



INTERNATIONAL MASTERS IN NATURAL LANGUAGE  
PROCESSING AND HUMAN LANGUAGE TECHNOLOGY

# Automatic Disambiguation of-*mente* ending Adverbs in Brazilian Portuguese

Gaia Olivia da Paixão Fernandes

**Main Advisor:** Dr. Jorge Baptista

**Co-Advisor:** Dr. Dolors Català

Universidade do Algarve  
Faculdade de Ciências Humanas e Sociais

Universitat Autònoma de Barcelona  
Facultat de Filosofia i Lletres



June 2011



## Abstract

Many researches on Word Sense Disambiguation (WSD) have been conducted in the last fifty years and there still much work to do to solve the ambiguity problem. This dissertation presents and discusses the linguistic and computational difficulties of dealing with adverbs, mainly the *-mente* ending adverbs in Brazilian Portuguese. To achieve this aim, first, recent works and grammars were analysed in order to provide a general review of how these words are treated in linguistic field. The conclusion is that the adverbs are poorly treated and their classification is usually based only on their semantic features. These works were compared to a noteworthy study on *-ment* French adverbs undertaken by Molinier and Levrier (2000) that is based on the Lexicon-Grammar methodological and theoretical framework developed by Gross (1986). The authors attempt to provide a syntactic-semantic systematic classification of these adverbs in French. This was the inspiration for proposing a systematic, in other words, consistent and reproductive, classification for the *-mente* Portuguese adverbs. All adverbs were selected from the NILC/São Carlos corpus and only the most frequent ones were manually classified, more precisely 532. It was observed that 98 of them were ambiguous. This means that they could be assigned to more than one class, and consequently had different meanings. Fifty-three of them were taken into consideration to perform two main supervised machine learning experiments using the Weka workbench. The aim of these experiments was to compare the performance of nine classifiers, to learn more about the behavior of the adverbs and try to establish useful features for their disambiguation. To perform these experiments a training dataset was built with 2650 instances (53 *-mente* adverbs) and each one of them was manually classified into one of the 12 classes proposed and the part-of-speech of all words were automatically annotated. A total of 17 features were proposed and extracted. In general, satisfactory results were obtained. This study may improve the treatment of *-ment* ending adverbs in parsers and Machine Translation (MT) systems.

**Key-words:** Adverbs, Classification, Word Sense Disambiguation, Machine Learning, Automatic Disambiguation



## Resumo

Nos últimos cinquenta anos, foram realizadas muitas pesquisas na área da Desambiguação Lexical de Sentido (DLS), porém ainda há muito trabalho para ser feito a fim de que os problemas de ambiguidade sejam totalmente resolvidos. Esta dissertação apresenta e discute as dificuldades linguísticas e computacionais do tratamento dos advérbios do Português do Brasil, mais especificamente dos advérbios em *-mente*. Para atingir este propósito, primeiro, foram analisadas gramáticas e trabalhos mais recentes com o intuito de apresentar uma revisão geral de como são tratadas estas palavras na área da linguística. Observou-se que os advérbios são tratados superficialmente e que a sua classificação geralmente é baseada somente nas suas características semânticas. Estes trabalhos foram comparados com um estudo realizado por Molinier and Levrier (2000) sobre os advérbios terminados em *-ment* em francês, que por sua vez baseou-se na linha teórica e metodológica do Léxico-Gramática desenvolvida por Gross (1986). Os autores propõem uma classificação sintático-semântica para este tipo de advérbios em francês. Uma classificação sistemática, ou seja, consistente e reproduzível, para os advérbios terminados em *-mente* do português do Brasil foi proposta com base no último estudo realizado. Os advérbios analisados neste trabalho foram selecionados a partir do corpus NILC/São Carlos e somente os mais frequentes (532) foram analisados. Observou-se que 98 destes advérbios são ambíguos, isto significa que podem ser classificados em mais de uma classe, possuindo diferentes significados. Destes advérbios, foram levados em consideração 53 deles para realizar dois experimentos supervisionados de aprendizagem automática de máquina usando o programa Weka. O objetivo destes experimentos era comparar o rendimento de nove algoritmos e analisar o comportamento dos advérbios para tentar estabelecer características úteis para a sua desambiguação. Dessa forma, foi construído um conjunto de dados de teste com 2650 exemplos e cada um deles foi classificado manualmente numa das 12 classes propostas. Todas as palavras foram etiquetadas (classes gramaticais) automaticamente através de um anotador categorial. Um total de 17 características foram propostas e extraídas. Em geral, os resultados obtidos foram satisfatórios. Este trabalho visa melhorar o tratamento dos advérbios terminados em *-mente*, principalmente, em analisadores sintáticos e sistemas de tradução automática.

**Palavras-chave:** Advérbios, Classificação, Desambiguação Lexical de Sentido, Aprendizado de Máquina, Desambiguação Automática



## Resumen

En los últimos cincuenta años, se han desarrollado muchas investigaciones en el área de la Desambiguación del Sentido de las Palabras, sin embargo todavía hay mucho trabajo por hacer para resolver el problema de la ambigüedad. En este estudio, presentamos y discutimos las dificultades lingüísticas y computacionales del tratamiento de los adverbios del Portugués de Brasil, particularmente los terminados en *-mente*. Hemos expuesto los diferentes enfoques de las gramáticas y de los trabajos portugueses más recientes sobre esta categoría gramatical. En general, los adverbios son tratados superficialmente y su clasificación está basada esencialmente en sus características semánticas. Después, hemos comparado estas visiones con un estudio realizado por Molinier and Levrier (2000) sobre los adverbios franceses terminados en *-ment*. Los autores proponen una clasificación sintáctico-semántica para este tipo de adverbios basándose en el marco teórico y metodológico del Léxico-Gramática desarrollado por Gross (1986). Siguiendo estos principios, nos hemos planteado llevar a cabo una clasificación sistemática, es decir, consistente y reproducible, de los adverbios tratados en este estudio para tratar los posibles problemas de ambigüedad. Por ello, hemos extraído 1936 adverbios del corpus NILC/São Carlos de los cuales hemos clasificado sólo los más frecuentes (532) y representativos. Nos hemos centrado sobre 53 de ellos por su ambigüedad, pueden ser asignados a más de una clase, y tienen diferentes significados. Hemos seleccionado frases donde aparecían estos adverbios a fin de realizar dos experimentos supervisados de aprendizaje automático usando el programa Weka. El objetivo de estos experimentos era comparar el rendimiento de nueve clasificadores, entender el comportamiento de los adverbios e intentar establecer características útiles para su desambiguación. Para dichos experimentos, hemos construido un conjunto de datos de prueba con 2650 ejemplos (53 adverbios terminados en *-mente*), cada uno de ellos fue clasificado manualmente en una de las 12 clases propuestas y las categorías gramaticales de todas las palabras fueron etiquetadas de manera automática. También propusimos 17 características que fueron posteriormente extraídas automáticamente. Los resultados fueron bastante satisfactorios. Esperamos que este estudio pueda suponer un punto de partida para mejorar el tratamiento de los adverbios terminados en *-mente* en analizadores sintácticos y sistemas de traducción automática.

**Palabras-clave:** Adverbios, Clasificación, Desambiguación del Sentido de las Palabras, Aprendizaje Automático, Desambiguación Automática





## Acknowledgements

I gratefully acknowledge the support, guidance and patience of my main advisor Prof. Dr. Jorge Baptista (UALG) from the Universidade do Algarve (UALG). He always found a way to spend some time to guide me and correct this dissertation, even when I went to Barcelona.

I am equally indebted to my co-advisor Prof. Dr. Dolors Català from the Universitat Autònoma de Barcelona (UAB) for her assistance, for always giving me helpful suggestions and for being patient to correct this work.

Last but not the least, I would like to thank all my classmates, especially Jorge Guilherme Carmargo de Souza, Sónia Jacinto and Mariano Felice who helped me a lot while I was writing this dissertation. Thank you for the endless discussions, moral support, for sharing your knowledge with me and for helping me to carry out some tasks.

This project was supported by the European Commission, Education & Training, Erasmus Mundus: EMMC 2008-0083, Erasmus Mundus Masters in NLP & HLT programme.





# List of Abbreviations

<b>Adj</b>	Adjective
<b>Adv-mente</b>	<i>-mente</i> ending adverbs of Brazilian Portuguese
<b>ASC</b>	AttributeSelectedClassifier (Classifier)
<b>BN</b>	BayesNet (Classifier)
<b>Det</b>	Determinant
<b>FC</b>	FilteredClassifier (Classifier)
<b>IBk</b>	k-nearest-neighbor classifier (Classifier)
<b>J48</b>	C.45 decision tree learner (Classifier)
<b>Kstar</b>	Nearest neighbor with generalized distance function (Classifier)
<b>LT</b>	LADTree (Classifier)
<b>MF</b>	Focus adverbs (Adverb Class)
<b>ML</b>	Machine Learning
<b>MP</b>	Viewpoint adverbs (Adverb Class)
<b>MQ</b>	Quantification adverbs (Adverb Class)
<b>MS</b>	Subject-oriented manner adverbs (Adverb Class)
<b>MT</b>	Machine Translation
<b>MT</b>	Time adverbs (Adverb Class)
<b>MV</b>	Proper manner adverbs (Adverb Class)
<b>N</b>	Noun
<b>NB</b>	NaiveBayes (Classifier)
<b>NLP</b>	Natural Language Processing
<b>PAa</b>	Disjunctive adverbs of attitude: evaluation (Adverb Class)
<b>PAh</b>	Disjunctive adverbs of attitude: habit (Adverb Class)
<b>PAm</b>	Disjunctive adverbs of attitude: modal (Adverb Class)
<b>PAs</b>	Disjunctive adverbs of attitude: subject-oriented attitude (Adverb Class)
<b>PC</b>	Conjunctive adverbs (Adverb Class)
<b>Prep</b>	Preposition
<b>PS</b>	Disjunctive adverbs of style (Adverb Class)
<b>SMO</b>	Sequential Minimal Optimization (Classifier)
<b>Weka</b>	Waikato Environment for Knowledge Analysis



# Contents

<b>1</b>	<b>Introduction</b>	<b>1</b>
1.1	Motivation . . . . .	1
1.2	Objectives . . . . .	4
1.3	Structure of the dissertation . . . . .	5
<b>2</b>	<b>Related Work</b>	<b>7</b>
2.1	Adverbial Structures in Linguistics . . . . .	7
2.1.1	Types of Adverbial Structures . . . . .	7
2.1.2	General Properties . . . . .	9
2.1.2.1	Scope . . . . .	9
2.1.2.2	Mobility and Position . . . . .	13
2.1.2.3	Relation with Interrogative Pro-Forms . . . . .	18
2.1.3	Classifications of Adverbial Structures . . . . .	21
2.1.3.1	A Lexicon-Grammar Classification . . . . .	23
2.2	Adverbial Structures in NLP . . . . .	36
2.3	A Brief Overview of WSD . . . . .	39
2.3.1	The Weka Package . . . . .	42
<b>3</b>	<b>Classification of <i>Adv-mente</i></b>	<b>43</b>
3.1	Corpus . . . . .	43
3.2	Selection of <i>Adv-mente</i> . . . . .	44
3.3	Classification Issues . . . . .	45
3.4	Classification Results . . . . .	52
3.5	Comparison with Electronic Dictionaries . . . . .	56

<b>4</b>	<b>WSD: An Experiment on <i>Adv-mente</i></b>	<b>60</b>
4.1	General observations . . . . .	60
4.2	<i>Adv-mente</i> used in the experiment . . . . .	62
4.3	Building the Training Data . . . . .	62
4.4	Annotations . . . . .	64
4.4.1	Hand-labeled annotation . . . . .	65
4.4.2	Automatic annotation . . . . .	67
4.5	Features and Extraction of Values . . . . .	70
4.6	ML Experiments . . . . .	74
4.6.1	Training and Testing Data . . . . .	74
4.6.2	Learning Algorithms . . . . .	75
4.6.3	Description of Experiments . . . . .	77
4.6.4	Evaluation: baseline and other measures . . . . .	78
<b>5</b>	<b>Results and Discussion</b>	<b>80</b>
5.1	Experiment 1 . . . . .	80
5.2	Experiment 2 . . . . .	88
<b>6</b>	<b>Conclusions and Future Work</b>	<b>92</b>
6.1	Future Work . . . . .	94
	<b>References</b>	<b>96</b>
	<b>Appendices</b>	<b>103</b>
<b>A</b>	<b>Classification of Brazilian Portuguese <i>Adv-mente</i></b>	<b>104</b>
<b>B</b>	<b>Comparison with AEDic and HEDic</b>	<b>116</b>



# List of Tables

2.1	Classification of adverbs proposed by Molinier and Levrier (2000)	36
3.1	Count of lemmas and tokens of all words, adverbs and <i>Adv-mente</i> of the NILC/São Carlos corpus. Source: Linguateca.	44
3.2	Frequency Classes - Distribution of <i>Adv-mente</i> of the NILC/São Carlos corpus	45
3.3	General Distribution based on the Classification of the 532 <i>Adv-mente</i>	52
3.4	Frequency of each class based on the classification of all 532 <i>Adv-mente</i>	53
3.5	Frequency of each class based on the classification of the <i>Adv-mente</i> of group 1	54
3.6	Frequency of each class based on the classification of the <i>Adv-mente</i> of group 2, 3 and 4	54
3.7	Combinations of classes in group 2	55
3.8	Combinations of classes in group 3	55
3.9	Combinations of classes in group 4	56
3.10	Comparison between the <i>Adv-mente</i> of Brazilian Dictionaries and the most frequent <i>Adv-mente</i> of the NILC/São Carlos corpus	58
4.1	Ambiguous <i>Adv-mente</i> selected to be used in the ML experiment	62
4.2	Number of <i>Adv-mente</i> constructions identified in the 2650 sentences	66
4.3	Number of instances per class originally selected <i>Adv-mente</i>	67
4.4	The features proposed, their respective definitions and values	71
4.5	Illustrative example of POS extraction	74



4.6	MFS Baseline . . . . .	79
5.1	Weka classifiers accuracy: 1st experiment . . . . .	81
5.2	The four best classifiers: precision, recall and f-measure . . . . .	82
5.3	Standard Deviation of Classifiers for each <i>Adv-mente</i> . . . . .	83
5.4	Standard Deviation of the performance of each Classifier . . . . .	84
5.5	Selected features: Weka's CfsSubsetEval . . . . .	86
5.6	Classifiers' performance comparison - NB, BN and J48 . . . . .	87
5.7	Classifiers' performance comparison - LT, KStar and IBk . . . . .	87
5.8	Classifiers' performance comparison - ASC, FC, SMO . . . . .	87
5.9	Weka classifiers accuracy: 2nd experiment <i>Adv-mente</i> . . . . .	88
5.10	Classifiers: Classes with precision score equal to 0.00 <i>Adv-mente</i> . . . . .	89
5.11	Weka classifiers accuracy: 22% split for training . . . . .	90
5.12	Selected features: performance of the classifiers . . . . .	91



# Chapter 1

## Introduction

Adverbs are a major part-of-speech that can be defined as a heterogeneous word class composed of semantically, syntactically and morphologically different words (Guimier, 1996, p. 1). The *-mente* (*-ly*) ending adverbs are part of this highly controversial class. One feature shared by *-mente* ending adverbs is that they are morphologically homogeneous, being in most cases, derived from adjectives. At first glance, this feature makes this subclass of adverbs an extremely attractive case of study. However, despite this attractive feature, one should be aware of important semantic and syntactic differences that exist among them (Molinier and Levrier, 2000).

The topic of the present thesis is these *-mente* ending adverbs (more specifically the Brazilian Portuguese *-mente* ending adverbs), hereafter *Adv-mente*, and its choice was mainly inspired by the study of Molinier and Levrier (2000). In this first chapter, the motivations (Section 1.1) and goals (Section 1.2) for this thesis are presented. Finally, the outline of the thesis is given (Section 1.3).

### 1.1 Motivation

Adverbs are a significant part of the lexicon of many languages. In Brazilian Portuguese, about 1.456,426 adverbs have been registered in the 32,3M words corpus NILC/São Carlos<sup>1</sup> (Aluísio and Pinheiro, 2003), which corresponds approximately

---

<sup>1</sup><http://www.linguateca.pt/acesso/corpus.php?corpus=SAOCARLOS>

to 4,50% of the total number of words. This is a significant number and reveals that adverbs are largely used. In addition, in the NILC/São Carlos corpus 2,867 different adverbs (lemmas) were counted and approximately 1,936 (67,52%) of them are derived, *-mente* ending adverbs (hereafter referred to as *Adv-mente*). Therefore, most of the adverbs (lemmas) in the corpus are *Adv-mente*.

In spite of their frequency in texts and their sheer number in the language's lexicon, general grammars, such as Bechara (2001) and Cunha and Cintra (1984), usually only provide a broad semantic classification of this class of words and comment on the morphological relation with the base adjective. In Brito (2003), the author argues that more features should be taken into consideration in order to classify the adverbs, however she does not propose any other classification system. Those classifications are normally absent from current dictionaries Houaiss and Villar (2001) and Almeida Costa and Sampaio e Melo (1999). Usually, Portuguese dictionaries register non-derived adverbs (e.g. *apenas* 'only', *não* 'no'), and only include *Adv-mente* whose meaning cannot be directly related to the base adjective (e.g. *alto* 'high', *altamente* 'highly').

For example, the version 1.0 of the Houaiss Electronic Dictionary (Houaiss and Villar, 2001) of Brazilian Portuguese has only 47 entries of *Adv-mente* (e.g. *infelizmente* 'unfortunately' not to be related to *infeliz* 'unfortunate'). Nevertheless, the latest version of the same dictionary (Houaiss and Mauro, 2009) presents 124 entries of *Adv-mente*. More surprisingly, it is the number of *Adv-mente* entries in the the Aurélio Electronic Dictionary of Brazilian Portuguese (Buarque de Holanda Ferreira, 2004). It has 257 entries of *Adv-mente*<sup>2</sup>. Even that dictionaries seem to include and treat more cases of *Adv-mente*, this number is still insignificant if we compare it with the number of different *Adv-mente* that can be found in a corpus (e.g. NILC/São Carlos Corpus).

Taking into consideration the specific case of Portuguese *Adv-mente*, few works explore the syntactic and semantic aspects of the Portuguese *Adv-mente*. Even though some studies on adverbs of European Portuguese such as Costa (2008) and (Palma, 2009) have been produced, none of them focus on *-mente* ending adverbs. In the former study (Costa, 2008), the author provides a broad description of

---

<sup>2</sup>A discussion about the treatment of the *Adv-mente* in both Brazilian dictionaries will be presented in Chapter 3.

adverbs, whereas the latter proposes a linguistic description and a formal classification of compound adverbs (Palma, 2009). Some studies on French *-ment* ending adverbs have been undertaken in the last years (Guimier, 1996) and (Molinier and Levrier, 2000) with the purpose of providing a syntactic-semantic description and classification of French adverbs ending in *-ment*. To our knowledge, there is still no systematic, syntactic-semantic classification of adverbs (and in particular of the *Adv-mente* that is the largest subclass of adverbs) in Portuguese. Such classification should be based on the use of these adverbs in real texts and should provide a detailed analysis of the possible relations that the adverbs can establish with other structures in a sentence, like the classification that has been made for French equivalent *-ment* ending adverbs (Molinier and Levrier, 2000). The classification proposed by these authors is furthermore necessary since the same *-ment* ending adverb may present more than one possible meaning. The same happens to Brazilian Portuguese *-mente* ending adverbs as the different sentences (1.1) and (1.2) of the same adverb illustrate:

(1.1) *Incrivelmente*, o Pedro sobreviveu ao acidente de carro

Incredibly, Peter survived a car accident

(1.2) O Pedro é *incrivelmente* bonito

Peter is incredibly handsome

In sentence (1.1), the adverb *incrivelmente* refers to the whole sentence and expresses a value judgment about the fact that Peter survived a car accident. In sum, the subject of the enunciation expresses that it was hard to believe that Peter survived the accident. However, in sentence (1.2), the adverb *incrivelmente* refers to the adjective *bonito* ‘handsome’, quantifying it. This adverb gives the information that Peter was ‘to a very high degree’ or extremely handsome.

