 1999). The second approach, on the other hand, is an independent tool that can be integrated in complex systems that need morphossyntactic disambiguation.

3.1. Probabilistic-based approach

The multilingual Festival system provides a part-of-speech tagging module, where the morphological analysis component is totally lexicon based, and the part-of-speech tagging algorithm is a language independent n-gram based trainable tool. This tool is based on Hidden Markov Models (HMMs) and uses the Viterbi algorithm to predict the sequence of tags.

Two specific resources were hence needed by this module: a lexicon and a set of n-gram models.

3.2. Hybrid approach

The developed hybrid approach comprehends three modules: a morphological analysis module, a linguistic-oriented disambiguation rules module and a probabilistic-based disambiguation module.

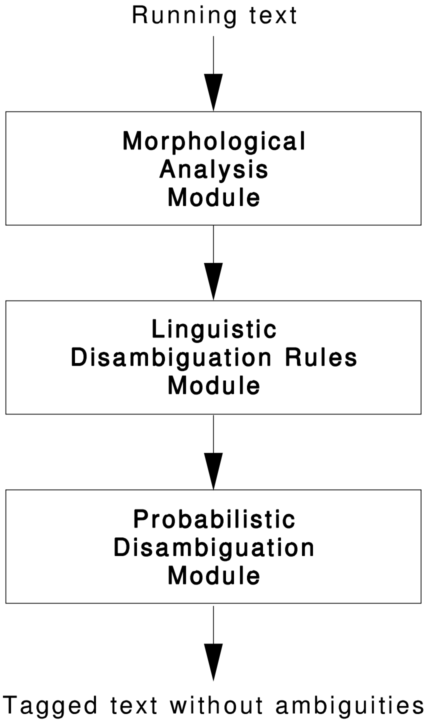
Running text
Morphological Analysis Module
Linguistic Disambiguation Rules Module
Probabilistic Disambiguation Module
Tagged text without ambiguities

Figure 1: Processing sequence.

As can be observed in figure 1, the input of the morphological analysis module is running text that is tagged with all possible part-of-speech tags for each word. Then the linguistic-oriented disambiguation rules module resolves all possible ambiguities, removing possibilities from the previous set of tags for each word. Finally, the probabilistic-based disambiguation module resolves the remaining ambiguities, giving as result the fully disambiguated text.

The morphological analysis module adopted is Palavroso, a broad coverage morphological analyzer developed at INESC (Medeiros, 1995). This analyzer was developed to address specific problems of Portuguese language like compound nouns, enclitic pronouns and adjectives degree. As a result it gives all possible part-of-speech tags for a given word. If a word is not known, it tries to guess possible part-of-speech tags, always giving an answer.

The linguistic-oriented disambiguation rules module is still in development and is based on local grammars. It is inspired in the work of (Voutilainen, 1995). Figure 2 illustrates the rule format. There is an input trigger for the rule, followed by an if-condition that, if satisfied, causes an action to be performed. The rules can also have an else section with an action to perform when the if-condition fails. The work on a set of rules is currently in progress. The rule presented in figure 2 is merely an example of a possible rule which tries to disambiguate the past participle from adjective in Portuguese, given the tag of the previous word. When the input token has an adjective/verb ambiguity (`AMB="A= V="`), if the previous token is tagged as a verb (`-1/TAG="V="`), then the resulting tag is verb.

```
Input: AMB = "A= V="
If
    (-1/TAG="V=")
Then
    "V="
```

Figure 2: Disambiguation rule.

The probabilistic-based disambiguation module is also based on HMMs and uses the Viterbi algorithm to find the most likely sequence of tags for the given sequence of words, and the forward algorithm to compute the lexical probabilities. The forward algorithm is presented in (Allen, 1995). The forward probability ($\alpha_i(t)$) is the probability of producing the $w_1, \cdots, w_t$ word sequence and ending on the state w_t/T_i, where T_i is the i^{th} tag of the tagset.

$$\alpha_i(t) = P(w_t/T_i, w_1, \cdots, w_t)$$

Then we can derive the probability of a word w_t being an instance of lexical category T_i as

$$P(w_t/T_i|w_1, \cdots, w_t) = \frac{P(w_t/T_i, w_1, \cdots, w_t)}{P(w_1, \cdots, w_t)}$$

Estimating the value of $P(w_1, \cdots, w_t)$ by summing over all possible sequences up to any state at position t, we obtain:

$$P(w_t/T_i|w_1, \cdots, w_t) \cong \frac{\alpha_i(t)}{\sum_{j=1,N} \alpha_i(t)}$$

4. Linguistic resources

4.1. *Corpus*

The *corpus* used for training and testing was developed in the LE-PAROLE project (Bacelar et al., 1997). This project in the Language Engineering area was financed by the European Commission, in the context of the Telematics Applications of Common Interest program. Institutions from 15 European countries have participated in this project, whose aim was to develop the initial core of a set of written language resources for the European Community countries. Harmonized reference *corpora* and generalist *lexica* were developed according to a common model for the 12 European languages involved.

