Using Morphossyntactic Information in TTS Systems: Comparing Strategies for European Portuguese

Ricardo Ribeiro[1], Luís Oliveira[2], and Isabel Trancoso[2]

[1] INESC-ID Lisboa/ISCTE
[2] 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. To improve the quality of the speech produced by a Text-to-Speech (TTS) system, it is important to obtain the maximum amount of information from the input text that may help in this task. 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. In this paper, we present the development of a morphossyntactic tagging system and analyze its influence on the performance of a TTS system for European Portuguese.

1 Introduction

The information obtained by a morphossyntactic tagging system can be relevant in several areas of natural language processing. For example, knowing the part-of-speech of a given word allow us to predict which words (or word-types) can occur in its neighborhood. That kind of information is useful in the language models used for speech recognition. Morphossyntactic information can also be used by automatic term acquisitions systems or information retrieval systems to select special words (or word-types) or to know which affixes a given word can take. In the same way, a morphossyntactic tagger can help a Text-to-Speech (TTS) system improve the quality of the produced speech.

The first stage of a TTS 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.

There are three basic phases in the Text Analysis module: document structure detection; text normalization; and linguistic analysis. The one that concerns us in this paper is the inclusion of a morphossyntactic tagger in the linguistic analysis.

The information obtained by a morphossyntactic tagging system is relevant to the Phonetic and Prosodic Analysis modules. Concerning the Phonetic Analysis module, in Portuguese, as in other languages, the pronunciation of a word

N.J. Mamede et al. (Eds.): PROPOR 2003, LNAI 2721, pp. 143–150, 2003.

can depend on the word class (or part-of-speech, lexical tag, morphossyntactic class, etc.). For example, the word “almoço” is pronounced “almoço” (close “o”) if used as a noun, and pronounced “alMOço” (open “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. Morphossyntactic information may also influence the performance of the Prosodic Analysis module, contributing 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 [13]. 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” [7]. 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.

The next section describes the part-of-speech tagging system developed for Portuguese. Section 3 describes the *corpus* and the tagset we have used for developing the system, and the lexicons involved. Before concluding, we compare the results obtained by the developed system with the ones achieved by other taggers based on different approaches, considering the effects of the different classes of errors on the performance of the complete TTS system.

2 Morphossyntactic Tagging System

The morphossyntactic tagging process we have implemented consists of the two sequential steps illustrated in figure 1.

The separation between morphological analyzis and ambiguity resolution was motivated by the fact that neolatin languages, such as Portuguese, are highly inflectional when compared with English. In this sense, morphological analysis can be relevant. In fact, on the one hand, linguistic oriented systems are usually based on the elimination of the ambiguity previously introduced by a lexical analysis process, and, on the other hand, in data-driven approaches, information is derived from *corpora* and due to data sparseness word forms may not appear with all possible tags or even not occur at all [8,10].

The morphological analysis module adopted is Palavroso, a broad coverage morphological analyzer [9] 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 disambiguation module, developed in the context of this work, is MARv (Morphossyntactic Ambiguity Resolver). MARv’s architecture comprehends two

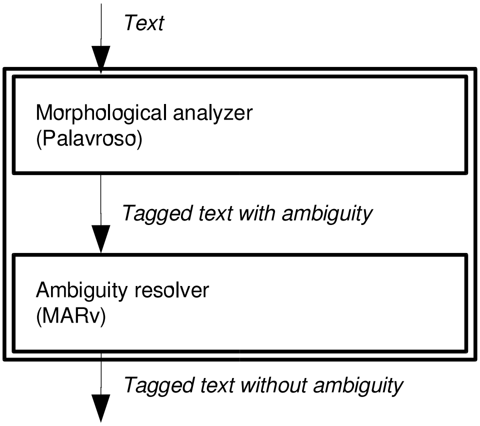

Fig. 1. Architecture of the morphossyntactic tagging system

modules: a linguistic-oriented disambiguation rules module and a probabilistic disambiguation module. The ambiguity is first reduced by the disambiguation rules module and then the probabilistic module produces a fully disambiguated output.