This ambiguity leads to the following questions:

1. Is it possible to determine useful features by analysing *Adv-mente* in context to disambiguate them? If so, which are these features?
2. Is it feasible to determine automatically which sense of the *Adv-mente* is being used in a sentence based on these features?

The research that was undertaken by Molinier and Levrier (2000) helped in establishing some linguistic features, e.g. syntactic and semantic, of *Adv-mente*. Other features were defined during the task of building a training corpus based on *Adv-mente* in context (see Section 4.5).

Baptista and Català (2009b) conducted an initial research<sup>3</sup> on the automatic disambiguation of Portuguese and Spanish *-mente* ending adverbs. This work was the motivation for expanding and investigating further the ambiguity of Portuguese *-mente* ending adverbs based on a NLP approach.

## 1.2 Objectives

The objectives of this study are twofold. The first one consists in proposing a systematic syntactic-semantic classification of the most frequent Brazilian Portuguese *-mente* ending adverbs. The second goal concerns the possible applications of this linguistic study to the field of Natural Language Processing (NLP). The study will verify whether *Adv-mente* instances in texts can be automatically classified. The use of this information may be applied to Word Sense Disambiguation (Stevenson, 2003; Jurafsky and Martin, 2009) since the ambiguity of *Adv-mente* is probably connected to the classification issue (Baptista and Català, 2009b). In addition, the disambiguation of *Adv-mente* may be used to improve Machine Translation (MT) systems (Jurafsky and Martin, 2009) as each meaning of the same *Adv-mente* may have different translations. This topic will not be covered in this thesis, but rather suggested. The examples (1.3) and (1.4) illustrate a case of *Adv-mente* ambiguity:

(1.3) *O Pedro comeu **ligeiramente**, pois estava atrasado para ir ao trabalho*

Peter ate **quickly**, because he was late to go to work

(1.4) *O Pedro achou a comida **ligeiramente** apimentada*

Peter found the food **slightly** spicy

In example (1.3), the adverb *ligeiramente* means that Peter ate fast and it refers to the action of eating, whereas in example (1.4) the same adverb means ‘a

---

<sup>3</sup>A more precise explanation about this work can be found in Section 2.2

little’ and it refers to the adjective *apimentada* ‘spicy’. In sum, in Portuguese, we use the same adverb, *ligeiramente*, to refer to two different situations (‘fast’ and ‘a little’), but, in English, there are two different adverbs (‘quickly’ and ‘slightly’) that convey each meaning.

Both Portuguese examples were typed in Google Translator<sup>4</sup> in order to obtain their respective translation into English. The translations proposed by the tool were the following:

(1.5) \**Peter ate **slightly**, because he was late to go to work*

(1.6) *Peter found the food **slightly** spicy*

As it can be seen, Google Translator was able to identify only one of the two different meanings of the adverb *ligeiramente*, and consequently it failed in translating (example 1.5) the first sentence proposed (example 1.3).

### 1.3 Structure of the dissertation

Excluding this introductory chapter, there are five more chapters in the present thesis. In Chapter 2, some recent linguistic studies are introduced to present how adverbial structures (Section 2.1), in particular *-mente* ending adverbs, have been treated. These studies present many aspects of adverbial structures. For this reason, the following topics were selected to be covered due to their relevance to this research: types of adverbial structures and their semantic and formal properties (Section 2.1.2). Moreover, an overview of the classification of *-ment* ending adverbs in French proposed by Molinier and Levrier (2000) is presented (Section 2.1.3). Section 2.2 presents a general view about the treatment given to adverbial structures in the field of Natural Language Processing (NLP). This chapter also presents a brief overview of the Word Sense Disambiguation problem (Section 2.3) and a tool, namely Weka, that was used to carry on the machine learning experiment is described (Section 2.3.1).

---

<sup>4</sup>Google Translator is a free online tool offered by Google to translate texts. It is available in <http://translate.google.com/?hl=pt-br&ie=UTF-8&tab=wT#>

Chapter 3 describes the first task performed in this study. The corpus used and the *Adv-mente* selected to be analysed are presented (Section 3.1 and 3.2). Furthermore, some issues regarding the classification of the *Adv-mente* are discussed and the results of this task are evaluated (Section 3.3 and 3.4). A comparison between the *Adv-mente* selected from the corpus and the *Adv-mente* found in two Brazilian electronic dictionaries is drawn (Section 3.5).

Chapter 4 describes the machine learning experiments undertaken as an attempt to disambiguate the *Adv-mente*. Firstly, some general observations and the ambiguous *Ad-mente* selected to be analysed are presented (Section 4.1 and 4.2). Secondly the building of training dataset is described (Section 4.3), including the selection of a sample of sentences to compile a corpus and the annotations that were made. Furthermore, the features proposed and their extraction from the corpus are explained (Section 4.5). The experiments carried out with Weka are described (Section 4.6).

In Chapter 5 the results of the machine learning experiments are presented and evaluated. Finally, in Chapter 6, conclusions of this study and possible future works arising from it (Section 6.1) are discussed.



# Chapter 2

## Related Work

In this section, a concise discussion of how adverbial structures have been treated in previous works will be presented. The following topics will be covered: types of adverbial structures and a general description of the main formal and semantic properties of those structures, in a critical overview of different authors and their respective works. Special focus will be given to the adverbs ending in *-mente* (*Adv-mente*) within the preceding topics, which explains why most of the examples given are about *Adv-mente*.

### 2.1 Adverbial Structures in Linguistics

#### 2.1.1 Types of Adverbial Structures

To delimit the word class of adverbs is a tough task. Grammarians such as Bechara (2001) or Cunha and Cintra (1984) try to group together completely different words (e.g. *lá* ‘there’, *só* ‘just/only’, *baixo* ‘low’, *altamente* ‘highly’) without relating their status as adverbs and their obvious different, morphological, semantic and syntactic aspects. As a result, it is hard to establish useful common features among words that belong to this class. One approach used to define the class of adverbs is based on its function. According to Bechara (2001) and Cunha and Cintra (1984), adverbs mainly modify verbs, and may also modify adjectives, other adverbs and sometimes whole sentences. Bechara (2001, p. 288) also points out that some adverbs can modify nouns. In Costa (2008) and Molinier and Levrier (2000), it is

also pointed out the characteristic that some adverbs may function as complements of verbs (essential arguments), as the example (2.1) provided by Costa (2008, p. 14) illustrates:

- (2.1) *O Pedro mora ali* # \**O Pedro mora*  
Peter lives there # \*Peter lives

The point is that these attempts to define the word class of adverbs are confused with the existence of structures that can function in the same way as simple adverbs, namely adverbials (or adverbial complements or adjuncts).

In Molinier and Levrier (2000), an interesting and clear distinction between adverbs (word class) and adverbials is presented concerning the French language. This explanation can be also used in the case of Portuguese. For Molinier and Levrier (2000), the class of adverbs is formed by any invariable word or group of words that do not belong to any of the following three classes: prepositions, conjunctions and interjections (Molinier and Levrier, 2000, p. 23). Based on a negative definition, which considers what adverbs cannot be, the authors conclude that this is a residual class. On the other hand, Gross (1986, p. 11) argues that there are three main groups of morphologically different structures, adverbials, that modify verbs or phrases:

1. Adverbs (one single word):
  - non-derived adverbs: *lá* ‘there’, *só* ‘just/only’
  - derived adverbs: *altamente* ‘highly’
2. Compound adverbs (adverbial phrases): *de propósito* ‘on purpose’
3. Adverbial clauses: *quando eu a vi* ‘when I saw her’

Based on this last view, it is possible to note that adverbs constitute a word class that can function as adjuncts. For this reason, it may be more useful to think about the syntax of adverbial structures, independent of the morphology of a word or string of words involved in these structures. This notion is known as *Extended Adverbs* and it can be found in Gross (1986, pp. 11-12). This notion was

adopted by Molinier and Levrier (2000) and it is based on the Lexicon-Grammar methodological and theoretical framework developed by Gross (1986).

Moreover, the different adverbial structures described above may also be grouped in semantic classes. This means that an adverb, a compound adverb and an adverbial clause can all convey the same overall meaning. For instance, in the following four examples all adverbial structures convey the meaning of **time**:

(2.2) *O Pedro veio **ontem***

Peter came yesterday

(2.3) *Não tenho visto a Maria **recentemente***

I haven't seen Mary recently

(2.4) *Eu o verei **depois de amanhã***

I will see you the day after tomorrow

(2.5) ***Quando o Pedro chegou**, eu estava dormindo*

When Peter arrived, I was sleeping

In the following Section 2.1.2, general semantic and formal (syntactic) properties shared by all these adverbial structures, and that are relevant to this study, will be presented.

## 2.1.2 General Properties

### 2.1.2.1 Scope

Adverbial structures may modify or even determine other structures. The idea of determinative adverbs can be found in Gross (1977). For Molinier and Levrier (2000), this determinative value can be performed by focus adverbs as they may determine a noun phrase, for instance:

(2.6) *A loja tem riquíssimo acervo sessentista, **principalmente** bandas americanas de garagem, como Flower Pot Men e Thirteenth Floor Elevators<sup>1</sup>*

---

<sup>1</sup>This example was extract from the NILC/São Carlos corpus available at <http://www.linguateca.pt/>.

As already mentioned, in Section 2.1.1, grammarians such as Bechara (2001) and Cunha and Cintra (1984) usually define the range of structures that an adverb deals with as being verb modifiers, adjective and adverb intensifiers and sentence modifiers. This is said to be the function of adverbs based on the relation that they establish with those structures (verbs, adjectives, adverbs, noun phrases and sentences). It is unclear if the authors consider this property relevant to characterise the adverbs.

In Costa (2008), the property of scope over the entire sentence or its constituents is described as being essential to distinguish two main semantic classes of adverbs: adverbs may modify the whole sentence or the predicate. According to the author this is a semantic property, even having a syntactic basis.

According to this author, there are two main tests that can be done in the base sentence - examples (2.7), (2.9), (2.11) and (2.13) - in order to determine the scope of an adverb, and consequently the two different semantic classes of adverbs proposed by him: cleft sentence and negative sentence. The first test consists in extracting the adverb in a cleft sentence<sup>2</sup>, as examples(2.8) and (2.10) illustrate. The second criterion consists in putting the base sentence in the negative and extract the adverb to the very end of the sentence. This criterion is used to verify if adverbs can be affected by the negative particle when they appear in the final position of a sentence, as it is shown in examples (2.12) and (2.14).

- Cleft sentence:

(2.7) *O Pedro leu o livro ontem*

Peter read the book yesterday

(2.8) ***Foi ontem que o Pedro leu o livro***

It was yesterday that Peter read the book

(2.9) *Infelizmente, tenho de me ir embora*

Unfortunately, I have to leave

(2.10) ***\*É infelizmente que tenho de me ir embora***

It is unfortunately that I have to leave

---

<sup>2</sup>Structure of a cleft sentence: Ser Advérbio que Frase / It “to be” (is/was) Adverb that Sentence

- Negative sentence:

(2.11) *O João corre depressa*

Peter runs quick

(2.12) *O João **não** corre **depressa***

John does not run quick

(2.13) \**Infelizmente, o João corre*

Unfortunately, John runs

(2.14) *O João **não** corre, **infelizmente***

John does not run, unfortunately

In the examples<sup>3</sup> (2.8) and (2.12) *ontem* ‘yesterday’ and *depressa* ‘quick’ are adverbs that modify the verb phrase, whereas in the examples (2.10) and (2.14) *infelizmente* ‘unfortunately’ is an adverb that modifies the entire clause. In addition, sentence (2.14) cannot be acceptable unless a pause (or comma in writing) separates the adverb *infelizmente* ‘unfortunately’ from the main clause. Adverbs that modify the verb phrase present the opposite behavior, since a pause cannot separate them from the sentence, as the following example illustrates:

(2.15) \**O João corre, depressa*

Peter runs, quick

According to Costa (2008), distinguishing these two classes of adverbs, the author does not suffice to describe all relations that an adverb can establish with other structures. Some adverbs do not modify neither a sentence nor a verb phrase, namely adverbs that express quantification, focus and polarity, i.e., *excessivamente* ‘excessively’, *especificamente* ‘specifically’ and *sim* ‘yes’. For this reason, the author treats these classes of adverbs apart by giving other criteria to identify them. For further information about these classes, see Costa (2008, pp. 63-75).

In fact, the relation established between adverbial modifiers and other structures is as an important feature to Molinier and Levrier (2000). They define clear

---

<sup>3</sup>All examples were taken from Costa (2008, pp. 42-43).

formal criteria to describe these relations. However, their approach differs in some points from the one described by Costa (2008). For Molinier and Levrier (2000), all modifiers can be divided into two main classes:

- Sentence-external modifiers - the entire proposition is modified:

(2.16) *Felizmente*, o Pedro virá à escola

Fortunately, Peter will come to school

- Sentence-internal modifiers - adverbs are integrated into the proposition establishing different relations with structures inside the sentence:

(2.17) *O Pedro* entrou *ilegalmente* no país

Peter entered the country illegally

Two formal criteria are used to distinguish these two classes, which are similar but different to those used by Costa (2008):

- *Sentence-external modifiers* must present the following properties:

- (i) a modifier can be detached to the initial position of a negative sentence

(2.18) *Felizmente*, o Pedro *não* vir à escola

Fortunately, Peter will not come to school

- (ii) the same modifier cannot be extracted when using cleft sentences

(2.19) \***É felizmente que** o Pedro virá à escola

It is fortunately that Peter will come to school

- while typical *sentence-internal modifiers* present the opposite behavior:

- (i) a modifier cannot be detached to the initial position of a negative sentence

(2.20) \***Ilegalmente**, o Pedro *não* entrou no país

Illegally, Peter did not enter the country

- (ii) the same modifier can be extracted when using a cleft sentence

(2.21) *Foi ilegalmente que o Pedro entrou no país*

It was illegally that Peter entered the country

Not all adverbial structures of this class present both criteria simultaneously. This and other issues regarding the classification proposed by Molinier and Levrier (2000) will be discussed in 2.1.3.

It is possible to conclude that the two criteria (cleft and negative sentence) described by Costa (2008) are not exactly equal to the other two ones proposed by Molinier and Levrier (2000) that have just been discussed. In the former study, the author initially proposes a general criteria that could be used to classify all adverbs into two semantic classes, but, in fact, his criteria cannot be applied in all cases.

On the one hand Costa (2008) proposes more classes of adverbs that are explained apart in order to cover other cases and that cannot be analysed based on the two criteria proposed by him. On the other hand, Molinier and Levrier (2000) succeed in defining a general syntactic criteria based on the scope of the adverbs with the purpose of dividing them into two major classes. Inside of each main class, the authors established subclasses based on syntactic and semantic criteria to define the different subgroups of *-ment* ending French adverbs. For this reason, the criteria proposed by Molinier and Levrier (2000) seems to be more systematic than the one proposed by Costa (2008).

### 2.1.2.2 Mobility and Position

It is widely accepted that most adverbial structures are able to occupy different positions in a sentence. In the works that have been presented until now, all of them describe this characteristic, but each one uses different approaches to explain this property.

Traditionally, grammars (Cunha and Cintra, 1984) explain that the position of adverbials may be determined based on their relation with other constituents. In Cunha and Cintra (1984) adverbials that modify adjectives, participles or other adverbs appears right before those constituents (e.g. *Ele é **tão** bonito* ‘He is so handsome’). In the case of adverbials that modify verbs, a semantic approach is also adopted to determine the position of those adverbs. For instance, manner

adverbs that modify verbs usually appear after the verb (e.g. *O menino moveu-se lentamente* ‘The boy moved slowly’), whereas time and place adverbials can appear before or after the verb:

(2.22) *Ontem* (,) (eu) fui à escola

Yesterday, I went to the school

(2.23) (Eu) (,) *ontem* (,) fui à escola *ontem*

\*I yesterday went to the school

(2.24) (Eu) fui *ontem* à escola

I went yesterday to the school

(2.25) (Eu) fui à escola *ontem*

I went to the school yesterday

In these last examples, the varied position of the adverb *ontem* may be related to the salient information agreed with different elements of the clause. Notice also that subject is facultative. Commas are optional when they are used before the verb to separate it from the adverb (examples (2.22) and (2.23)) and unacceptable after it (examples (2.24) and (2.25)).

The authors present only a limited number of rules that do not systematically explain why adverbial structures have these properties.

Bechara (2001, p. 290), calls attention to the fact that the presence or absence of a strong relation between an adverbial structure and the verb may explain why some adverbials are more mobile than others. Nevertheless, the author does not go further in this discussion. This view is also discussed by Brito (2003, pp. 323-432), but in an unsystematic way. This is due to the fact that the authors do not dedicate a special topic to analyse the mobility aspect of adverbials. The authors explain some mobility issues throughout the section, and for this reason it is hard to infer clear approaches to deal with the adverbs mobility phenomena.

A similar discussion is presented by Costa (2008, pp. 77-109). However, the author provides a clearer approach to be used in order to treat the mobility of adverbial structures. His view is based on the idea that sentence-external modifiers



usually appear in positions before the verb (see examples in (2.26)), while sentence-internal modifiers normally appear in positions after the verb (see examples in (2.27)), as it can be seen in the following examples<sup>4</sup>:

(2.26) (a) ***Felizmente***, o João leu o livro

Fortunately, John read the book

(b) O João, ***felizmente***, leu o livro

John fortunately read the book

(c) \*O João leu ***felizmente*** o livro

John read fortunately the book

(d) \*O João leu o livro ***felizmente***

John read the book fortunately

(2.27) (a) °***Rapidamente***, o João leu o livro

Quickly, John read the book

(b) °O João, ***rapidamente***, leu o livro

John quickly read the book

(c) O João leu ***rapidamente*** o livro

John read quickly the book

(d) O João leu o livro ***rapidamente***

John read the book quickly

Although, this criterion seems to be useful, Costa (2008) says that, in the items (c) and (d) of example 2.26, if commas are used, the sentences become grammatical (example 2.28):

(2.28) (a) O João leu, ***felizmente***, o livro

John read fortunately the book

(b) O João leu o livro, ***felizmente***

John read the book fortunately

---

<sup>4</sup>The examples of (2.26) were taken from Costa (2008, p. 78).

Costa (2008) also considers the sentences (2.27.a) and (2.27.b) as being unacceptable and ungrammatical. In fact, both sentences are grammatical and acceptable, but their meaning differ from the meaning stated in examples (2.27.c) and (2.27.d).

This analysis covers many cases of the mobility phenomena of adverbs. However, Costa (2008, pp. 77-109) considers more exceptions rather than useful common characteristics shared by adverbs to discuss their mobility throughout his work. In this way, it is difficult to establish a systematic and concise way of identifying the mobility of adverbials and also to reproduce the criteria proposed in other examples.