The *corpus* used in the present work is a subset of about 290,000 running words of the collected 20 million running words *corpus* for European Portuguese. This subset was morphossyntactically tagged using Palavroso and manually disambiguated. The tagset had about 200 tags with information that varied from grammatical category to morphological features that could be combined to form composed tags (resulting in about 400 different tags). The information coded by the tagset is presented in table 1.

Category	Subcategory	Features
Noun	proper common	gender and number
Verb	main auxiliary	mood; tense; person; gender and number
Adjective		degree; gender and number
Pronoun	personal demonstrative indefinite possessive interrogative relative exclamative reflexive reciprocal	person; gender; number; case and formation
Article	definite indefinite	gender and number
Adverb		degree
Adposition		formation; gender and number
Conjunction	coordenative subordinative	
Numeral	cardinal ordinal	gender and number
Interjection		
Unique	mediopassive	
Residual	foreign abbreviation acronym symbol	
Punctuation		

Table 1: Morphossyntactic information.

The tagset was fully harmonized between all the languages involved. Each tag is an array, and each position of the array codes one of the features presented in table 1, saving the first for the grammatical category and the second for the subcategory. When a position (category, subcategory or feature) is not used, its code is replaced by an equal sign. For example, `R=r` means adverb with no subcategory, in *regular* degree.

This *corpus* was divided into training and test subsets. The training *corpus* has about 230,000 running words and it covers about 25,000 different word forms. The test *corpus* has about 60,000 running words, of which about 900 are words marked as errors, 21,000 are ambiguous (34.6%) and the remaining 38,000 are non-ambiguous. It includes around 10,000 different word forms, with 1.73 tags per word on average and 30.69% different ambiguous word forms.

The tagset used by the taggers was obtained by downsizing the LE-PAROLE tagset to 54 tags. Only the information about the grammatical category and subcategory was retained.

4.2. *Lexica*

The lexicon used by the probabilistic approach has about 21,000 entries with associated probabilities and about 1.4 tags per entry. All the information in the lexicon was obtained from the above training *corpus*. As this lexicon must be used by the POS tagging module of Festival, the whole *corpus* was normalized, and all tokens involving digits, for instance, were converted to an alphabetic form.

All entries were processed by Palavroso and all the part-of-speech tags not occurring in the training *corpus* for an entry, were added to that entry. In order to avoid assigning null probabilities to these non-occurring tags, we used the add-one smoothing technique (Jurafsky and Martin, 2000). For the n-grams models, we used trigram models also obtained from the normalized training *corpus*.

The probabilistic module of the hybrid approach also uses similar lexicon and trigram models. The lexicon, however, is larger (about 25,000 entries), due to the fact that it was derived from the training *corpus* without normalization.

In order to analyse the influence of the taggers in the Phonetic Analysis module, we used the main lexicon of the Portuguese version of Festival. This lexicon contains about 79,000 different entries, each characterized by POS tags and corresponding pronunciation. It includes 76 different types of ambiguities. The most frequent are adjective/common noun, adjective/verb, and common noun/verb.

Tag	Description
A=	adjective
Cc	coordenative conjunction
I	interjection
Mc	cardinal numeral
Mo	ordinal numeral
Nc	common noun
Np	proper noun
Pd	demonstrative pronoun
Pp	personal pronoun
R=	adverb
S=	adposition
Td	definite article
V=	verb
Xf	foreign word

Table 2: Tags description.

However, the number of ambiguities that have influence in the Phonetic Analysis module, causing different pronunciations, is only 16. In table 3 they are presented with the percentage of different word forms of the lexicon with that kind of ambiguity. In order to simplify the next tables with results, table 2 shows the abbreviation tags involved in the disambiguation of homograph words.

Ambiguity	Different word forms (%)
A= Nc V=	0.876%
A= Np V=	0.009%
A= V=	2.957%
Cc Nc	0.001%
I R= V=	0.001%
Mc Mo	0.005%
Mc Mo Nc	0.001%
Mo Nc	0.001%
Mo V=	0.005%
Nc Np V=	0.051%
Nc Pd Pp Td	0.003%
Nc R= V=	0.007%
Nc V=	3.936%
Np Xf	0.023%
R= V=	0.013%
S= V=	0.017%

Table 3: Ambiguities that influence the Phonetic Analysis module.