The disambiguation rules module is based on a set of contextual rules developed specifically for Portuguese. The rules have the following structure: an input trigger section; an *if*-condition; and an action section.

```
Input: AMB = ‘‘A= Nc V=’’
If
     (-1/TAG = ‘‘S=’’)
then
     SELECT ‘‘Nc’’
```

Fig. 2. Disambiguation rule

As shown in figure 2, the input trigger consists of a simple condition where it is verified if the observed input matches an ambiguity class (`AMB`) or a given word. If the rule is triggered, the *if*-condition is evaluated. The terms involved have the following format:

$$(position\_relative\_to\_the\_observed\_input/keyword\ [\ = | \neq\ ]\ value)$$

where *keyword* can be `TAG`, `AMB` or `WORD`. The actions to be performed may be of two types: a selection (`SELECT`) of a single tag or a removal (`REMOVE`) of a set of tags. The actual set of rules includes 35 rules [6].

The probabilistic-based disambiguation module is based on Markov models 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 [1]. 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:

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

An in depth description of this system can be found in [11].

3 Linguistic Resources

3.1 *Corpus*

The *corpus* used for training and testing was developed in the LE-PAROLE European project [2] in which harmonized reference *corpora* and generalist *lexica* were built 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.

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=n` means adverb with no subcategory, in *normal* degree.

This *corpus* was subdivided 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

Table 1. Morphossyntactic information

Category	Subcategory	Features	Tag
Noun	proper common	gender and number	Np Nc
Verb	main auxiliary	mood; tense; person; gender and number	V=
Adjective		degree; gender and number	A=
Pronoun	personal demonstrative indefinite possessive interrogative relative exclamative reflexive	person; gender; number; case and formation	Pp Pd Pi Po Pt Pr Pe Pf
Article	definite indefinite	gender and number	Td Ti
Adverb		degree	R=
Adposition		formation; gender and number	S=
Conjunction	coordinative subordinative		Cc Cs
Numeral	cardinal ordinal	gender and number	Mc Mo
Interjection			I
Unique	mediopassive		U
Residual	foreign abbreviation acronym symbol		Xf Xa Xy Xs
Punctuation			O

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 down-sizing the LE-PAROLE tagset to 54 tags. Only the information about the grammatical category and subcategory was retained.

3.2 *Lexica*

The lexicon used by the probabilistic module of the disambiguation system has about 25,000 entries with associated probabilities. All the information in the lexicon was obtained from the above training *corpus*.

In order to analyze 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 morphossyntac-

Table 2. Ambiguities that influence the Phonetic Analysis module

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

Table 3. Evaluated taggers

Identification	Description	Approach
A	Markov models tagger integrated in Festival speech synthesis system [3]	Probabilistic
B	Transformation-based tagger, developed by [5]	Symbolic learning/Rule-based

Table 4. Overall success rates

System	Success rate
A	92.05%
B	95.17%
C	**94.23%**

tic tags and corresponding pronunciation. It includes 76 different types of ambiguities. The most frequent are adjective/common noun, adjective/verb, and common noun/verb. However, the number of ambiguities that have influence in the Phonetic Analysis module, causing different pronunciations, is only 16. In Table 2 they are presented with the percentage of different word forms of the lexicon with that kind of ambiguity.

4 Experimental Results

To analyze the performance of the developed system, two other taggers were adapted for European Portuguese (table 3) and a comparative evaluation was made.

The following tables present the success rates achieved by the taggers. The system presented in Sect. 2 is identified with the letter C. Table 4 shows the overall success rates and Table 5 discriminates the success rate for morphossyntactic descriptions (MSD) that comprehend content words.