For this reason, the approach used by Molinier and Levrier (2000, pp. 31-34) seems to be more coherent and concise, as it will be presented in the following paragraphs. The authors present the mobility as an important property shared by adverbial structures and base their view on Gross (1986).

For Gross (1986, p. 17), the mobility property of adverbs is an important criterion to distinguish adverbs from essential arguments (complements), since adverbial structures can vary their position in sentences. Molinier and Levrier (2000, pp. 31-34) present the following most common positions that adverbials can occupy:

1. **Initial position (IP):** the adverbial comes before any other sentence element and it is followed by a comma

(2.29) *Provavelmente, o Pedro não comprou a saia para a Maria*

Probably, Peter did not buy the skirt to Mary

2. **Medial position subject-verb (MPSV):** the adverbial comes between the subject and the verb and it comes between commas

(2.30) *O Pedro, provavelmente, não comprou a saia para a Maria*

Peter probably did not buy the skirt to Mary

3. **Medial position after the verb (MPAV):** the adverbial comes immediately after the verb, between the verb and its complement

(2.31) *O Pedro não comprou, **provavelmente**, a saia para a Maria*

Peter did not buy probably the skirt to Mary

4. **End position (EP)**: the adverbial comes after all elements of the sentence usually preceded by a comma

(2.32) *O Pedro não comprou a saia para a Maria, **provavelmente***

Peter did not buy the skirt to Mary probably

According to Gross (1986, pp. 17-18), the mobility that adverbs can present in a sentence is related to the tie or lack of tie that those adverbs have with other structures in the proposition or the entire proposition. For this reason, the sentence-external modifiers, such as *provavelmente* ‘probably’ (see examples (2.29), (2.30), (2.31) and (2.32)), usually have a high mobility because they modify the entire sentence. Their detached position is made clear by the use of commas that separate these adverbs from the sentence they modify.

Other adverbials also present a high mobility in the sentence: **time** (*recentemente* ‘recently’) and **viewpoint** (*judicialmente* ‘judicially’) adverbial structures.

On the other hand, adverbial structures that express, for example, **quantification** or **intensification** (*completamente* ‘completely’) are not so mobile as the previous ones and tend to appear only next to the element they operate on (Molinier and Levrier, 2000, pp. 31-34). **place** (*aqui* ‘here’) adverbial structures also do not present a high mobility in sentences. If **place** adverbials do not appear in the basic position of adjuncts - at the end of the sentence without comma - they can appear between commas, before the verb.

Moreover, Gross (1986, pp. 17-18) highlights the fact that the high mobility of adverbials sometimes might influence their meaning. Consider, for example, the *Adv-mente naturalmente* ‘naturally’. Depending on its position in the sentence, it can be a sentence-external or a sentence-internal modifier.

(2.33) (a) ***Naturalmente**, o João conversou com a sua namorada sobre a viagem*

Naturally, John talked to his girlfriend about the trip

- (b) *O João conversou **naturalmente** com a sua namorada sobre a viagem*  
John talked naturally to his girlfriend about the trip

In example (2.33.a), the adverb *naturalmente* modifies the whole sentence and expresses the speaker's attitude in relation to the truth value of what it is being said (proposition). On the contrary, in (2.33.b), *naturalmente* can be defined as a manner adverb integrated into the proposition, because it is modifying the verb *conversar* 'to talk'. Furthermore, it is subject-oriented because it can be related to the adjective *natural* and it also can be paraphrased by other structures (e.g.: *O João conversou [...] de forma natural* 'John talked [...] in a natural way'), which is not the case in (2.33.a).

The following example illustrates that if the adverb *naturalmente* of the example (2.33.b) was used between commas, its scope would change and the adverb would then modify the whole sentence:

- (2.34) *O João conversou, **naturalmente**, com a sua namorada sobre a viagem*  
John talked, naturally, to his girlfriend about the trip

In view of the previous examples, it is possible to conclude that the mobility of adverbial structures depends on the relation that they establish with their scope. When adverbial structures modify the verb or another constituent inside the sentence, those adverbials presents more restrictions on mobility. Hence, it is necessary to analyse the internal organization of the sentence and also the semantic compatibilities between the adverbial structure and the constituent.

### 2.1.2.3 Relation with Interrogative Pro-Forms

In Bechara (2001, pp. 293-294) and Cunha and Cintra (1984, pp. 539-540), the following interrogative pro-forms are presented: *onde?* 'where?' (**place**), *como?* 'how?' (**manner**), *quando?* 'when?' (**time**), *por que?* and *por quê?* 'why' (**cause**). According to the authors, these pro-forms are interrogative adverbs with a pronominal base. They are used in direct (2.35) and indirect (2.36) interrogative sentences to express one of the meanings previously presented in parenthesis. For example<sup>5</sup>:

---

<sup>5</sup>The example was taken from Cunha and Cintra (1984, p. 539).

(2.35) *Onde está o livro?* (**place**)

Where is the book?

(2.36) *Ignoro onde está o livro* (**place**)

I do not know where the book is

For Molinier and Levrier (2000), these interrogative pro-forms<sup>6</sup> can be used to determine if other adverbial structures (derived, non-derived and compound adverbs) are linked to the verb, so that these adverbial structures can be classified in different syntactic-semantic classes (**place**, **manner** and **time**).

**Place** adverbial structures answer questions introduced by the interrogative adverbs (introduced or not by prepositions), i.e. *onde* ‘where?’ and so forth:

(2.37) (a) *Onde você foi?*

Where did you go?

(b) *Eu fui ao Brasil*

I went to Brazil

**Time** adverbial structures answer questions that have interrogative adverbs like (introduced or not by prepositions, i.e. *até quando* ‘until when?’) *quando?* ‘when?’, *quanto tempo?* ‘how long?’, *quantas vezes?* ‘how many times?’; they correspond to **date** (2.38.a) and (2.38.b), **duration** (2.39.a) and (2.39.b) and **frequency** (2.40.a) and (2.40.b) adverbials, respectively:

(2.38) (a) *Quando você foi ao Brasil?*

When did you go to Brazil?

(a) *Eu fui ao Brasil no mês passado*

I went to Brazil last month

(2.39) (a) *Quanto tempo você leva para terminar o seu dever de casa ?*

How long do you take to finish your homework?

---

<sup>6</sup>The pro-form *por que?* ‘why’ (**cause**) is not considered by the authors, because it cannot be associated to adverbial nominal groups (Molinier and Levrier, 2000, p. 34).

- (a) *Eu levo **três horas** para terminar o meu dever de casa*  
I take three hours to finish my homework

- (2.40) (a) ***Quantas vezes** você leu esse livro?*  
How many times have you read that book?

- (a) *Eu o li umas **dez vezes***  
I have read it about ten times

**Manner** adverbial structures answer questions that have the interrogative adverb *como?* ‘how?’ and *de que forma/ modo/ maneira?* ‘in which way/ manner?’ (2.41.a) and (2.41.b):

- (2.41) (a) ***Como/ De que maneira** a Maria comeu a salada?*  
How/ In which manner did Mary eat the salad?  
(b) *A Maria comeu a salada **lentamente***  
Mary ate the salad slowly

There are other interrogative pro-forms (Molinier and Levrier, 2000, pp. 34-36) that will not be discussed here, since they follow the same logic, but they express other syntactic-semantic classes. Although this criterion may be useful in many cases, Gross (1986) highlights the fact that it is not always possible to use this criteria, because some adverbs cannot be related to interrogative pro-forms. For example<sup>7</sup>:

- (2.42) (a) ***Como/quantas vezes/quando?** o Pedro desculpou a Joana?*  
How/ How many times/ When did Peter forgive Joana?  
(b) *\*O Pedro desculpou a Joana **por esta vez***  
Peter forgave Joana for this time/ for the time being

The meaning of *por esta vez*, while time-related, cannot be entirely covered by the interrogative pro-form, hence it cannot be a good answer to such question.

In addition, there are cases in which the same adverb can be associated to more than one interrogative pro-form:

---

<sup>7</sup>This example was taken from Palma (2009).

(2.43) (a) *Como/ Em quanto tempo? a Maria comeu a comida?*

How/ How much time did Mary eat the food?

(b) *A Maria comeu a comida rapidamente*

Mary ate the food quickly

In this case (2.43.a) and (2.43.b), the meaning of the **manner** adverbs *rapidamente* is also time-related, hence there is an ambiguous status regarding the two pro-forms.

### 2.1.3 Classifications of Adverbial Structures

Traditionally, adverbial structures are classified according to their meaning, in other words, a semantic approach is adopted.

In the grammars of Bechara (2001, pp. 290-292) and Cunha and Cintra (1984, pp. 538-539), seven classes are established: **affirmation**, **doubt**, **intensity**, **space**, **manner**, **negation** and **time**. There are also three classes (**order**, **exclusion**, **designation**) that are treated apart, as it is said that they do not modify verbs, adjectives or other adverbs. The semantic aspect alone may not be sufficient to develop a systemic and clear classification. For this reason, grammarians' classification are poor, because it does not take into consideration other important aspects of the adverbial structures, such as the scope which the adverbs operate on, their mobility position and their relation with interrogative pro-forms (see Section 2.1.2).

In Brito (2003, pp. 418-419), the author presents the same seven semantic classes presented in Bechara (2001, pp. 290-292) and Cunha and Cintra (1984, pp. 538-539). However, the author criticises this traditional view of classifying adverbs based only on semantic criteria. Hence, the author proposes some additional morphological (e.g. degree of adverbs) and syntactic (e.g. transitivity, scope and paraphrases) criteria that could be useful in the task of classifying adverbs. Nevertheless, Brito (2003) only presents and discusses different properties. The author does not propose any systematic way of using these properties to establish a more in-depth classification of the adverbs, which would not only consider the semantic aspects of adverbs as it happens in traditional grammars. As a result,

it seems unfeasible to apply randomly the properties proposed by Brito (2003) in order to obtain a thorough classification .

Costa (2008, pp. 41-75) also presents the classification of adverbial structures based on their meaning. It is interesting to note that the author does not base his classification on the traditional semantic criteria. He uses some formal properties (e.g. the use of cleft sentences and the extraction of the adverb in a negative sentence) to distinguish, according to his view, two different semantic classes of adverbial structures: (1) adverbs that operate on the entire sentence and (2) adverbs that operate on verbs (see Section 2.1.2.1). In addition, he uses semantic and other syntactic criteria to define subclasses inside these two main classes. However, Costa (2008) considers that there are types of adverbials (e.g. **quantity** adverbs) that do not belong to any of these two groups, since they establish relations with other types of constituents such as noun phrases. For this reason some adverbs are treated apart (Costa, 2008, pp. 62-75). A systematic classification should cover more cases than exceptions, and this does not seem to be considered by Costa (2008).

Molinier and Levrier (2000, pp. 44-54) also use formal features to propose a classification of *-ment* ending adverbs. The scope feature is also considered as an important criterion to distinguish different types of adverbs. In the case of Molinier and Levrier (2000), this feature is used firstly to classify all adverbs into two main groups: (see Section 2.1.2.1). From this first generic classification, other features (e.g. meaning, paraphrases, mobility and relation with pro-forms) are used to define subclasses. Based on this approach, the authors intend to cover all possible types of adverbials instead of dealing with exceptions as in Costa (2008) or random properties as in Brito (2003).

In conclusion, neither the traditional approach nor the one proposed by Costa (2008) seems to succeed in providing an in-depth classification. On the one hand, the traditional approach such as in Cunha and Cintra (1984) and Bechara (2001) provides a feeble classification only based on the semantic criterion or even do not propose any formal classification as in Brito (2003), only discussing random properties. On the other hand, in Costa (2008), the criteria proposed cover only a part of the adverbials and others are treated as exceptions. Due to this fact, a more thorough and efficient classification proposed by Molinier and Levrier (2000),



in other words, that cover all cases of adverbials based on clear and systematic criteria has been adopted in the present study and will be briefly presented in the next section.

### 2.1.3.1 A Lexicon-Grammar Classification

Molinier and Levrier (2000) have proposed a systematic syntactic-semantic approach to classify the French adverbs ending in *-ment*. Their classification follows the theoretical and methodological principles of the Lexicon-Grammar, developed by Gross (1986) based on the Transformational Operator Grammar of Harris (1976, 1991). In spite of its original application to French, it can also be adapted to Portuguese adverbs, providing that some peculiarities of this language are taken into consideration. In fact, the classification can be extended to other adverbial structures as Molinier and Levrier (2000) abundantly illustrate and Palma (2009) suggests, namely to compound adverbs.

As already mentioned in section 2.1.2.1, according to Molinier and Levrier (2000), *Adv-mente* are classified into two main classes: (a) sentence-external modifiers and (b) sentence-internal modifiers. In order to determine if an adverbial belongs to class (a) or (b), Molinier and Levrier (2000) proposes two general syntactic properties, (i) and (ii):

- (i) detachment of adverbials to the initial position of a negative sentence
- (ii) extraction of adverbials in cleft sentences

Consider the two following examples, typical of each class:

(2.44) ***Aparentemente**, o João comeu os legumes*

Apparently, John did ate the vegetables

(2.45) *O João falou comigo **calmamente***

John did speak with me calmly

Based on the two properties previously mentioned, *sentence-external adverbials* such as in (2.44) (noted *P*) present the following behavior:

(a) they *can* be detached to the initial position of a negative sentence

(2.46) ***Aparentemente**, o João **não** comeu os legumes*

Apparently, John did not eat the vegetables

and,

(b) they *cannot* be extracted in cleft sentences

(2.47) \****Foi aparentemente** que o João comeu os legumes*

It was apparently that John ate the vegetables

*Sentence-external adverbs* must present the behavior (a) and (b) *simultaneously* to be classified as *P*.

On the other hand, the behavior of *sentence-internal adverbials* (noted *M*) may vary regarding the properties (i) and (ii). Both properties, (i) and (ii), are not presented by all sentence-internal adverbials simultaneously. Many *sentence-internal adverbials* such as in (2.45) typically present the following behavior:

(c) they *cannot* be detached to the initial position of a negative sentence

(2.48) \****Calmamente**, o João **não** falou comigo*

Calmly, John did speak with me

and,

(d) they *can* be extracted in cleft sentences

(2.49) \****Foi calmamente que** o João falou comigo*

It was calmly that John spoke with me

The *Adv-mente calmamente* ‘calmly’ in examples (2.48) and (2.49) present the typical behavior of sentence-internal modifiers.

Nevertheless, there are *sentence-internal adverbs*, i.e. adverbs that have a strong relation with verbs such as:

*O João assediou **sexualmente** a Maria*

John harassed sexually Mary

which, because of this particular relation manifest a less typical syntactic behavior, namely:

- (e) they *cannot* be detached to the initial position of a negative sentence

(2.50) \***Sexualmente**, o João **não** assediou a Maria

Sexually, John did not harass Mary

but like *sentence-external adverbials (P)*

- (f) they *cannot* be extracted in cleft sentences

(2.51) \***Foi sexualmente que** o João assediou a Maria

It was sexually that John assaulted Mary

Finally, other *sentence-internal adverbials* present the opposite behavior:

*O João trabalha **atualmente** na fábrica*

John works currently at the factory

- (g) they *can* be detached to the initial position of a negative sentence

(2.52) **Atualmente**, o João **não** trabalha na fábrica

Currently, John works at the factory

- (h) they *can* be extracted in cleft sentences

(2.53) **É atualmente que** o João trabalha na fábrica

It is currently that John works at the factory

Adverbials of **time** and **viewpoint** usually present the behaviors (g) and (h), which have been just described.

As the behavior of *sentence-internal adverbials* may vary regarding properties (i) and (ii), some of them may be confused with sentence-external adverbs. Hence, other properties may also be taken into consideration.

Another property proposed by Molinier and Levrier (2000) is the possibility/impossibility of paraphrasing the *Adv-mente*. This paraphrase can be made:

- (a) by way of an equivalent, morphologically related adjective modifying operator-nouns (Gross, 1986, pp. 11-13):

**‘de (modo+maneira+forma+jeito<sup>8</sup>) Adj’**

- (2.54) (a) *O Pedro comportou-se **naturalmente** durante o jantar*  
Peter behaved naturally during the dinner  
(b) *O Pedro comportou-se **de maneira natural** durante o jantar*  
Peter behaved in a natural way during the dinner

- (b) by way of a prepositional phrase, often syntactically constraint, involving an equivalent and morphologically related noun:

**‘Prep (Det) N Adj’**

- (2.55) (a) *O Pedro ordenou os livros **alfabeticamente***  
Peter ordered the books alphabetically  
(b) *O Pedro ordenou os livros **por ordem alfabética***  
Peter ordered the books alphabetically in alphabetical order

This property can be quite useful when dealing with some *Adv-mente*, but it is not always possible to make a paraphrase. For example, this feature is highly productive in the case of the subject-oriented manner adverbs (MS)

- Adjectival Paraphrase: **‘de (forma+maneira+modo) Adj’**

- (2.56) (a) *O Pedro leu **atentamente** a revista*  
Peter read attentively the magazine  
(a) *O Pedro leu a revista **de forma atenta***  
Peter read the magazine in an attentive way

However, in the case of the quantification adverbs (MQ), sometimes it is not possible to paraphrase *de Adv-mente*:

---

<sup>8</sup>The operator-noun *jeito* is only used in Brazilian Portuguese.

- Adjectival Paraphrase: ‘**de (forma+maneira+modo) Adj**’

(2.57) (a) *O Pedro está **perdidamente** apaixonado pela Maria*

Pedro is hopelessly in love with Mary

(b) \**O Pedro está apaixonado **de forma perdida** pela Maria*

Pedro is in love in a hopeless way with Mary

There are other two types of adjectival paraphrases pointed out by Molinier and Levrier (2000):

- Adv, Sentence = ‘**Que Sentence ser Adj**’: equivalence of an adverb construction to an adjectival, sentential subject construction. This property is usually observed with sentence-modifier disjunctive adverbs of attitude (PA)

(2.58) (a) ***Supreendentemente**, o Pedro não veio*

Surprisingly, Peter did not come

(b) *Que o Pedro não tenha vindo é surpreendente/ É surpreendente que o Pedro não tenha vindo*<sup>9</sup>

That Peter did not come it is surprising/ It is surprising that Peter did not come

- ‘**de um ponto de vista Adj**’: equivalence of the adverb to a paraphrase with *ponto de vista* ‘viewpoint’. This property is crucial for all **viewpoint** adverbs (MP)

(2.59) (a) ***Moralmente**, essa atitude não é aceitável*

Morally, this attitude is not acceptable

(b) ***De um ponto de vista moral**, essa atitude não é aceitável*

From a moral point of view, this attitude is not acceptable

Regarding the nominal paraphrases, there are two types:

1. ‘**com N**’: *prudently* ‘prudently’ = *com prudência* ‘with prudence’

---

<sup>9</sup>It is equivalent to the previous sentence. Only the order was changed.

2. ‘Prep N’: *manualmente* ‘manually’ = *à mão* ‘by hand’

The first nominal (1) paraphrase is common with deadjectival, predicative nouns (*prudência, prudente, prudentemente*), while the second (2) involves non-predicative nouns (e.g. *mão*) and is related to *Adv-mente* derived from relational adjectives (e.g. *manual/ manualmente*). Naturally, these paraphrases are subject to many morphological and historical idiosyncrasies that will not be discussed in this study.