Ambiguity	Probabilistic approach	Hybrid approach
A= Nc V=	9.96%	10.53%
A= Np V=	0.00%	0.00%
A= V=	14.37%	12.32%
Cc Nc	0.19%	0.07%
I R= V=	18.03%	13.11%
Mc Mo	1.75%	1.75%
Mc Mo Nc	0.40%	0.40%
Mo Nc	0.28%	0.37%
Mo V=	1.50%	2.40%
Nc Np V=	6.86%	9.80%
Nc Pd Pp Td	4.53%	7.10%
Nc R= V=	18.18%	16.36%
Nc V=	5.96%	4.29%
Np Xf	0.00%	0.00%
R= V=	28.37%	25.00%
S= V=	2.38%	2.54%

Table 6: Error rates obtained for the ambiguities shown in table 3.

5. Experimental results

Table 4 shows the overall POS error rates obtained with the two approaches and table 5 presents the error rates obtained for some relevant part-of-speech categories.

Approach	Error rate
Probabilistic	8.24%
Hybrid	7.17%

Table 4: Overall error rates.

POS	Probabilistic	Hybrid
Proper noun	22.69%	22.15%
Common noun	5.23%	3.80%
Verb	9.17%	4.42%
Adjective	10.87%	15.38%
Adverb	6.87%	5.56%

Table 5: Error rates for some relevant POS.

The error rate for proper nouns is not really very significant, since adding new entries to the lexicon will improve this rate. The high error rate obtained for adjectives may be explained by the relative large percentage of adjective/verb in past participle ambiguity.

It is important to observe that a significant part of the errors made by the taggers was obtained when trying to tag unknown words. In fact, the number of words in the test *corpus* that do not occur in the training *corpus* is around 4,400, corresponding to 3,200 different forms.

Table 6 further discriminates these error rates in terms of the different kinds of ambiguity relevant for homograph disambiguation.

Concerning the influence of part-of-speech tagging in the prosodic processing, we conducted several preliminary studies in the context of the different phrasing methods evaluated in (Viana et al., 2001). Our first experiment consisted of computing the percentage of errors in content/function word classification, to which the phrasing algorithms are mostly sensitive. The probabilistic approach resulted in 0.90% errors and the hybrid one in 0.65% errors.

Our second experiment consisted of verb classification, since it is relevant for correctly assigning the pitch contour. The probabilistic tagger failed to identify a verb in 9.17% of the occurrences, whereas the hybrid approach failed only in 4.42% of the times.

As a final remark it is possible to observe that the hybrid approach has a better overall performance. Regarding the influence on the Phonetic Analysis module, the probabilistic-based approach has better results in six kinds of ambiguity, but with no significant differences. Exception made to “Nc Np V=” and “Nc Pd Pp Td” ambiguities. In the same analysis, the hybrid approach has also better results in six kinds of ambiguity, but with larger differences in four of them. Regarding the influence on the Prosodic Analysis module, the hybrid approach has clearly a better performance than the probabilistic-based one. The error rate is smaller both in terms of content/function word classification and in terms of verb identification.

6. Conclusions and future work

This study allowed us to have an idea of what type of disambiguation errors are mostly relevant in the context of TTS systems for deriving the correct pronunciation of homograph words. Further work is still necessary in order to optimize the rule-based module and also in order to obtain a broader lexical coverage. Future work will concentrate on these issues and also on evaluating more thoroughly the impact of disambiguation errors on prosodic phrasing.

7. Acknowledgments

The authors wish to thank Maria do Céu Viana for many helpful discussions.

8. References

James Allen. 1995. *Natural Language Understanding*. The Benjamin/Cummings Publishing Company, Inc.

Fernanda Bacelar, José Bettencourt, Palmira Marrafa, Ricardo Ribeiro, Rita Veloso, and Luzia Wittmann. 1997. Le-parole - do corpus à modelização da informação lexical num sistema multifunção. In *Actas do XIII Encontro da Associação Portuguesa de Linguística*, Portugal.

A.W. Black, P. Taylor, and R. Caley, 1999. *The Festival Speech Synthesis System*. University of Edimburgh.

X. Huang, A. Acero, and H. Hon. 2001. *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development*. Prentice Hall.

Daniel Jurafsky and James H. Martin. 2000. *Speech and Language Processing*. Prentice Hall.

José Carlos Medeiros. 1995. Processamento morfológico e correcção ortográfica do português. Master's thesis, Instituto Superior Técnico - Universidade Técnica de Lisboa, Portugal.

M.C. Viana, L.C. Oliveira, and A.I. Mata. 2001. Prosodic phrasing: human and machine evaluation. In *Proc. 4th ISCA Workshop on Speech Synthesis*, Scotland.

A. Voutilainen, 1995. *Constraint Grammar: a Language-Independent System for Parsing Unrestricted Text*, chapter Morphological disambiguation. Mouton de Gruyter.