The best overall success rate was achieved by the transformation-based tagger (B). Concerning the identification of content words, the differences for proper

Table 5. Success rates achieved in identifying content words

MSD	A	B	**C**
Proper noun	76.84%	93.69%	**89.19%**
Common noun	94.73%	95.24%	**97.07%**
Verb	90.38%	96.11%	**96.93%**
Adjective	89.11%	86.99%	**85.23%**
Adverb	93.12%	96.52%	**95.06%**

Table 6. Error rates obtained for the ambiguities presented in Table 2

Ambiguity	A	B	**C**
A= Nc V=	9.96%	13.03%	**10.34%**
A= Np V=	0.00%	0.00%	**0.00%**
A= V=	14.37%	11.00%	**10.70%**
Cc Nc	0.19%	0.02%	**0.10%**
I R= V=	18.03%	4.92%	**13.11%**
Mc Mo	1.35%	0.00%	**1.35%**
Mc Mo Nc	0.40%	0.08%	**0.40%**
Mo Nc	0.05%	0.05%	**0.14%**
Mo V=	1.50%	0.00%	**2.40%**
Nc Np V=	6.86%	1.96%	**9.80%**
Nc Pd Pp Td	4.53%	2.47%	**6.96%**
Nc R= V=	18.18%	1.82%	**7.27%**
Nc V=	4.85%	3.24%	**2.82%**
Np Xf	0.00%	0.00%	**0.00%**
R= V=	0.48%	0.00%	**0.00%**
S= V=	0.79%	0.32%	**0.16%**

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

In order to stress the influence of the taggers on the performance of the TTS system, the presented values are error rates. 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 [13]. Our first experiment consisted of computing the percentage of errors in content/function word classification, to which the phrasing algorithms are mostly sensitive. The system A made 1.18% errors, the developed system (C) had error rate of 0.64% and the best result was obtained by the system B. Our second experiment consisted of verb classification, since it is relevant for correctly assigning the pitch contour. The best result was achieved by the system C, failing to identify a verb 3.07% of the times where the system with best overall success rate (B) had an error rate of 3.89%.

5 Conclusions

This paper reported the work done in the development of a morphossyntactic tagging system for the Portuguese language, an area where the scarce resources still demand for new contributions ([4,12]). The developed system was compared with other taggers that implemented other approaches to this problem and the results were positive (an analysis of some of the available systems for Portuguese can be found in [11]). This study also allowed us to understand what are the disambiguation errors that influence the performance of the TTS system and which are the most relevant ambiguity classes.

References

1. J. Allen. *Natural Language Understanding.* The Benjamin/Cummings Publishing Company, Inc, 1995.
2. F. Bacelar, J. Bettencourt, P. Marrafa, R. Ribeiro, R. Veloso, and L. Wittmann. LE-PAROLE – Do corpus à modelização da informação lexical num sistema multifunção. In *Actas do XIII Encontro da APL*, Portugal, 1997.
3. A.W. Black, P. Taylor, and R. Caley. *The Festival Speech Synthesis System.* University of Edimburgh, 1999.
4. A.H. Branco and J.R. Silva. EtiFac: A facilitating tool for manual tagging. In *Actas do XVII Encontro Anual da APL*, pages 1427–1431, Lisboa, Portugal, 2002. APL e Colibri.
5. E. Brill. Transformation-based error-driven learning and natural language processing. *Computational Linguistics*, 21(4), 1995.
6. Caroline Hagège. Personal communication, 2001.
7. X. Huang, A. Acero, and H. Hon. *Spoken Language Processing: A Guide to Theory, Algorithm, and System Development.* Prentice Hall, 2001.
8. É. Laporte. *Tratamento das Línguas por Computador*, chapter Resolução de ambiguidades. Caminho, 2001.
9. José Carlos Medeiros. Processamento morfológico e correcção ortográfica do português. Master's thesis, Instituto Superior Técnico, Portugal, 1995.
10. C. Oravecz and P. Dienes. Efficient stochastic part-of-speech tagging for hungarian. In *Proc. of the Third LREC*, pages 710–717, Las Palmas, Espanha, 2002. ELRA.
11. Ricardo Ribeiro. Anotação morfossintáctica desambiguada do português. Master's thesis, Instituto Superior Técnico, Portugal, 2003.
12. T.B. Sardinha. Compilação e anotação de um *corpus* de português de linguagem profissional. *The Especialist*, 21(1):111–147, 2000.
13. M.C. Viana, L.C. Oliveira, and A.I. Mata. Prosodic phrasing: human and machine evaluation. In *Proc. of the 4^{th} ISCA Workshop on Speech Synthesis*, Scotland, 2001.