Sentence types (i.e. declarative, interrogative, imperative, exclamatory) may also be a useful property, especially in the case of *sentence-external adverbials*, since there is a restriction between sentences types and adverbials classes. For instance, *sentence-external adverbials* that express **modality**, **evaluation** or that are subject-oriented appear only in declarative sentences and they cannot be used in interrogative sentences:

- (2.60) (a) **Felizmente**, a Maria passou no exame de direção  
Fortunately, Mary passed her driving test  
(b) \***Felizmente**, a Maria passou no exame de direção?  
Fortunately, Mary passed her driving test?

On the contrary, sentence-internal adverbials of time can be used in interrogative sentences:

- (2.61) (a) O João depende **financeiramente** do pai  
John depends financially on his father  
(b) **Financeiramente**, o João depende do pai?  
Financially, does John depend on his father?

Based on these and other properties Molinier and Levrier (2000) have determined and described specific subclasses for each one of the two main class of adverbs. The authors describe many other features, mainly syntactic and semantic, inside of the subclasses in order to define clearly each one of them. In the present study, only the main characteristics discussed by the authors are highlighted, placing greater emphasis on the syntactic features.

The *sentence-external modifier class* (*P*) is divided in three main subclasses: conjunctives adverbs (PC), disjunctives adverbs of style (PS), disjunctives adverbs of attitude (PA).

### 1. PC

The main characteristic of these adverbs is that they cannot occur at the absolute beginning of the discourse.

(2.62) *O Pedro ficou doente e, **consequentemente**, não pode viajar*

Peter got sick and, consequently, he could not travel

### 2. PS

The adverbs of this subclass have a relation with metalinguistic operators and they also have their scope on those operators. Consider the following examples:

(2.63) ***Honestamente**, isso será um problema*

Honestly, this will be a problem

= (it is equivalent to the following paraphrase)

(2.64) *Eu considero **honestamente** que isso será um problema*

I honestly believe that this will be a problem

### 3. PA

The disjunctive adverb of attitude class (PA) is closely related to the PS class. For this reason, Molinier and Levrier (2000) explain that the PA adverbs are complementary to PS adverbs.

This subclass can be divided into four categories: evaluative adverbs (PAa), adverbs of habit (PAh), modal adverbs (PAm) and subject-oriented attitude adverbs (PAs).

(PAa) the speaker expresses an opinion about or his attitude towards the content of the proposition

(2.65) *Felizmente, o Pedro não gastou muito dinheiro*

Fortunately, Peter did not spend a lot of money

(PAh) it accords an habitual aspect to the process in the proposition, because of this they always present verbs in the present or imperfective tense with habitual value

(2.66) *Habitualmente, o Pedro dorme à meia-noite*

Habitually, Peter sleeps at midnight

(PAm) it attributes a modal value to the proposition; different modalities may be involved

(2.67) *Aparentemente, o Pedro comeu muito*

Apparently, Peter ate a lot

(PAs) they present the typical properties of sentence-external adverbs but with a scope on the subject

(2.68) *Inteligentemente, o Pedro foi embora*

Intelligently, Peter went away

In order to determine a PAs construction, an adjectival paraphrase can be used, such as:

(2.69) *O Pedro foi inteligente em ir embora*

Peter was intelligent because he went away

Taking into consideration the *sentence-internal modifiers*, they can be divided into six main subclasses: subject-oriented manner adverbs (MS), proper manner adverbs (MV), adverbs of quantification (MQ), viewpoint adverbs (MP), adverbs of time (MT) and focus adverbs (MF).

### 1. MS

The subject-oriented manner adverbs (MS) have a relation with base adjectives. Consider the following example:

(2.70) *O Pedro leu atentamente a revista*

Peter read attentively the magazine



This sentence can be split into two sentences:

(2.71) *O Pedro leu a revista # O Pedro estava atento*

Peter read the magazine # Peter was attentive

Note that by splitting the main sentence, the adverb becomes an adjective (*atento* ‘attentive’) that establish a relation with the subject of the sentence.

In addition, an adjectival (2.72) or nominal (2.73) paraphrase can be used in the place of MS *Adv-mente*

(2.72) *O Pedro leu a revista de forma atenta*

Peter read the magazine in an attentive way

(2.73) *O Pedro leu a revista com atenção*

Peter read the magazine \*with attention<sup>10</sup>

## 2. MV

The manner adverbs that modify verbs (MV) usually present the following behavior:

*O Pedro ordenou os livros alfabeticamente*

Peter ordered the books alphabetically

(2.74) \***Alfabeticamente**, o Pedro **não** ordenou os livros

Alphabetically, Peter did not ordered the books

(2.75) **Foi alfabeticamente que** o Pedro ordenou os livros

It was alphabetically that Peter ordered the books

MV adverbs cannot be extracted in negative sentences (2.74), because they have a strong relation with the verb in the sentence.

The MV adverbs do not have their scope on the subject, and this is the difference between them and the MS adverbs. Thus, it is not possible to use an

---

<sup>10</sup>This is a literal translation. It is not possible to use this construction in English.

adjectival paraphrase with MV adverbs (i.e. *\*O Pedro era alfabético*/ ‘Peter was alphabetical’). However, it is possible to use an appropriate nominal paraphrase started by a preposition, for instance:

(2.76) *O Pedro ordenou os livros **por ordem alfabética***

Peter ordered the books in alphabetical order

The MV class is a large class of adverbs and these adverbs usually appear in positions after the verb, without commas. Moreover, they allow the use of prepositional paraphrases (e.g. *alfabeticamente* ‘alphabetically’/ **por ordem alfabética** ‘in alphabetical order’).

Finally, most MV adverbs can answer the question *como?* ‘how?’.

### 3. MQ

Adverbs of quantification (MQ) cannot be extracted in negative sentences (2.77) and (2.79). Depending on the MQ adverb they may (2.78) or may not (2.80) be detached to the initial position in cleft sentences.

*O Pedro ama **profundamente** a Maria*

Peter is deeply in love with Mary

(2.77) *\***Profundamente**, o Pedro não ama a Maria*

Deeply, Peter is not in love with Mary

(2.78) *É **profundamente** que o Pedro ama a Maria*

It is deeply that Peter is in love with Mary

*A geladeira está **praticamente** vazia*

The fridge is practically empty

(2.79) *\***Praticamente**, a geladeira não está vazia*

Practically, the fridge is not empty

(2.80) *\*É **praticamente** que a geladeira está vazia*

It is practically that the fridge is empty

MQ adverbs function as adverbial determiners and quantifiers. In general, they answer to the question *quanto?* ‘how many/how much’.

*O Pedro está **perdidamente** apaixonado pela Maria*

Peter is madly in love with Maria

(2.81) *O quanto o Pedro está apaixonado pela Maria*

How much is Peter in love with Maria?

(2.82) *perdidamente ou muito*

madly or extremely

#### 4. MP

The viewpoint adverbs (MP) usually can be extracted in negative sentences (2.83) and also can be detached to the initial position in cleft sentences (2.84). This last property is the reason why MP *Adv-mente* are classified as sentence-internal modifiers.

*O Pedro é **politicamente** ingênuo*

Peter is politically naive

(2.83) ***Politicamente**, o Pedro **não** é ingênuo*

Politically, Peter is not naive

(2.84) ***É** **politicamente** **que** o Pedro é ingênuo*

It is politically that Peter is naive

MP adverbs have a morphological property: they establish a relation with relational adjectives (i.e. *politicamente* ‘politically’ and *político* ‘politic’). It is possible to paraphrase the MP *Adv-mente* with ‘*de um ponto de vista Adj*’ ‘from a point of view Adj’ (Molinier and Levrier, 2000, pp. 41-42):

(2.85) *O Pedro é ingênuo **de um ponto de vista político***

Peter is naive from a political point of view

5. MT

The adverbs of time (MT) answer questions such as *quando?* ‘when?’ (2.86), *quanto tempo?* ‘how long?’ (2.87) and *com que frequência?* ‘how often?’ (2.88):

(2.86) (a) *O Pedro foi à escola ontem*

Peter went to the school yesterday

(b) **Quando** *o Pedro foi à escola?* ‘When did Peter go to the school?’

R: *ontem* ‘yesterday’

(2.87) (a) *O Pedro leva uma hora para tomar banho*

Peter takes one hour to take a shower

(b) **Quanto tempo** *o Pedro leva para tomar banho?* ‘How long does Peter take to take a shower?’

R: *uma hora* ‘one hour’

(2.88) (a) *O Pedro lê o jornal diariamente*

Peter reads the newspaper daily

(b) **Com que frequência** *o Pedro lê o jornal?* ‘How often does Peter read the newspaper?’

R: *diariamente* ‘daily’

Finally, MT adverbs are different from sentence-external adverbs of habit (PAh), because the former are integrated into the proposition, whereas the latter modify the whole sentence. Unlike the PAh adverbs, the MT present the following behavior:

*O Pedro viaja para Brasília mensualmente*

Peter travels to Brasília monthly

(2.89) **Mensualmente**, *o Pedro não viaja para Brasília*

Monthly, Peter does not travel to Brasília

(2.90) *É mensualmente que o Pedro viaja para Brasília*

It is monthly that Peter travels to Brasília

The possibility of being detached to the initial position in cleft sentences (2.90) is a crucial characteristic to distinguish MT from PAh adverbs.

## 6. MF

One of the main characteristics of focus adverbs (MF) is that they determine a noun phrase:

(2.91) *O Pedro escreve principalmente poesia*

Peter writes principally poetry

They also present the following behaviors:

- they *cannot* be detached to the initial position of a negative sentence

(2.92) \***Principalmente**, o Pedro **não** escreve poesia

Principally, Peter do not write poetry

and,

- most of them *cannot* be extracted in cleft sentences

(2.93) \***É principalmente que** o Pedro escreve poesia

It is principally that Peter writes poetry

Note that if the noun phrase is extracted together with the adverb in a cleft sentence then the sentence becomes acceptable:

(2.94) *É principalmente poetry que o Pedro escreve*

It is principally poetry that Peter writes

This indicates that the adverb has a strong relation with the noun phrase.

This section presented a brief overview of the classification of adverbs, that was firstly made for the French adverbs ending in *-ment*, proposed by Molinier and Levrier (2000). This classification can also be used to analyse the Portuguese

adverbs, as it was illustrated by the examples here given. However, it is important to take into consideration the peculiarities of Portuguese and the different constructions that adverbs may present in this language, in order to propose a systematic syntactic-semantic classification these adverbs.

Table 2.1 summarizes the classification presented in this section:

Class	Definition	Example
PC	Conjunctive adverbs	<i>consequentemente</i>
PS	Disjunctive adverbs of style	<i>honestamente</i>
PAa	Disjunctive adverbs of attitude: evaluation	<i>felizmente</i>
PAh	Disjunctive adverbs of attitude: habit	<i>habitualmente</i>
PAm	Disjunctive adverbs of attitude: modal	<i>aparentemente</i>
PAs	Disjunctive adverbs of attitude: subject-oriented attitude	<i>inteligentemente</i>
MS	Subject-oriented manner adverbs	<i>atentamente</i>
MV	Proper manner adverbs	<i>alfabeticamente</i>
MQ	Quantification adverbs	<i>demasiadamente</i>
MP	Viewpoint adverbs	<i>politicamente</i>
MT	Time adverbs	<i>diariamente</i>
MF	Focus adverbs	<i>principalmente</i>

Table 2.1: Classification of adverbs proposed by Molinier and Levrier (2000)

The next section (2.2) will present some research on adverbial structures undertaken in the NLP field.

## 2.2 Adverbial Structures in NLP

In the last decades, the NLP field has grown and many studies have been carried out to improve NLP systems. However, adverbs still do not receive much attention in this field. Small research has been undertaken regarding the treatment of Portuguese adverbial structures, in particular. In this section some researches on adverbs are presented.

A significant breakthrough in the computational treatment of adverbs can be noted, if some old and recent researches on varied topics about the adverbs that were undertaken in different languages are considered.

A research conducted by Defrise (1989) on French adverb *presque* (‘almost’) is a clear example that shows why it is important to study the linguistic features of adverbs in order to process them computationally. The author argues that, at that time, most NLP systems did not have enough information of word classes such as the adverbs. For this reason, these systems could not properly process

(e.g. analysis, disambiguation and correct lexical choices) words that belonged to these classes.

Conlon and Evens (1992) argue in favor of the development of linguistics studies on English adverbs to be used for parsers, language understanding and information retrieval systems. This study reveals the need to establish a systematic and controlled approach to deal with adverbial structures computationally as

“it is very difficult for Natural Language Processing (NLP) systems to handle adverbs because of the large number of syntactic roles that adverbs can assume in sentences” (Conlon and Evens, 1992, p. 1192)

The authors also cite other properties that might be related to the difficulty in treating the adverbs, such as the different scopes of adverbs, the peculiar positional and selectional restrictions, and the fact that some adverbs can change their meaning when they are used in different positions in a sentence. Basically, they try to establish a few semantic and syntactic features of some types of adverbs (e.g. manner adverbs) that could be helpful in their treatment. To achieve this, they use the information of some adverbs taken from an English dictionary to establish useful features. For instance, they show that information about the scope of adverbs (the relation of adverbs with constituents in a sentence) could be used to treat lexical ambiguity in language understanding systems and could also help word choice in text generation. In sum, the aim of this study is to motivate the analysis of adverbs to be applied in different NLP systems and highlight its importance rather than providing a robust approach to handle the adverbs.

Since then, many theories and systems have been proposed and developed, allowing that more computational linguistic studies could be conducted. Nimb (2002), for example, discusses how to build an ontology on Danish temporal adverbs and which information regarding these adverbs should be included in a computational lexicon. Another aim is to propose which characteristics previously studied can be incorporated into a lexicon model called SIMPLE (Lenci et al., 2000).

Català and Baptista (2007a) discuss the limitations of pattern-matching finite-state methods when applied to lexical analysis of multiword expressions. In order to illustrate their view, the authors describe and apply these techniques to Spanish

compound adverbs. In another research Català and Baptista (2007b) analyse the Spanish adverbial frozen expressions. In the last work, the authors' aim is to provide a linguistic description of these expressions so that it can be computationally adopted.

Zan et al. (2010) try to recognize automatically Modern Chinese adverbs' usage based on statistical approaches. They opted for this approach, because many problems were observed in previous works, in which a rule-based method was applied to handle the case of recognition of Chinese adverbs' usage. The authors conclude that better results can be achieved if the statistical methodology is applied, rather than the rule-based one.

Comparative studies on Portuguese and Spanish adverbs were performed regarding different topics. Baptista and Català (2002) discusses the problems of processing Portuguese and Spanish compound temporal adverbs. Their purpose is to provide an automatically translation of these compound temporal adverbs from one language into the other based on a rule-based approach.

Baptista and Català (2009a,b) attempt to disambiguate Portuguese and Spanish *-mente* ending focus adverbs (MF) adopting a rule-based approach in order to improve the performance of the Portuguese Grammar (Mamede et al., 2007) developed for the Xerox Incremental Parser - XIP (Ait-Mokhtar et al., 2002), in particular the chunking rules. This rule-based approach tries to identify the different constructions that *-mente* ending focus adverbs may have besides the MF constructions analysed by Molinier and Levrier (2000) and their respective meanings. In addition, finite-state tools are used to capture distributional patterns for the disambiguation of MF adverbs. The authors conclude that there is still much research to be done in order to achieve better results.

Some NLP researches that focus specifically on Portuguese adverbs were carried out in the past few years.

Bick (2000) and Afonso (2002), for example, did researches on how adverbs and others words related to them should be treated by parsers. In its presentation of the parsing system PALAVRAS for Portuguese, Bick (2000) dedicates a small section to describe how adverbs are dealt by the parser. He explains that adverbials are very heterogeneous regarding rules such as word order, context patterns and lexical potential. For this reason, the author affirms that in order to apply these rules in



his parser, it was necessary to analyse carefully many features (e.g. syntactic and semantic). A detailed description of these features can be found in his work.

Afonso (2002) discusses how to identify and treat automatically the first element in a coordination of *Adv-mente*, that in Portuguese is characterised by the reduction of the suffix *-mente* in the first element (e.g. *científica e civicamente*). This is a problem of rule-based approach to part-of-speech disambiguation. One of the arguments presented is that the search and analysis of relevant syntactic-semantic patterns in corpora may help to determine how to define the first element of the *Adv-mente* coordination (adjective or adverb). Finally, this study aims to include, in the parser PALAVRAS, the features identified in order to improve and evaluate its performance regarding this coordination phenomena.

Finally, some researches on Portuguese temporal adverbs have been done. The works of Baptista et al. (2008), Hagège et al. (2009) and Hagège et al. (2010) are about Portuguese time expressions, including adverbial structures of time. In these works, the authors discuss how to identify, classify and normalize these times expressions with the purpose of implementing this information in NLP systems. Their final aim is to provide a proper automatic recognition of these expressions in texts, so that systems may be able to associate them with events and states of things that they modify.

Based on the works previously presented, it is possible to conclude that adverbial structures are not easy to be computationally processed. There are still many linguistic features concerning these structures that must be carefully studied in order to be implemented in NLP systems, and consequently to improve their performance.

## 2.3 A Brief Overview of WSD

In the previous sections, linguistic and computational works regarding the adverbs, mainly *-mente* ending adverbs, were discussed. Now a brief review of what is Word Sense Disambiguation, its importance, applications and main characteristics will be made in order to situate this NLP topic in the context of the present dissertation.

Word Sense Disambiguation (WSD) is a NLP task that consists in analysing

and determining the sense of words in context. It is an important task that must be taking into consideration as many words have multiple meanings and not always it is easy to determine the nature of many word senses (Jurafsky and Martin, 2009, p. 671).

Jurafsky and Martin (2009) affirms that Weaver (1955) was one of the first persons who introduced the discussion of the need for disambiguating words, in the context of machine translation. Weaver (1955) suggested that a word could be disambiguated “by looking at a small window around it” (Jurafsky and Martin, 2009, p. 710). Considerable research on WSD has been undertaken ever since and many different approaches to deal with it has been developed in the past forty years.

Specia (2007, pp. 49-124) discusses the approaches applied to WSD based on the classification proposed by Ide and Véronis (1998). The author explains that these approaches can be classified based on the NLP methods they use. As a result, she considers three main methods and many different techniques used in each one of them. These methods and their respective techniques are broadly presented below:

1. **Method based on linguistic and/or extralinguistic knowledge:** disambiguation models (e.g. rules) are manually defined or extracted from online linguistic tools/resources.
  - Manually codified knowledge: knowledge is manually extracted
    - Symbolic approach: e.g. semantic networks techniques
    - Connectionist approach: e.g. semantic priming techniques
  - Pre-codified knowledge: knowledge is extracted from online lexical resources (e.g. electronic dictionaries and thesauri)
2. **Empiric Method based on corpus:** disambiguation models are obtained from the superficial knowledge based on information from the corpus and also from statistical or machine learning techniques.
  - Supervised approach: a hand-labeled corpora is used
  - Minimally Supervised approach: a very small hand-labeled corpora is used
  - Unsupervised approach: a non hand-labeled corpora is used
3. **Hybrid Method:** it combines characteristics of both 1 and 2 previous methods.

For Specia (2007) (based on Sparck-Jones and Galliers (1996)) and Jurafsky and Martin (2009), there are two main ways of evaluating the methods applied to WSD: (a) extrinsic evaluation (in vivo) and (b) intrinsic evaluation (in vitro). The extrinsic evaluation means that the WSD component is evaluated regarding the whole context of the application in which it was used, hence it is more complex and time-consuming to be implemented. In the intrinsic evaluation, the WSD component is evaluated if it were an independent system, and for this reason it is more adopted than the first one (Jurafsky and Martin, 2009, p. 678).

Considering only the intrinsic evaluation, a WSD component can be evaluated for example by means of the measures of *precision*, *recall*, *f-measure* and *accuracy* (Manning and Schütze, 2001).

*Precision* is the proportion of selected items that system identified correctly, whereas *recall* is the proportion of the target items that the system selected. The *f-measure* is a combination of the *precision* and *recall* measures. It indicates the general performance of the system. *Accuracy* means the percentage of words correctly tagged in comparison to the total of ambiguous words in a test set.

The results of these evaluations are usually contrasted with an *upper* or a *lower bound (baseline)* measure in order to better assess the performance of the system. In the case of WSD, the upper bound is usually human performance, e.g. annotator agreement (Manning and Schütze, 2001, p. 233). One of the simplest examples of *baseline* consists in determining the most frequent sense (MFS) for each word (Jurafsky and Martin, 2009, p. 678). As a result, a minimum value is established and the performance of the system can be evaluated in comparison to it.

Finally, WSD can be used also to improve other NLP tasks such as text classification, machine translation, question answering, information retrieval and dialogue systems. Furthermore, it is also used in part of speech tagging or syntactic analysis. An extensive and in-depth review of WSD is beyond the scope of this dissertation. Further reading on this subject can be found in works such as Stevenson (2003), Stevenson and Wilks (2003) and Jurafsky and Martin (2009), among others.

### 2.3.1 The Weka Package

The method that was chosen to deal with *Adv-mente* ambiguity problem in this dissertation was the *empiric method based on corpus*, more specifically the machine learning technique. This technique has been widely applied for handling the WSD problem, as the information to disambiguate words can be extracted automatically from the corpus. More reasons for this choice are explained in Chapter 4.

In order to apply this method and its techniques to the disambiguation experiments<sup>11</sup> of this dissertation, the free workbench Waikato Environment for Knowledge Analysis known as Weka<sup>12</sup> was used and a training dataset based on sentences extracted from a corpus was built (see Section 4.3).

The Weka package is a machine learning workbench developed by Witten and Frank (2005). It has a varied collection of ML algorithms and data preprocessing tools. The interfaces of Weka help the user to have access to all facilities available in this program such as the preparation of the input data, the use of existing methods on new datasets, the visualisation of the result of learning, the statistical evaluation that is provided (e.g. *precision*, *recall* and *f-measure*) of learning schemes, among other facilities.

---

<sup>11</sup>These experiments are detailed explained in Chapter 4.

<sup>12</sup>Weka is available for download at <http://www.cs.waikato.ac.nz/ml/weka/>.

# Chapter 3

## Classification of *Adv-mente*

This chapter presents the corpus used in this study, highlighting the proportion of adverbs in general and *Adv-mente* identified in this corpus (see Section 3.1). The way how the *Adv-mente* were selected and the reason why 532 of them were classified (see Section 3.2) are explained.

Some issues regarding the classification task are explained (see Section 3.3). The results obtained after classifying all the 532 *Adv-mente* are presented and discussed (see Section 3.4).

Finally, a comparison between the 532 *Adv-mente* of the corpus that were classified and the *Adv-mente* found in two Brazilian electronic dictionaries is made. The aim of this comparison is to discuss how electronic dictionaries treat this type of adverbs and verify how many of the 532 *Adv-mente* are treated by both dictionaries (see Section 3.5).

### 3.1 Corpus

The *Adv-mente* that were analysed in this study were extracted from the NILC/São Carlos corpus<sup>1</sup>. This corpus is a collection of Brazilian Portuguese texts of different genres (e.g. didactic, epistolary and short essays written by students), mainly journalist texts. These last texts were extracted from a Brazilian newspaper known as Folha de São Paulo<sup>2</sup>.

---

<sup>1</sup><http://www.linguateca.pt/aceso/corpus.php?corpus=SAOCARLOS>

<sup>2</sup>The online version of this newspaper is available at <http://www.folha.uol.com.br/>

It is one of the corpora available at Linguateca<sup>3</sup> web page and can be downloaded or consulted through an interface developed by the AC/DC project (Santos and Bick, 2000; Santos and Sarmiento, 2002). This corpus was syntactically annotated by PALAVRAS parser that was developed by Bick (2000).

As already mentioned, the NILC/São Carlos corpus contains approximately 32,3M of words and 1,456,426 of these words (approximately 4.5%) are adverbs. In the Linguateca web page, a list of the lemmas is available, indicating all adverbs and their respective frequency in the corpus. According to this list, from the 2,867 different adverbs (lemmas) there included, only the *Adv-mente* were selected. The 1,936 different *Adv-mente* (lemmas) occur 103,638 times in the corpus (7,11% from all adverbs). Table 3.1 summarizes these figures:

	Corpus (total)	Adverbs	Adv-mente (%)
<b>TYPES (lemmas)</b>	397.552	2.867	1.936 (67.52%)
<b>TOKENS (words)</b>	32.342.456	1.456.426	103.638 (7.11%)

Table 3.1: Count of lemmas and tokens of all words, adverbs and *Adv-mente* of the NILC/São Carlos corpus. Source: Linguateca.

## 3.2 Selection of *Adv-mente*

Unfortunately, for reasons of time and scope, it was not possible to describe the linguistic features of all *Adv-mente*, given the number of *Adv-mente* in the NILC/São Carlos corpus. Hence, the distribution of these adverbs has been calculated (see Table 3.2), in order to select those that would be analysed.

Considering the 1.936 *Adv-mente* (LexAdv-mente and %Adv-menteLex) and their respective frequency in the corpus (FrequencyTxt and %Adv-menteTxt), they have been divided into six main frequency classes (Frequency Class).

Based on the results presented in Table 3.2, all *Adv-mente* with a frequency equal or above 10 were analysed and classified. As a result, a total of 532 from the 1.936 *Adv-mente* were classified which correspond to 97.05 % of the total occurrences of *Adv-mente* in the NILC/São Carlos corpus (see Appendix A).

<sup>3</sup>Broadly speaking, “Linguateca is a distributed network for fostering the computational processing of the Portuguese language” (<http://www.linguateca.pt/>). More information about this project is available in Santos (2009), among other papers available at Linguateca web page.

Frequency Class	LexAdv-mente	%Adv-menteLex	FrequencyTxt	%Adv-menteTxt
Frq > 101	158	8.16%	88.125	85.03%
Frq=51-100	76	3.93%	5.197	5.01%
Frq=21-50	153	7.90%	5.103	4.92%
Frq=16-20	60	3.10%	1.072	1.03%
Frq=11-15	70	3.62%	889	0.86%
Frq=1-10	1.419	73.30%	3.252	3.14%
Total	1.936	100.00%	103.638	100.00%

Table 3.2: Frequency Classes - Distribution of *Adv-mente* of the NILC/São Carlos corpus

### 3.3 Classification Issues

The classification of all 532 *Adv-mente* of the NILC/São Carlos corpus was made by means of the introspective method (Gross, 1975, 1988) and contrasted with data - random sample of sentences - from the NILC/São Carlos corpus by means of the AC/DC interface.

In order to carry out this task, the classification (PC, PS, PAa, PAh, PAm, PAs, MS, MV, MQ, MT, MP, MF) originally proposed by Molinier and Levrier (2000)<sup>4</sup> was taken into consideration. The goal of this section is to discuss some of the problems that were encountered during the process of classifying the 532 *Adv-mente*. Some adjustments made when performing the classification task of the *Adv-mente* are presented below.

In many cases the Brazilian *-mente* ending adverbs do not function in the same way that their corresponding in French *-ment* ending adverbs. For example, the *Adv-mente*<sup>5</sup> *decentemente* ‘decently’ was classified as MS (3.1). Its corresponding in French<sup>6</sup>, *déceamment* was classified as PS (3.2) and MS (3.3), by Molinier and Levrier (2000). Consider the following examples:

- *decentemente/déceamment*:

(3.1) *Eu estava vestido **decentemente** para sair à rua, só me faltava chapéu para a cabeça e porta franca para a retirada* (MS)

(3.2) ***Déceamment**, tu ne peux pas le faire* (PS)

<sup>4</sup>This study was briefly reviewed in Chapter 2.

<sup>5</sup>All Brazilian Portuguese examples were taken from NILC/São Carlos corpus.

<sup>6</sup>All French examples were taken from Molinier and Levrier (2000, pp. 428-519).

(3.3) *Elle se conduit **déce**mmment* (MS)

On the one hand, in Brazilian Portuguese, it is perfectly acceptable to use the *Adv-mente decentemente* (3.1) as a subject-oriented manner adverb (MS) just as in French (3.3). In this way, it is possible to split the main sentence and transform the adverb in an adjective, which can be related to the subject of the sentence, as example (3.4) illustrates:

(3.4) *Eu estava vestido (...) # Eu estava **decente***

I was dressed (...) # I was decent

On the other hand, in Portuguese it is unacceptable to use the *Adv-mente decentemente* as a disjunctive adverb of style (PS) unlike in French (3.2). *Decentemente* does not establish a relation with any metalinguistic operator, and consequently does not have its scope on this operator, as it can be seen in example (3.5):

(3.5) \**Eu considero **decentemente** que eu estava vestido para sair à rua (...)*

Another issue observed and that was also pointed out by Molinier and Levrier (2000, pp. 155-157) was the ambiguity that some *Adv-mente* may present inside the class MV. A clear example of this phenomenon is the *Adv-mente fixamente* ‘fixedly’, as the different paraphrases (3.6) and (3.7) illustrate:

(3.6) *O Pedro olhou **fixamente** para o quadro* (O olhar do Pedro estava fixo no quadro)

Peter stared fixedly at the painting (Peter looked continuously at the painting)

(3.7) *O Pedro pregou o quadro **fixamente** à parede* (O quadro estava fixo à parede)

Peter hung the painting up firmly on the wall (The painting was fixed on the wall)



In sentence (3.6), the adverb *fixamente* ‘fixedly’ refers to the fixed way that Peter was looking at the painting, only characterising Peter’s action. However, in sentence (3.7), the adverb *fixamente* ‘firmly’ not only refers to the action of hanging up, but also to the painting and gives an information about it: this painting was strongly fixed on the wall. The extra information that is given by the *fixamente* in example (3.7) makes it slightly different from the same *Adv-mente* presented in example (3.6).

This issue was not taken into consideration in the present study, because it is necessary to do an in-depth research on this subject in order to analyse better the phenomenon of ambiguity of *Adv-mente* in the same class in Portuguese. As a result, if an *Adv-mente* could present this phenomenon, just one general class was considered such as in the case of the *Adv-mente fixamente* ‘fixedly/firmly’ that was classified only as MV (see Appendix A).

Moreover, other problems were faced to classify some *Adv-mente*, specially the ambiguous ones. In many cases, it was difficult to determine the possible constructions of the *Adv-mente* due to the ambiguous or not so well-formulated sentences found in the corpus.

The *Adv-mente religiosamente* is an example of this difficulty. Consider the following sentences:

(3.8) *Seja como for, o Natal certamente traz a marca de uma ancestral anarquia contendo simbolismos tão misteriosos e universalmente intrincados quanto inesperados e, de certa forma, **religiosamente** irreverentes*

(3.9) *Mas continuava a cumprir **religiosamente** seu ritual: banhar-se de manhã e orar à tarde no templo*

(3.10) *O ministro israelense acrescentou que Jerusalém está fechada política e aberta **religiosamente***

In sentence (3.8), *religiosamente* seems to function as MP (viewpoint). This can be confirmed by using the paraphrase of this adverb class (see Section 2.1.3.1) as the example (3.11) shows:

(3.11) *Seja como for, o Natal certamente traz a marca de uma ancestral anarquia contendo simbolismos tão misteriosos e universalmente intrincados quanto*

*inesperados e, de certa forma, irreverentes de um ponto de vista religioso*

By analysing this sentence, the paraphrase used is totally acceptable, hence it proves that the adverb in question can be classified as MP. However, the *Adv-mente religiosamente* in example (3.9) does not seem to present the same behavior as in sentence (3.8).

Sentence (3.9) confirms that *religiosamente* can also be classified as MV, because it can establish a close relation to the verb. In this case, *religiosamente* cannot be detached to the initial position when using a negative sentence (3.12). In addition to this feature, it can be clearly paraphrased by the operator-noun in a manner construction (3.13).

(3.12) \*Mas **religiosamente**, não continuava a cumprir seu ritual: banhar-se de manhã e orar à tarde no templo

(3.13) *Mas continuava a cumprir **de forma religiosa** seu ritual: banhar-se de manhã e orar à tarde no templo*

In sentence (3.10), *religiosamente* is clearly a MP. Unlike in sentence (3.12), in this case this *Adv-mente* can be detached to the initial position when using a negative sentence (3.14) and paraphrased by ‘*de um ponto de vista Adj*’ ‘from a point of view Adj’ (3.15).

(3.14) *Religiosamente, (...) Jerusalém está aberta **religiosamente***

(3.15) *O ministro israelense acrescentou que Jerusalém está fechada política e aberta **do ponto de vista religioso***

Some difficulties were experienced when using the criteria proposed by Molinier and Levrier (2000) to distinguish the MV from the MS class and the MQ from the MF class.

Few *Adv-mente*<sup>7</sup> were classified as both MS and MV. The *Adv-mente* of both classes have similar behaviors regarding the two generic syntactic properties: detachment of adverbials to the initial position of a negative sentence and extraction of adverbials in cleft sentences (see Section 2.1.3.1).

<sup>7</sup>The *Adv-mente* are: *naturalmente* ‘naturally’, *claramente* ‘clearly’, *precisamente* ‘precisely’, *seguramente* ‘surely/safely’, *firmemente* ‘firmly’ and *cruelmente* ‘cruelly’.

In order to distinguish the MV from MS one of the features used was the paraphrases. The *Adv-mente* of the MV class only accept prepositional paraphrases with the respective noun of the adverb (e.g. *manualmente* ‘manually’ - *à mão* ‘by hand’), whereas the *Adv-mente* of the MS class accept adjectival paraphrases (e.g. *cuidadosamente* ‘carefully’ - *de maneira cuidadosa* ‘in a careful way’) or paraphrases with a predicative noun starting by the preposition *with* (e.g. *com cuidado* ‘with care’. The adverbs of both classes modify the verb. But the MS adverbs also establish a strong relation with the subject of the sentence, whereas the MV ones do not present this characteristic. Consider the example below:

(3.16) *Armstrong prosseguiu **cuidadosamente** durante a descida da escada de nove degraus*

Armstrong continued carefully while going down the nine-step stairs

It cannot be detached in a negative sentence (3.17), but it can be extracted in a cleft sentence (3.18):

(3.17) \***Cuidadosamente**, Armstrong não prosseguiu (...)

Carefully, Armstrong did not continue (...)

(3.18) *Foi **cuidadosamente** que Armstrong prosseguiu (...)*

It was carefully that Armstrong continued (...)

If the two paraphrases previously mentioned are used in the place of *cuidadosamente* to verify if this *Adv-mente* is a MV ou MS, we obtain the following results:

(3.19) *Armstrong prosseguiu **com cuidado** (...)*

Armstrong continued with care (...)

(3.20) *Armstrong prosseguiu **de maneira cuidadosa** (...)*

Armstrong continued in a careful way (...)

In sentences (3.19) and (3.20) both paraphrases are acceptable and both confirm that *cuidadosamente* is a manner sentence-internal adverb *Adv-mente*. Moreover, we can split the sentence and transform the *Adv-mente* in an adjective showing that it functions as a subject-oriented adverb (MS):

- (3.21) *Armstrong prosseguiu # Armstrong foi cuidadoso (...)*  
Armstrong continued # Armstrong was careful (...)

Now, consider the next examples:

- (3.22) *Veras disse que atuará firmemente no caso*  
Veras said that he will act firmly in the case

- (3.23) *O motorista deve fazer as curvas segurando o volante firmemente com as duas mãos*

The driver should make a turn holding the steering wheel firmly with both hands

By making use of the paraphrases, we can use *com firmeza* ‘with firmness’<sup>8</sup> and *de uma maneira firme* ‘in a firm way’. In sentences (3.22) and (3.23), both paraphrases are acceptable. Are these *Adv-mente* related to the verb or to the subject? This was a problem faced in many cases of *Adv-mente*.

If the sentences are split and the *Adv-mente* transformed into an adjective, we can see that sentence (3.24) is acceptable and roughly equivalent to (3.22), but in sentence (3.25) the adjective has nothing to do with the meaning of the original sentence (3.23).

- (3.24) *Veras disse que atuará no caso # Veras será firme*  
Veras said that he will act in the case # Veras will be firm

- (3.25) *O motorista deve fazer as curvas segurando o volante com as duas mãos # ?O motorista é firme*

The driver should make a turn holding the steering wheel firmly with both hands

In sum, the adjective of sentence (3.24) is clearly related to the subject, hence the *Adv-mente* is classified as MS. However in sentence (3.25), we realize that this adjective could be related to the subject, but in this case it changes the meaning

---

<sup>8</sup>It is a rare construction in English, because this is a literal translation.

of the sentence. In fact in the main sentence (3.23) the meaning is that the driver was holding the steering wheel in a firm way. In this case, *firmly* has its scope on the verb, being classified as MV.

All these analysis had to be taking carefully into consideration to distinguish the MS from MV *Adv-mente*, or to determine if an *Adv-mente* could be classified as both MS and MV.

A similar problem was observed concerning the MF and MQ classes, of course, considering their intrinsic features, described in Section 2.1.3.1. Some difficulties were faced in order to determine with which element in the sentence the *Adv-mente* established a relation and the nature of this element, since both classes may function as determinants (Baptista and Català, 2009b). This criteria diverge from the one proposed by Molinier and Levrier (2000), as the authors consider that only MF can function as determinants.

Finally, some difficulties were encountered when classifying the *Adv-mente* with a low frequency in the NILC/São Carlos corpus, namely the ones with a frequency lower or equal to 30. As it was expected, the few number of examples of these *Adv-mente* in context was an obstacle to determine all possible constructions of these adverbs.

When necessary, a search for more examples of the less frequent *Adv-mente* was done on Google<sup>9</sup> in order to analyse more possible constructions of these adverbs.

If the results of this search could not help in providing more illustrative examples, the last solution adopted was to consult the French list of *-ment* ending adverbs made by Molinier and Levrier (2000, pp. 428-519) as an attempt to contrast the constructions of *Adv-mente* in both languages. However, this comparison sometimes was not possible since some adverbs are specific of each language, for instance, the *Adv-mente proutamente* ‘soon’ does not have a direct *-ment* adverb corresponding in French.

The problems here described illustrate how difficult it is to use all features proposed by Molinier and Levrier (2000) when analysing the *Adv-mente* in context. If this task is difficult for a human-linguist, how could a learning algorithm produce better results? It would be really hard for a machine to paraphrase adverbs, detached them to the initial position of a sentence or even to extract them in cleft

---

<sup>9</sup>The Brazilian web page is [www.google.com.br](http://www.google.com.br).

sentences. It seems that Molinier and Levrier (2000) analysed many adverbs out of context, giving simple examples to explain the features, but they did not to analyse the real usage of these adverbs. As there is a huge number of *Adv-mente* and to classify, and consequently disambiguate them manually is time-consuming, an attempt to perform these tasks automatically was made. Therefore, at a first approach, only the linguistic features that could be extracted automatically and many others that were proposed based on the analysis of these adverbs in context were defined to be used in some learning-based experiments (see Section 4).

### 3.4 Classification Results

Considering the 532 *Adv-mente* that were classified, some *Adv-mente* could belong to one or more classes. The general result is shown in Table 3.3.

Group	Number of Classes	N° of <i>Adv-mente</i>	FrequencyTxt	%FrequencyTxt
Group 1	1	434	71.587	71.17%
Group 2	2	81	17.989	17.89%
Group 3	3	14	8.639	8.59%
Group 4	4	3	2.361	2.35%
Total		532	100.576	100,00%

Table 3.3: General Distribution based on the Classification of the 532 *Adv-mente*

On the one hand, a total of 434 non-ambiguous *Adv-mente* were classified. As they can belong to only one of the twelve classes they were grouped together (group 1). Their total frequency represent 71,17% of all 532 *Adv-mente* analysed in the corpus.

On the other hand, 98 *Adv-mente* can be classified into two or more classes. Therefore, the *Adv-mente* of groups 2, 3 and 4, may be classified into two, three or four classes, respectively. The frequency of the *Adv-mente* of these three groups correspond to 28,83% of the total frequency of the 532 *Adv-mente* analysed.

In spite of the small number of *Adv-mente* that can be classified into two or more classes, many of them present a highly significant frequency in the corpus, for example, the *Adv-mente especialmente* ‘specially/especially’ that can be classified as MQ or MF and has a frequency of 13,24% (2.490). This is the fourth most frequent *Adv-mente* in the corpus (see Appendix A). If a frequent *Adv-mente* might present different constructions, each one with a different scope and, consequently,

with a different meaning, it seems relevant to determine correctly the construction it shows in each instance.

Table 3.4 illustrates the frequency of each class (PC, PS, PAa, PAh, PAm, PAs, MS, MV, MQ, MT, MP and MF) identified in the 532 *Adv-mente* that were analysed.

Class	Frequency	%Frequency
PC	23	3.54%
PS	16	2.46%
PAa	18	2.77%
PAh	9	1.38%
PAm	24	3.69%
PAs	5	0.77%
MS	135	20.77%
MV	218	33.54%
MQ	83	12.77%
MT	46	7.08%
MP	52	8.00%
MF	21	3.23%
Total	650	100%

Table 3.4: Frequency of each class based on the classification of all 532 *Adv-mente*

A total of 650 constructions were determined. The most frequent classes are MV (33.54%) and MS (20.77%). This result corroborates the observations made by Molinier and Levrier (2000, p. 295) and the general intuition that *-mente* ending adverbs are mostly manner adverbs. The authors conclude in their study that the largest class of *-ment* ending French adverbs is MV. The same conclusion can be made concerning the Brazilian Portuguese *Adv-mente*. The class PAs has the lowest frequency, representing a 0.77% of the total.

Table 3.5 shows the frequency of each class considering only the 434 non-ambiguous *Adv-mente* of group 1. As it was mentioned before, in this group each *Adv-mente* can be classified into only one of the classes. For this reason, 434 constructions were determined. The most frequent classes in group 1 reflects the general result shown in Table 3.3, hence MV (35.25%) and MS (23.73%) are the classes with the highest frequencies in group 1. No construction of the PAs class was identified in this group. The PS class has the second lowest frequency (0.69%) in group 1.

The results for the frequency of each class regarding the groups 2, 3 and 4 - the ones that present ambiguity classes - are given in Table 3.6. Group 2, 3 and 4 present a total of 161, 42 and 12 constructions, respectively.

Class	Group 1	%Group 1
PC	11	2.53%
PS	3	0.69%
PAa	7	1.61%
PAh	6	1.38%
PAm	18	4.15%
PAs	0	0%
MS	103	23.73%
MV	153	35.25%
MQ	52	11.98%
MT	41	9.45%
MP	28	6.45%
MF	12	2.76%
Total	434	100%

Table 3.5: Frequency of each class based on the classification of the *Adv-mente* of group 1

In general, the MV and MS classes usually present the highest frequencies in all three groups. Some classes of sentence-external modifiers (P) such as PAh have a low frequency in all three groups or even do not appear at all.

Class	Group 2	%Group 2	Group 3	%Group 3	Group 4	%Group 4
PC	9	5.56%	1	2.38%	2	16.67%
PS	6	3.70%	6	14.29%	1	8.33%
PAa	9	5.56%	2	4.76%	0	0%
PAh	3	1.85%	0	0%	0	0%
PAm	3	1.85%	3	7.14%	0	0%
PAs	4	2.47%	0	0%	1	8.33%
MS	22	13.58%	7	16.67%	3	25.00%
MV	54	33.33%	9	21.43%	2	16.67%
MQ	19	11.73%	11	26.19%	1	8.33%
MT	5	3.09%	0	0%	0	0%
MP	24	14.81%	0	0%	0	0%
MF	4	2.47%	3	7.14%	2	16.67%
Total	162	100%	42	100%	12	100%

Table 3.6: Frequency of each class based on the classification of the *Adv-mente* of group 2, 3 and 4

Taking into consideration the different class combinations of all 98 ambiguous *Adv-mente* of groups 2, 3 and 4, Tables 3.7, 3.8 and 3.9 illustrate the different class combinations identified in each one of the three groups, respectively.

A total of 21 different class combinations were identified in group 2 (see Table 3.7). The most frequent class combinations are *MP*, *MV*(29.63%) and *MQ*, *MV*(14.81%). In both class combinations, the class *MV* is present. This same class appear in more six class combinations: *MS*, *MV* / *MT*, *MV* / *PAa*, *MV* / *PAh*, *MV* / *PAm*, *MV* / *PC*, *MV*. In this group, the *MV* class is the most frequent



2 classes combinations	N° of <i>Adv-mente</i>	%Frequency
MF, MS	1	1.23%
MP, MV	24	<b>29.63%</b>
MQ, MF	2	2.47%
MQ, MS	2	2.47%
MQ, MV	12	<b>14.81%</b>
MS, MT	1	1.23%
MS, MV	2	2.47%
MT, MS	1	1.23%
MT, MV	2	2.47%
PAa, MQ	3	3.70%
PAa, MS	3	3.70%
PAa, MV	3	3.70%
PAh, MT	1	1.23%
PAh, MV	2	2.47%
PAm, MS	1	1.23%
PAm, MV	2	2.47%
PAs, MS	4	4.94%
PC, MS	2	2.47%
PC, MV	7	8.64%
PS, MF	1	1.23%
PS, MS	5	6.17%
<b>Total</b>	81	100%

Table 3.7: Combinations of classes in group 2

ambiguity class.

Table 3.8 shows the 9 different class combinations identified in group 3. The most frequent class combination is PS, MQ, MS (21.43%). In fact, the frequency of all the 9 class combinations do not present a significant difference as it can be seen in Table 3.8.

3 classes combinations	N° of <i>Adv-mente</i>	%Frequency
MQ, MF, MV	1	7.14%
PAm, MQ, MV	1	7.14%
PC, MF, MS	1	7.14%
PS, MV, MQ	1	7.14%
MF, MQ, MS	1	7.14%
PAa, MQ, MV	2	14.29%
PAm, MS, MV	2	14.29%
PS, MQ, MV	2	14.29%
PS, MQ, MS	3	21.43%
<b>Total</b>	14	100%

Table 3.8: Combinations of classes in group 3

Group 4 (see Table 3.9) has three different class combinations and all of them have the same frequency (33.33%). Note that, the class MV and MF is present in two of the three class combinations, whereas the class MS appears in all three

combinations.

4 classes combinations	N° of <i>Adv-mente</i>	%Frequency
PAs, MQ, MS, MV	1	33.33%
PC, MF, MS, MV	1	33.33%
PC, PS, MF, MS	1	33.33%
<b>Total</b>	<b>3</b>	<b>100%</b>

Table 3.9: Combinations of classes in group 4

These results demonstrate that the classes MV and MS are usually present in the most frequent class combinations in groups 2, 3 and 4. This means that these two classes, in general, are the most ambiguous classes based on the classification of the 532 *Adv-mente*. Finally, it is worthy to note that these same classes (MV and MS) are also the most frequent ones concerning the non-ambiguous *Adv-mente* (see Table 3.5).

### 3.5 Comparison with Electronic Dictionaries

The 532 *Adv-mente* analysed and classified were compared to the *Adv-mente* found in two Brazilian dictionaries: Aurélio Electronic Dictionary of Brazilian Portuguese (Buarque de Holanda Ferreira, 2004) and Houaiss Electronic Dictionary (Houaiss and Mauro, 2009), hereafter referred to as AEDic and HEDic, respectively. The aim of this comparison is to validate this work by showing if the *Adv-mente* selected from the NILC/São Carlos corpus are included in these dictionaries and if so to analyse how they are treated.

The AEDic has 257 entries of *Adv-mente*, whereas the HEDic has 127 entries of the same kind of adverbs. These numbers are inferior to the 1.936 *Adv-mente* identified in the NILC/São Carlos corpus. One reason for the low number of *Adv-mente* entries may be explained by the fact that many *Adv-mente* in both dictionaries are included indirectly in the entry of their respective base adjective. For instance, the word *claro* ‘clear’, in Portuguese, can be a noun, an adjective, an adverb or an interjection. There is an *Adv-mente* related to this word, namely *claramente*. In the AEDic, this *Adv-mente* is not treated as an independent entry, but it is used in the microstructure of the word *claro* to explain its meaning of when used as an adverb, as the it can be seen in the example below:

“**Claro:** *advérbio. Com clareza; claramente: falar claro; ver claro; “Falemos claro, amigo. Viemos à caça do sujeito, e por força que o havemos de levar.”* (José de Alencar, *O Gaúcho*, p. 127).” (Buarque de Holanda Ferreira, 2004)

Most of the *Adv-mente* are not even included in dictionaries, for example, the *Adv-mente igualmente* ‘equally’ neither has an entry nor is treated inside the entry of its base adjective *igual* ‘equal’.

It is unknown which criteria are used by the lexicographers to include or not *Adv-mente* in dictionaries, and once included if this *Adv-mente* should be treated as an independent entry or not (macrostructure). Furthermore, the information given (microstructure) about the *Adv-mente* in both dictionaries is basically made by means of its relation to the base adjective (1), paraphrases (2), equivalence to other *Adv-mente* (3) and / or examples usually taken from literary works (4), as the follow example of the *Adv-mente tortamente* ‘askew’ illustrates:

“**tortamente:** [Do f. de torto + -mente.] *Advérbio. 1. De modo torto; impropriamente, torcidamente, tortuosamente: “Pedro e Paulo falavam-lhe por exclamações, por exortações, por súplicas, a que ela respondia mal e tortamente”* (Machado de Assis, *Esaú e Jacó*, p. 258).” (Buarque de Holanda Ferreira, 2004)

1. **its relation to the base adjective:** [Do f.<sup>10</sup> de torto + -mente.]
2. **paraphrases:** de modo torto
3. **equivalence to other *Adv-mente*:** impropriamente, torcidamente and tortuosamente
4. **examples taken from literary works:** “Pedro e Paulo falavam-lhe por exclamações, por exortações, por súplicas, a que ela respondia mal e tortamente” (Machado de Assis, *Esaú e Jacó*, p. 258)

All this information is not always provided. It was observed that in many cases in both dictionaries some of them was given such as in the following example:

---

<sup>10</sup>The abbreviation used in the dictionary for feminino ‘feminine’.

“*virgemente*: adv. à maneira de virgem; como virgem ETIM<sup>11</sup> virgem + -mente” (Houaiss and Mauro, 2009)

In this example, it is only provided two paraphrases and the relation of the adverb *virgemente* to its base adjective *virgem*. This means that the information given about the *Adv-mente* may vary, so that it is not possible to obtain the same information (the four ones previously described) about all of them.

In both dictionaries, if a *Adv-mente* is said to have more than one meaning, there are sub-entries for each different definition (see the example of  *pessoalmente* ‘personally’ below). In these cases, apart from the information previously mentioned (that may be used or not), no extra explanation is usually given in order to distinguish the different meanings of the ambiguous *Adv-mente*, as the following example illustrates.

“*pessoalmente*: adv. 1. sem a intermediação de nada ou de ninguém; em pessoa, diretamente [fui ouvi-la p.] [quero conhecê-lo p., não pela internet] 2. de uma perspectiva pessoal; particularmente [p. prefiro ir às compras a fazer turismo cultural] ETIM pessoal + -mente” (Houaiss and Mauro, 2009)

This example shows that the information given about the *Adv-mente* in AEDic and HEDic seems to be insufficient to help the reader to understand the usage of these adverbs and to distinguish the different meanings of ambiguous *Adv-mente*. This result is not surprising. If adverbs are superficially treated by grammarians as it was discussed in section 2.1, it was not expected that lexicographers would give a more complex and in-depth analysis of the *Adv-mente*.

Table 3.10 shows how many of the 532 most frequent *Adv-mente* of the NILC/São Carlos corpus are treated in the dictionaries in question.

<b>HEDic</b>	62	11.65%
<b>AEDic</b>	44	8.27%
<b>HEDic and AEDic</b>	31	5.83%
<b>NOT APPEARING</b>	395	74.25%
<b>TOTAL</b>	532	100%

Table 3.10: Comparison between the *Adv-mente* of Brazilian Dictionaries and the most frequent *Adv-mente* of the NILC/São Carlos corpus

<sup>11</sup>ETIM is the abbreviation for etimologia ‘etymology’.

From the 532 *Adv-mente* of the NILC/São Carlos corpus analysed in this study, 62 (11.65%) of them only appear in the HEDic, 44 (8.27%) only appear in the AEDic and 31 (5.83%) are present in both dictionaries. Based on this, only 25.75% of the 532 *Adv-mente* analysed are treated in at least one of the two dictionaries, hence most of the 532 *Adv-mente* (395 that corresponds to 74.25%) cannot be found neither in HEDic nor in AEDic.

These results show that many *Adv-mente* are not treated by neither AEDic nor by HEDic. However, what it is interesting to note is that all the 23 most frequent *Adv-mente* (see Appendix B) of the NILC/São Carlos corpus (see Appendix B) are treated in at least in one of the dictionaries. As the frequency of the *Adv-mente* decreases, the probability that an adverb with a low frequency will appear in one of the dictionaries also decreases.

This may be an evidence that Brazilian dictionaries treat only the most frequent *Adv-mente* in the language. Nevertheless, it is still unknown why so few adverbs are treated in those dictionaries as independent entries. This comparison corroborates the general intuition that most of the *Adv-mente* are not treated in dictionaries. Few of them are considered and their information is poor, not organized in a clear system and most of the examples given does not reflect the real use of these adverbs. For this reason, it would be difficult to use this information for analysing the *Adv-mente* in order to disambiguate them automatically.

# Chapter 4

## WSD: An Experiment on *Adv-mente*

This chapter describes the methodology and the machine learning (ML) experiments that were carried out. First some general observations are explained (Section 4.1). The ambiguous *Adv-mente* that were selected to be analysed are presented (Section 4.2). Then, the tasks of building the training data and annotating the instances of this dataset are described (Sections 4.3 and 4.4). The features proposed, their respective values and the task of extracting them are discussed (Section 4.5). Finally, two ML experiments are presented. They were divided according to the way that the information of the training dataset was used (Section 4.6).

### 4.1 General observations

In the previous chapter, a classification for the *Adv-mente* was proposed and the problem of ambiguous cases was explained. In sum, 98 from the 532 *Adv-mente* of the NILC/São Carlos corpus that were manually classified present at least two constructions, and consequently each construction convey a different meaning. Now the aim is to verify whether it is possible to automatically determine both the classification of ambiguous *Adv-mente* instances and their correct sense in a sentence, since the ambiguity of *Adv-mente* is connected to the classification issue (Baptista and Català, 2009b).

As it was presented in section 2.3, there are different NLP methods to handle the WSD problem. It seems that adopting, for example a *pre-codified knowledge method*, in which information of the words could be extracted from resources such as electronic dictionaries, would probably not be the best one to handle the ambiguity problem of *Adv-mente*. In section 3.5, it was explained some problems concerning the treatment of these adverbs based on the analysis of two Brazilian electronic dictionaries. In sum, more than half of the 532 *Adv-mente* selected from the NILC/São Carlos corpus are not treated as independent entries (macrostructure) in any of the two dictionaries; some of them *Adv-mente* are treated inside the microstructure of their respective base adjective; others are totally ignored; and the information given about the *Adv-mente* seems to be insufficient to help in performing a disambiguation task. Empiric Method based on corpus disambiguation models are obtained from the superficial knowledge based on information from the corpus and also from statistical or machine learning techniques.

For this reason, it was decided to adopted an *empiric method based on corpus* which is also one of the methods that can be applied to WSD task. Based on this, the disambiguation of *Adv-mente* which present more than one construction may be achieved by using machine learning-based techniques (Witten and Frank, 2005). Depending on the type of the data used, a supervised, minimally supervised or unsupervised approach may be adopted (Jurafsky and Martin, 2009).

It was decided to adopted a supervised approach to carry out the experiments on the disambiguation of *Adv-mente*. For Jurafsky and Martin (2009, p. 673), “if we have data that has been hand-labeled with correct word senses, we can use a supervised learning approach to the problem of sense disambiguation - extracting features from the text that are helpful in predicting senses and then training a classifier to assign the correct sense given these features. The output of training is thus a classifier system capable of assigning sense labels to unlabeled words in context”. As a result, sentences selected from the NILC/São Carlos corpus were annotated and features were extracted in order to build a training dataset that could be processed by a machine learning tool.

To carry out the ML experiments the Weka tool (Witten and Frank, 2005; Hall et al., 2009) was used. In Weka, a varied number of different learning algorithms and data preprocessing tools are available which facilitates their application to

dataset previously built.

## 4.2 *Adv-mente* used in the experiment

As it was explained in section 3.3, some difficulties were faced when classifying the *Adv-mente* with a low frequency in the NILC/São Carlos corpus due to the few number of examples available in the corpus.

For this reason, from the 98 ambiguous *Adv-mente*, it was decided that only the *Adv-mente* with a frequency equal or higher to 50 would be used in the ML experiment, therefore 53 *Adv-mente* were selected (see Table 4.1).

Group	N° of <i>Adv-mente</i>	N° of <i>Adv-mente</i> Selected
Group 2	81	39
Group 3	14	12
Group 4	3	2
Total	98	53

Table 4.1: Ambiguous *Adv-mente* selected to be used in the ML experiment

From the twelve classes proposed for the *Adv-mente* (PC, PS, PAa, PAh, PAm, PAs, MS, MV, MQ, MT, MP, MF), the PAs class, subject-oriented disjunctive adverbs of attitude, was the only one that was not considered in the experiment, because none of the 53 ambiguous *Adv-mente* selected present a PAs construction. This means that a total of eleven classes were taken into consideration to carry out the experiments.

## 4.3 Building the Training Data

In order to use a machine learning method, first, it is necessary to determine a training data.

A sample of sentences containing the 53 *Adv-mente* was selected from the NILC/São Carlos corpus to build the training data to be used in the experiments. This corpus was consulted by means of the AC/DC interface in the Linguatca website (see section 3.1). The first 50 sentences obtained in the search result were selected for each one of the 53 ambiguous *Adv-mente*, thus a total of 2650 sentences were taken from the corpus. In each sentence, only one *Adv-mente* was considered.



Each sentence corresponds to an instance of the training data. This sample of 2650 instances will be hereafter referred to as NSCSAMPLE.

Each one of these 53 *Adv-mente* has a different frequency in the corpus, and may be the training dataset should ideally reflect these frequency numbers. Nevertheless, due to the experiments that were decided to be conducted it was decided to select the same number of sentences of each one of the *Adv-mente*. In section 4.6, this decision is explained and the experiments mentioned are described.

Some problems concerning the sentences obtained in the search results were observed.

The first problem concerned **the length of the sentence**. Some sentences were extremely long and this made its analysis more difficult and time-consuming, as the example (4.1) illustrates:

(4.1) *Tratando-se de uma atividade dependente de autorização governamental e sujeita a fiscalização oficial, surgiram em vários países, **especialmente** em países sob regime totalitário ou tecnologicamente atrasados, em lugar de ligas que surgissem aos radioamadores informações mundiais sobre o radioamadorismo, grupos que procuraram funcionar como intermediários dos governos e dos radioamadores, apresentando às autoridades os radioamadores como impúberes que só por eles podem ser administrados, e aos radioamadores, que eles precisam pagar ao grupo por esta administração.*

Secondly, some sentences presented some **errors**, for example, in the place of appearing the verb *estava* it was found *'tava*, as it can be seen in example (4.2). This and other errors generated by the AC/DC system are described in the web page of the AC/DC project<sup>1</sup>.

(4.2) *par=Brasil-94b-pol-2: Dimenstein argumentou que uma das indicações de que não **'tava** movido por questões partidárias ou políticas contra o PT foi o agradecimento que lhe foi feito **pessoalmente** por Lula.*

**Repetition of sentences** was the other problem encountered. In some cases, in the results obtained some sentences appeared twice. For this reason, only one

---

<sup>1</sup><http://dinis.linguateca.pt/ACDC/>.

of them was considered. This happened, for example, in the case of the *Adv-mente claramente*, examples (4.3) and (4.4):

(4.3) *par=16054: Perguntou se o tal lance não provava **claramente** que não bastavam um árbitro e dois bandeirinhas para decidir dúvidas fundamentais no jogo.*

(4.4) *par=16180: Perguntou se o tal lance não provava **claramente** que não bastavam um árbitro e dois bandeirinhas para decidir dúvidas fundamentais no jogo.*

Although NILC/São Carlos corpus has a collection of Brazilian Portuguese texts, mainly journalistic, some sentences written in Spanish were found in the corpus, as in example (4.5). This was the last problem identified. Aluísio and Pinheiro (2003) explains that some sentences written in foreign languages might be found in the corpus. The sentences in Spanish were not considered.

(4.5) *par=40072: Del 1 al 6 noviembre, la SLHCT organizó en Tikal, Guatemala la Conferência Pan-americana de Matematica, Astronomia y Formas de Pensamiento Pre-colombinos, juntamente con el Grupo de Estudios de Etnomatematicas del Departamento de Matematica de la Universidade de San Carlos de Guatemala.*

If any of the first 50 sentences obtained in the search result for each *Adv-mente* presented one of the problems previously described, thus another one had to be selected to replace it. The new sentence was selected from the following 20 search results of each *Adv-mente*.

## 4.4 Annotations

After selecting the sentences, an annotation task was performed in two stages: a hand-labeled annotation to assign the construction of each *Adv-mente* in context (4.4.1) and an automatic annotation to obtain morpho-syntactic tags of all tokens (4.4.2).

### 4.4.1 Hand-labeled annotation

Firstly, the construction (PC, PS, PAa, PAh, PAm, MS, MV, MQ, MT, MP, MF) of each *Adv-mente* in all 2650 sentences was determined, based on the classification previously made (see chapter 3) and on the analysis of its context. This task was manually undertaken as, to our knowledge, there is no system that can do it automatically. Table 4.2 shows the number of cases identified for each one of the possible constructions of 53 *Adv-mente*. Note that for each one of these *Adv-mente*, a sample of 50 sentences were previously selected.

ID	Adv-mente	Nº of Constructions identified of each <i>Adv-mente</i>							
		Sense 1	Nº	Sense 2	Nº	Sense 3	Nº	Sense 4	Nº
1	especialmente	MQ	4	MF	46	-	-	-	-
2	normalmente	PAh	30	MV	20	-	-	-	-
3	exclusivamente	MF	50	MS	0	-	-	-	-
4	igualmente	PC	23	MV	27	-	-	-	-
5	oficialmente	MP	7	MV	43	-	-	-	-
6	pessoalmente	PS	5	MF	45	-	-	-	-
7	claramente	MS	9	MV	41	-	-	-	-
8	infelizmente	PAa	50	MS	0	-	-	-	-
9	independentemente	PC	43	MS	7	-	-	-	-
10	perfeitamente	MQ	30	MV	20	-	-	-	-
11	propriamente	MQ	13	MV	37	-	-	-	-
12	profundamente	MQ	21	MV	29	-	-	-	-
13	fortemente	MQ	37	MV	13	-	-	-	-
14	altamente	MQ	50	MV	0	-	-	-	-
15	juntamente	PC	50	MV	0	-	-	-	-
16	essencialmente	MQ	29	MF	21	-	-	-	-
17	originalmente	MT	50	MV	0	-	-	-	-
18	formalmente	MP	11	MV	39	-	-	-	-
19	felizmente	PAa	50	MS	0	-	-	-	-
20	paralelamente	PC	18	MV	32	-	-	-	-
21	sexualmente	MP	29	MV	21	-	-	-	-
22	curiosamente	PAa	48	MV	2	-	-	-	-
23	ligeiramente	MQ	29	MV	21	-	-	-	-
24	rigorosamente	MQ	17	MV	33	-	-	-	-
25	levemente	MQ	30	MV	20	-	-	-	-
26	duramente	MQ	22	MV	28	-	-	-	-
27	consequentemente	PC	50	MS	0	-	-	-	-
28	sinceramente	PS	28	MS	22	-	-	-	-
29	profissionalmente	MP	16	MV	34	-	-	-	-
30	mentalmente	MP	22	MV	28	-	-	-	-
31	firmemente	MS	32	MV	18	-	-	-	-
32	comercialmente	MP	11	MV	39	-	-	-	-
33	genericamente	MP	16	MV	34	-	-	-	-

Table 4.2: Number of *Adv-mente* constructions identified in the 2650 sentences

Table 4.2 – continued from previous page

ID	Adv-mente	N° of Constructions identified of each <i>Adv-mente</i>							
		Sense 1	N°	Sense 2	N°	Sense 3	N°	Sense 4	N°
34	brevemente	MS	20	MT	30	-	-	-	-
35	inversamente	PC	23	MV	27	-	-	-	-
36	conjuntamente	PC	4	MV	46	-	-	-	-
37	emocionalmente	MP	27	MV	23	-	-	-	-
38	visualmente	MP	28	MV	22	-	-	-	-
39	criminalmente	MP	0	MV	50	-	-	-	-
40	realmente	PS	5	MV	26	MQ	19	-	-
41	exatamente	MF	38	MQ	9	MS	3	-	-
42	justamente	PC	2	MF	46	MS	2	-	-
43	naturalmente	PAm	29	MS	2	MV	19	-	-
44	particularmente	PS	3	MQ	14	MF	33	-	-
45	verdadeiramente	PS	1	MQ	37	MV	12	-	-
46	seriamente	PS	2	MQ	31	MS	17	-	-
47	virtualmente	PAm	5	MQ	37	MV	8	-	-
48	seguramente	PAm	46	MS	0	MV	4	-	-
49	francamente	PS	15	MQ	22	MS	13	-	-
50	decididamente	PS	14	MQ	16	MS	20	-	-
51	superficialmente	PAa	2	MQ	8	MV	40	-	-
52	simplesmente	PC	1	PS	2	MF	47	MS	0
53	precisamente	PC	0	MF	49	MS	0	MV	1

Table 4.2: Number of *Adv-mente* constructions identified in the 2650 sentences

At least one of the constructions was not found in any of the 50 sentences selected for each one of the following 10 *Adv-mente*: ID 3, 8, 14, 15, 17, 19, 27, 39, 52 and 53. In the case of the 43 remaining *Adv-mente*, examples of all possible constructions of each *Adv-mente* were identified, though, in some cases, very few examples were found regarding some constructions (e.g. *justamente* PC=2 and MS=2).

Table 5.11 summarises the distribution of instances of the NSCSAMPLE per class.

On the one hand, very few instances of *Adv-mente* (30) were classified as disjunctive evaluative adverbs class (PAh). On the other hand, a huge number of instances, more precisely 857, were classified as being proper manner adverbs (MV). The second most frequent class of *Adv-mente* 475 found in this sample of instances was the subject-oriented manner adverb (MS). These results corroborates the observations made in section 3.4, in which the most frequent classes identified when performing the classification task were also the MV and MS classes.

Class	Number of instances
PC	214
PS	75
PAa	150
PAh	30
PAm	80
MS	147
MV	857
MQ	475
MP	167
MT	80
MF	375
<b>TOTAL</b>	2650

Table 4.3: Number of instances per class originally selected *Adv-mente*

#### 4.4.2 Automatic annotation

Secondly, a part-of-speech (POS) tagger for Portuguese was used to assign contextually appropriate grammatical descriptors to words in the sentences. The POS tagger chosen was the Lx-Tagger<sup>2</sup> (Branco and Silva, 2004). This tagger was developed and is maintained by the NLX-Natural Language and Speech Group<sup>3</sup>.

This tool can be downloaded or a demo version<sup>4</sup> available online can be used. The former version assigns a single morpho-syntactic tag<sup>5</sup> to every token (words and punctuation). The latter version provides not only a morpho-syntactic tag but also the lemma and inflectional information. Both versions were tested and the difference between the outputs of each one can be seen in examples (4.6) and (4.7):

(4.6) *Essa/DEM camada/CN gordurosa/ADJ é/V especialmente/ADV  
desenvolvida/PPA em/PREP animais/CN de/PREP clima/CN frio/ADJ  
.\*/PNT - **Version for download***

(4.7) *<p><s>Essa/DEM#fs camada/CAMADA/CN#fs  
gordurosa/GORDUROSO/ADJ#fs é/SER/V#pi-3s especialmente/ADV  
desenvolvida/DESENVOLVER,DESENVOLVIDO/PPA#fs em/PREP*

<sup>2</sup><http://lxcenter.di.fc.ul.pt/tools/en/LXTaggerEN.html>

<sup>3</sup>More information about this research group is available at <http://nlx.di.fc.ul.pt/>.

<sup>4</sup>The demo version is available at <http://lxcenter.di.fc.ul.pt/services/en/LXServicesSuite.html>.

<sup>5</sup>The tagset used is available in the web page of Lx-Tagger.

*animais/ANIMAL/CN#mp de/PREP clima/CLIMA/CN#ms*  
*frio/FRIO/ADJ#ms .\*/PNT </s.s></p>* - **Online version** -

The output (see 4.7) of the online version provides more information (e.g. lemma (nouns, adjectives and verbs) and person) that could be used to define some features to help in disambiguating *Adv-mente*. For this reason, the online version was used in this study. The disadvantage of using this version is that there is a limit of characters, hence few sentences can be tagged at the same time.

The main characteristics of the output generated by online version of the Lx-Tagger are:

- (a) the tag is attached to the token, using a / (slash) symbol as separator (e.g. *de/PREP*)
- (b) the lemma of nouns, adjectives and verbs are given (e.g. *frio/FRIO/ADJ#ms*)
- (c) morphological information is also given, using a # (hash) symbol as a separator (e.g. *é/SER/V#pi-3s*)
- (d) all contractions such as *do* and *neste* are expanded as follows *de\_/PREP o/DA#ms* and *em\_/PREP este/DEM#ms*
- (e) each individual token in multi-word expressions of closed POS classes gets the tag of that expression prefixed by "L" and followed by the number of its position within the expression (e.g. *de modo que = de/LCJ1 modo/LCJ2 que/LCJ3*)

This tagger was developed with a software called MXPOST<sup>6</sup> based on a large (600k token) and accurately hand-tagged corpus. It has obtained an accuracy of 96.24% using a 10-fold cross evaluation.

Even with a high accuracy, some problems were identified in the output of the tagged sentences that had to be manually corrected. Sometimes there was a double space between *Adv-mente* and the token that came immediately before it as example (4.8) illustrates:

---

<sup>6</sup>More information is available at [http://www.inf.ed.ac.uk/resources/nlp/local\\_doc/MXPOST.html](http://www.inf.ed.ac.uk/resources/nlp/local_doc/MXPOST.html).

(4.8) <p><s>As/DA#fp escolas/ESCOLA/CN#fp de/PREP  
 samba/SAMBA/CN#ms ,/PNT **normalmente/ADV** ,\*//PNT  
 só/ADV abrem/ABRIR/V#pi-3p as/DA#fp quadras/QUADRA/CN#fp  
 a/PREP partir/PARTIR/V#inf-nInf de/PREP julho/JULHO/CN#ms  
 ,\*//PNT quando/CJ começa/COMEÇAR/V#pi-3s a/DA#fs  
 seleção/SELEÇÃO/CN#fs de-/PREP o/DA#ms  
 samba-enredo/SAMBA-ENREDO/CN#ms para/PREP o/DA#ms  
 Carnaval/PNM seguinte/SEGUINTE/ADJ#gs .\*//PNT </s>

This double space produced an error when trying to extract the features automatically. As a result, they had to be removed.

Another problem was caused by an error produced by the Lx-Tagger. In some cases, this POS tagger did not tag correctly the *Adv-mente* and the tokens that came immediately before and after it, as it can be seen in example (4.9)

(4.9) <s>Logo/ADV abaixo/ADV ,\*//PNT  
**um sinceramente/ADV secreto/UM SECRETO/ADJ#ms**  
 amigo/AMIGO/CN#ms de-/PREP o/DA#ms doutor/DOUTOR/CN#ms  
 dá/DAR/V#pi-3s uma/UM#fs longa/LONGO/ADJ#fs  
 resposta/RESPOSTA/CN#fs em/PREP inglês/INGLÊS/CN#ms a/PREP  
 um/UM#ms Mr/PNM .\*//PNT Boss/PNM .\*//PNT </s>

To solve this problem, the output had to be carefully revised and the missing or wrong tags had to be manually inserted/corrected. One may think that the version for download might not present this errors, but the same problem was observed in this last version:

(4.10) Logo/ADV abaixo/ADV ,\*//PNT **um sinceramente secreto/DA**  
 amigo/CN de-/PREP o/DA doutor/CN dá/V uma/UM longa/ADJ  
 resposta/CN em/PREP inglês/CN a/PREP um/UM Mr/CN .\*//PNT  
 Boss/PNM .\*//PNT

Based on both annotations that were made and the linguistic information previously discussed in Chapter 2, some features were established with the purpose of being used in the ML experiment as explained in the next section.

## 4.5 Features and Extraction of Values

A total of 17 features were proposed in order to be used in the automatic classification (and consequently in the disambiguation) of instances of *Adv-mente*. In addition to the analysis of the context which surrounds the *Adv-mente*, some linguistic properties of these adverbs (see Section 2.1.3.1) were also taken into consideration.

The values of each feature were defined based on a detailed examination that was made regarding the elements that were being analysed. The values for every instance of *Adv-mente* in NSCSAMPLE were extracted automatically by means of a code written in Visual Basic for Applications 6.5. (VBA) that is included in many Microsoft Office applications. The application used was the Microsoft Office Excel, allowing that the values of the features for each instance in NSCSAMPLE were put in separated columns and rows in a spreadsheet of Excel. This spreadsheet was converted into an ARFF file, so that Weka could read it.

Table 4.4 summarises the features proposed, their respective definitions and values. A detailed description of each feature is presented in the items below.

Nº	Feature ID	Description	Values
1	<b>BEGSENT</b>	The <i>Adv-mente</i> is at the beginning at the sentence	TRUE, FALSE
2	<b>ENDSENT</b>	The <i>Adv-mente</i> is at the end of the sentence	TRUE, FALSE
3	<b>PRECOM</b>	There is a comma immediately before the <i>Adv-mente</i>	TRUE, FALSE
4	<b>FOLCOM</b>	There is a comma immediately after the <i>Adv-mente</i>	TRUE, FALSE
5	<b>BETCOM</b>	The <i>Adv-mente</i> is between comas	TRUE, FALSE
6	<b>PRESCOM</b>	The <i>Adv-mente</i> is preceded by a semi-colon	TRUE, FALSE
7	<b>CLOSEV</b>	The number of the distance between the <i>Adv-mente</i> and the closest verb	38 different numeric values referring to the distance
8	<b>PREVPOS</b>	The closest verb appears before or after the <i>Adv-mente</i>	BEFORE, AFTER, NVERB
9	<b>1PERSPS</b>	The closest verb is in the first person (singular or plural) form	TRUE, FALSE
10	<b>SER</b>	The closest verb is the verb SER	TRUE, FALSE
11	<b>ADVMODIF</b>	There is an adverbial modifier immediately before the <i>Adv-mente</i>	TRUE, FALSE

Table 4.4: The features proposed, their respective definitions and values



Table 4.4 – continued from previous page

N°	Feature ID	Description	Values
12	<b>POS-3</b>	The third part of speech previous to the <i>Adv-mente</i>	37 different values: 36 POS tags from the Lx-Tagger and 1 created (NPOS)
13	<b>POS-2</b>	The second part of speech previous to the <i>Adv-mente</i>	33 different values: 32 POS tags from the Lx-Tagger and 1 created (NPOS)
14	<b>POS-1</b>	The part of speech the appears immediately before the <i>Adv-mente</i>	41 different values: 40 POS tags from the Lx-Tagger and 1 created (NPOS)
15	<b>POS+1</b>	The part of speech the appears immediately after the <i>Adv-mente</i>	33 different values: 32 POS tags from the Lx-Tagger and 1 created (NPOS)
16	<b>POS+2</b>	The second part of speech following the <i>Adv-mente</i>	33 different values: 32 POS tags from the Lx-Tagger and 1 created (NPOS)
17	<b>POS+3</b>	The third part of speech following the <i>Adv-mente</i>	41 different values: 40 POS tags from the Lx-Tagger and 1 created (NPOS)

Table 4.4: The features proposed, their respective definitions and values

(A) **BEGSENT** and **ENDSENT**:

Both features present a binary value (TRUE or FALSE) that indicates whether or not the *Adv-mente* is at the beginning or end of the sentence. They were proposed based on the mobility and position property of adverbs (see Section 2.1.2.2). In general, sentence-external *Adv-mente* can appear at the beginning or end of a sentence as they operate on it. In the case of sentence-internal *Adv-mente* (M) as they establish a relation with other elements inside the proposition they usually do not appear in these positions, but depending on the (M) class their position inside the proposition may vary as well. MP adverbs, for example, often appear at the beginning of a sentence.

(B) **PRECOM**, **FOLCOM**, **BETCOM** and **PRESCOM**:

All four features also present a binary value (TRUE or FALSE) and were proposed based on the mobility and position property of adverbs (see Section 2.1.2.2). As some *Adv-mente* (e.g. sentence-external adverbials) can occupy different positions in a sentence, the punctuation (e.g. comma (,) and semi-colon (;)) is used to mark their mobility. If an *Adv-mente* has a strong relation with another structure in the proposition, normally, it cannot be separated

from it by means of a punctuation mark, such as in the case of MQ. The comma can appear immediately before (PRECOM) or after the *Adv-mente* (FOLCOM). There are still cases in which the *Adv-mente* occur between commas (BETCOM) or are preceded by a semicolon mark (PRESCOM).

(C) **CLOSEV, PREVPOS, 1PERSPS and SER:**

All these features are related to the closest verb to the *Adv-mente*. The CLOSEV gives the distance between the verb and the *Adv-mente*, counting the number of tokens that exist between than in the parsed sentence. Therefore, 38 different numeric values (from 0 to 52) were found. When a sentence does not have any verb the value 0 (zero) is given. This feature was considered as an attempt to distinguish manner adverbs that modify verbs (MV) from the other classes. These adverbs have a strong relation to the verb. As a result, it is expected that the smaller values may be able to identify verbs that are very close to the *Adv-mente* which in turn may potentially be a MV, whereas the larger values of CLOSEV would characterise the other classes of *Adv-mente*.

The nearest verb can appear before (value BEFORE) or after (value AFTER) the *Adv-mente* (feature PREVPOS). This information is relevant, for instance, in the case of MV *Adv-mente* that usually occur in a postverbal position. The value NVERB means that there is no verb in the sentence.

The 1PERSPS feature is related to a morphological information of the nearest verb. It has a binary value (TRUE or FALSE) that indicates whether or not this verb is in the first person, singular or plural. The important information is the person of the verb and its grammatical number is irrelevant. This feature was established in order to identify the PS adverbs (disjunctive adverbs of style), as it was observed that they usually appear together with verbs in the first person.

The last feature of this group is SER. It has also a binary value (TRUE or FALSE) in order to verify if the lemma of the closest verb is the verb *ser* ‘to be’. It was observed during the task of building the training dataset that this verb usually occurred with some specific classes of *Adv-mente*, for instance, the MF class.

(D) **ADVMODIF**:

This feature has a binary value (TRUE or FALSE) that indicates if there is or not an adverbial modifier (e.g. *mais* ‘more’) immediately before the *Adv-mente*. It seems that some constructions of *Adv-mente* can be modified by this kind of adverbial, whereas others cannot.

(E) **POS-3, POS-2, POS-1, POS+1, POS+2 and POS+3**:

In order to analyse the context in which the different constructions of the same *Adv-mente* occurs, the part of speech (POS) of the three tokens that comes before and after the *Adv-mente* were taking into consideration. The tags of the POS are the ones used by the Lx-Tagger<sup>7</sup>. The number of tokens to be analysed was decided based on the output generated by the tagger. It would not be possible to consider, for example, only the **POS-1** and **POS+1**, since this tagger expand the contractions (e.g. *no* = *em\_/PREP a/DA#fs*). For this reason, the POS of more tokens were extracted so that more information about the context of the *Adv-mente* could be analysed.

As this tool assigns a tag to every token, the punctuation marks (PNT tag) are also tagged. This tag was not considered in the extraction. The reason for this decision is that the punctuation marks that may be relevant or not in the process of disambiguation of the *Adv-mente* are treated in features apart (see (B)). For this reason, if one or more tokens in the three positions before and the three positions after the *Adv-mente* had a PNT tag, this last tag was not extracted giving priority to all other POS tags. Considering an example (4.11) of a tagged sentence, it can be seen that there is a comma immediately before the *Adv-mente normalmente*. The feature **PRECOM** will receive the value TRUE and in the POS features the PNT tag will not be extracted.

(4.11) <p><s>*Em\_/PREP a/DA#fs natureza/NATUREZA/CN#fs*  
*,/PNT normalmente/ADV os/DA#mp corpos/CORPO/CN#mp*  
*estão/ESTAR/V#pi-3p expostos/EXPOR,EXPOSTO/PPA#mp*  
*a\_/PREP a/DA#fs ação/AÇÃO/CN#fs de/PREP*  
*forças/FORÇA/CN#fp naturais/NATURAL/ADJ#gp ,\*/PNT*

<sup>7</sup>The tagset of this tool is available at <http://lxcenter.di.fc.ul.pt/services/en/LXServicesSuite.html>.

*como/CJ o/DA#ms atrito/ATRITO/CN#ms ,\*//PNT a/DA#fs*  
*força/FORÇA/CN#fs de\_/PREP a/DA#fs*  
*gravidade/GRAVIDADE/CN#fs e/CJ o/DA#ms*  
*empuxo/EMPUXO/CN#ms .\*//PNT </s\_s>*

Table 4.5 illustrates the difference between considering (I) and not considering (II) the PNT tag in the extraction.

Example	POS-3	POS-2	POS-1	<i>Adv-mente</i>	POS+1	POS+2	POS+3
(I)	DA	CN	PNT	normalmente	CN	V	PPA
(II)	PREP	DA	CN	normalmente	CN	V	PPA

Table 4.5: Illustrative example of POS extraction

Symbols such as <p>, <s> and </s\_s> (see example (4.11)) were discarded to avoid being extracted as a POS.

Besides the other tags provided by Lx-Tagger, one tag had to be created (NPOS) to deal with the cases in which there was not a token in any of the six positions, and consequently any POS could be extracted. For this reason, the NPOS tag replaced the empty value.

## 4.6 ML Experiments

In this section, the Weka<sup>8</sup> tools and the learning algorithms (or classifiers) that were chosen to carry out the ML experiment are presented. The training and testing dataset are explained. Lastly, a set of experiments that was carried out are described. All the results of these experiments are presented and discussed in Chapter 5.

### 4.6.1 Training and Testing Data

In section 4.3, the training data built for carrying out the machine learning experiments was explained. Besides it, a test data is also necessary to be defined in order to compute the performance of classifiers (learning algorithms). There are different ways of doing it and this depends on the amount of data available.

<sup>8</sup>The version of Weka used was the 3.6.

Witten and Frank (2005, p. 146) discuss that it is a challenge to measure the accuracy of classifiers when dealing with a limited amount of data, since it is hard to define the proportion of this data that can be used for training and testing the classifiers. They argue that in such cases a certain amount from the dataset can be held over for testing and the remainder can be used for training. The data can be split into two parts, being 50% of it used for training and the other 50% for testing. Nevertheless, this technique is hardly used, since a larger part of the data should be used for training.

The authors point out that among the methods available for predicting the error rate<sup>9</sup> of classifiers, the most widely used one is the 10-fold cross validation. In this standard statistical evaluation technique, the data is randomly divided into ten parts and each part is held over for testing, whereas the other nine-tenths parts are used for training. All this process is performed ten times on different training datasets and an average of the error rate is calculated.

As the training dataset that was built (NSCSAMPLE) (see Section 4.3) does not have a vast amount of instances and there is no other dataset available, this technique seems to be the best one to be adopted in this study. As a result, the NSCSAMPLE are used as both training and testing data using the 10-fold cross validation technique.

## 4.6.2 Learning Algorithms

As it was mentioned before (see Section 2.3.1), Weka has a large variety of learning algorithms (classifiers) implemented in its system. To carry out all the experiments, nine classifiers were chosen in order to compare their results and verify which one of them had the best performance when dealing with the classification, and consequently ambiguity problem of *Adv-mente*. Their choice was motivated by the approach that they adopt and also based on the algorithms that have been applied to WSD problems described in the NLP literature such as in Manning and Schütze (2001), Stevenson and Wilks (2003) and Jurafsky and Martin (2009).

The following scheme shows the classifiers there were chosen and give a brief

---

<sup>9</sup>Error rate indicates if the classifier was successful or not in classifying correctly each instance.

description<sup>10</sup> of how they function:

- **Bayesian classifiers (Bayes):**
  - (1) NaiveBayes (NB)- it is a standard probabilistic Naïve Bayes classifier
  - (2) BayesNet (BN) - learning Bayesian nets
- **Decision Trees classifiers (Trees):**
  - (3) J48 (J48) - decision tree algorithm C4.5
  - (4) LADTree (LT) - build alternating decision trees
- **Lazy classifiers (Lazy):**
  - (5) KStar (KStar) - nearest neighbor with generalized distance function
  - (6) IBk (IBk) - k-nearest-neighbor classifier
- **Metalearning algorithms (Meta):**
  - (7) AttributeSelectedClassifier (ASC) - reduce dimensionality of data by attribute selection
  - (8) FilteredClassifier (FC) - it runs a classifier on data that has been filtered
- **Functions classifiers (Function):**
  - (9) Sequential Minimal Optimization (SMO) - sequential minimal optimization algorithm for support vector classification

Besides these nine algorithms, others are also available in Weka, but they were not considered in this present study for reasons of time and scope. More information about these classifiers are available in books such as in Iraolak (2004), Witten and Frank (2005) and Araujo (2006).

---

<sup>10</sup>These descriptions were taken from Witten and Frank (2005, pp. 404-415).

### 4.6.3 Description of Experiments

Two different experiments were undertaken using the machine learning techniques available in Weka. In both, all the classifiers previously mentioned were used, with default parameter settings provided by Weka.

The aim of these experiments are twofold. First, different learning algorithms were used to compare their performance and verify which one was able to deal better with the ambiguity problem of *Adv-mente* and which are the most significant features. Secondly, the same algorithms were applied to different amount of instances selected from the NSCSAMPLE dataset in order to verify how good the *Adv-mente* could be disambiguated when adopting different approaches. Depending on the experiment, not all the 2650 instances of all 53 ambiguous *Adv-mente* were used. This decision depended on the objective of the experiment.

#### 1. Experiment 1:

In the first experiment, each ambiguous *Adv-mente* was analysed individually. As a result, each classifier was applied to a sample of only 50 instances. This is the reason why the same number of sentences were selected from the NILC/São Carlos corpus (see Section 4.3), so that it could be possible to analyse the same number of *Adv-mente*.

From the 53 *Adv-mente* (2650 instances), 10 of them (namely *exclusivamente*, *infelizmente*, *altamente*, *juntamente*, *originalmente*, *felizmente*, *consequentemente*, *criminalmente*, *simplesmente* and *precisamente*) were disregarded in this experiment due to the lack of examples in at least one of the classes (see Table 4.2). This means that a total of 40 ambiguous *Adv-mente* were used in this first experiment. Thirty-one of them present 2 possible constructions, whereas the remaining twelve have 3 possible constructions<sup>11</sup> the number of classes that were being considered varied according to the *Adv-mente* that was being analysed.

---

<sup>11</sup>There were only 2 *Adv-mente* (*simplesmente* and *precisamente*) from the 53 originally selected to carry out the machine learning experiments that presented 4 possible constructions. They were disregarded in this experiment due to the lack of examples to constitute one of their 4 classes.

This experiment was carried out in order to explore different classifiers for the disambiguation of *Adv-mente* and evaluate their performance.

## 2. Experiment 2:

The second experiment took into consideration all the 53 *Adv-mente* (and consequently all the 2650 instances of the NSCSAMPLE) and the 11 classes at the same time (see Table 4.2). The classifiers with the best performance verified in the first experiment were applied to all the 2650 instances in order to check the effectiveness of the features proposed and the accuracy of the classification of *Adv-mente* when considering all classes at the same time.

### 4.6.4 Evaluation: baseline and other measures

To evaluate the performance of these algorithms and the results, the experiments were analysed separately due to the different aims of each one and the way of using the 2650 instances.

For the experiment 1, the baseline was defined based on the Most Frequent Sense (MFS) criteria which is one of the simplest baseline approaches that is frequently adopted in WSD literature. It consists in assigning the most frequent sense for each ambiguous word taking into consideration its training set of occurrences. Hence, the MFS baseline of the *Adv-mente* was calculated based on the random sample of 50 occurrences of each word extracted from the NILC/São Carlos corpus. Each occurrence of *Adv-mente* was manually classified in order to determine its construction (see Table 4.2) based on the classification previously established in Chapter 3.1. Table 4.6 shows the MFS baseline of the 43 *Adv-mente* (1462 instances) used in this experiment.

The baseline results varied according to each *Adv-mente* in a range from 0.40 to 0.96. Even though six *Adv-mente*, namely (*especialmente*, *pessoalmente*, *curiosamente*, *conjuntamente*, *justamente* and *seguramente*), presented a very high baseline result, they were considered for this experiment. The lowest baseline belongs to the *Adv-mente decididamente* 0.40. The average of the baseline of the 43 *Adv-mente* is equal to 0.68.

Regarding the experiment 2, as all the 2650 instances (53 *Adv-mente*) were



<i>Adv-mente</i>	Class	Count	MFS	<i>Adv-mente</i>	Class	Count	MFS
especialmente	MF	46	0.92	mentalmente	MV	28	0.56
normalmente	PAh	30	0.60	firmemente	MS	32	0.64
igualmente	MV	27	0.54	comercialmente	MV	39	0.78
oficialmente	MV	43	0.86	genericamente	MV	34	0.68
pessoalmente	MF	45	0.90	brevemente	MT	30	0.60
claramente	MV	41	0.82	inversamente	MV	26	0.52
independentemente	PC	43	0.86	conjuntamente	MV	46	0.92
perfeitamente	MQ	30	0.60	emocionalmente	MP	27	0.54
propriamente	MV	37	0.74	visualmente	MP	28	0.56
profundamente	MV	29	0.58	realmente	MV	26	0.52
fortemente	MQ	37	0.74	exatamente	MF	38	0.76
essencialmente	MQ	29	0.58	justamente	MF	46	0.92
formalmente	MV	39	0.78	naturalmente	PAm	29	0.58
paralelamente	MV	32	0.64	particularmente	MF	33	0.66
sexualmente	MP	29	0.58	verdadeiramente	MQ	37	0.74
curiosamente	PAa	48	0.96	seriamente	MQ	31	0.62
ligeiramente	MQ	29	0.58	virtualmente	MQ	37	0.74
rigorosamente	MV	33	0.66	seguramente	PAm	46	0.92
levemente	MQ	30	0.60	francamente	MQ	22	0.44
duramente	MV	28	0.56	decididamente	MS	20	0.40
sinceramente	PS	28	0.56	superficialmente	MV	40	0.80
profissionalmente	MV	34	0.68	<b>Average</b>	-	-	0.68

Table 4.6: MFS Baseline

taken into consideration, a single baseline that was calculated based on the average of the MFS baseline of each one of these 53 adverbs. As a result, the baseline for this experiment is equal to 0.74.

Standard evaluation measures such as *precision*, *recall*, *f-measure* and *accuracy* were also used in both experiments. Their respective definition has already been explained in Section 2.3. Accuracy<sup>12</sup> may overestimate slightly the results, and consequently may not reflect the exact performance of the classifier. It gives a general result without considering the small cases Manning and Schütze (2001). *Precision* and *recall* are more sensitive to all cases, hence they present a more accurate result. The *f-measure* is a combination of these two last scores and may present the general performance. For this reason, these other measures were also considered.

<sup>12</sup>Accuracy differs from the precision measure, because it gives a measure for the overall performance of the classifier. To obtain the accuracy score the total number of correctly classified instances is divided by the total number of instances. Precision is here defined as the number of correctly classified instances of a specific class divided by the number of instances classified as belonging to this class (Manning and Schütze, 2001).

# Chapter 5

## Results and Discussion

This chapter presents the results obtained in each one of the two experiments. All experiments were performed using a 10-fold-cross validation. The evaluation measures described in the previous chapter (see Section 4.6.4) are used to assess the performance of the classifiers in order to determine which ones present the highest scores (experiment 5.1 and experiments 5.2). Depending on the results first results, other tests were undertaken in each experiment to verify the effectiveness of the features proposed and the 10-fold-cross-validation technique firstly adopted.

### 5.1 Experiment 1

A set of experiments were undertaken to determine the most accurate method for the disambiguation of each *Adv-mente*. As 10 from the 53 *Adv-mente* originally selected were disregarded due to the lack of examples in one their possible constructions, a total of 40 *Adv-mente* were used in this experiment. All the seventeen features proposed were considered in this first task. The accuracy obtained for the performance of each one of the 53 *Adv-mente* is given in Table 5.1. These results are compared to the MFS baseline previously defined for each *Adv-mente* (see Section 4.6.4), so that it is possible to determine which classifier(s) had the best performance(s).

<b>Adv-mente</b>	<b>MFS</b>	<b>NB</b>	<b>BN</b>	<b>J48</b>	<b>LT</b>	<b>KStar</b>	<b>IBk</b>	<b>ASC</b>	<b>FC</b>	<b>SMO</b>
especialmente	0.92	<b>0.96</b>	0.92	0.88	0.84	0.92	<b>0.96</b>	0.88	0.88	0.88

Table 5.1: Weka classifiers accuracy: 1st experiment

Table 5.1 – continued from previous page

Adv-mente	MFS	NB	BN	J48	LT	KStar	IBk	ASC	FC	SMO
normalmente	0.60	<b>0.82</b>	<b>0.82</b>	0.68	0.66	0.74	0.66	0.66	0.64	0.80
igualmente	0.54	0.66	<b>0.74</b>	0.56	0.54	0.66	0.66	0.58	0.56	0.52
oficialmente	0.86	0.88	<b>0.92</b>	0.78	0.80	0.88	0.84	0.80	0.84	0.86
pessoalmente	0.90	0.92	0.94	0.92	0.88	<b>0.96</b>	<b>0.96</b>	0.92	0.92	0.94
claramente	0.82	0.80	0.80	0.78	0.86	0.78	0.74	0.78	0.78	<b>0.90</b>
independentemente	0.86	0.90	<b>0.92</b>	<b>0.92</b>	0.86	0.88	0.88	0.86	<b>0.92</b>	<b>0.92</b>
perfeitamente	0.60	0.66	0.72	<b>0.82</b>	0.74	0.72	0.70	<b>0.82</b>	<b>0.82</b>	0.60
propriamente	0.74	0.80	0.80	<b>0.94</b>	0.88	0.80	0.80	<b>0.94</b>	<b>0.94</b>	<b>0.94</b>
profundamente	0.58	0.94	0.92	0.86	0.94	0.94	0.94	0.86	0.86	<b>0.96</b>
fortemente	0.74	0.84	0.84	0.84	<b>0.92</b>	0.86	0.88	0.84	0.84	0.84
essencialmente	0.58	0.86	0.88	0.88	0.88	<b>0.96</b>	<b>0.96</b>	0.88	0.80	0.90
formalmente	0.78	0.82	0.74	0.72	<b>0.92</b>	0.88	0.86	0.72	0.72	0.82
paralelamente	0.64	0.80	0.80	0.84	0.72	0.70	0.64	<b>0.90</b>	0.82	0.72
sexualmente	0.58	0.92	0.86	0.92	0.94	0.90	<b>0.96</b>	0.90	0.84	0.92
curiosamente	0.96	0.96	0.96	0.96	0.96	<b>1.00</b>	<b>1.00</b>	0.96	0.96	<b>1.00</b>
ligeiramente	0.58	0.90	<b>0.92</b>	0.90	0.80	0.90	0.90	0.90	0.90	0.90
rigorosamente	0.66	0.94	0.94	0.94	<b>0.98</b>	0.96	0.94	0.88	0.90	<b>0.98</b>
levemente	0.60	0.82	0.78	0.74	<b>0.88</b>	0.78	0.76	0.60	0.74	0.80
duramente	0.56	0.92	0.94	0.90	0.86	0.92	0.90	0.90	0.90	<b>0.98</b>
sinceramente	0.56	0.80	0.82	<b>0.90</b>	0.86	0.78	0.78	<b>0.90</b>	<b>0.90</b>	0.76
profissionalmente	0.68	0.80	0.80	0.86	<b>0.90</b>	0.78	0.78	0.86	0.86	0.86
mentalmente	0.56	0.72	0.74	0.68	0.76	0.74	0.66	0.68	<b>0.82</b>	0.66
firmemente	0.64	0.72	0.74	0.58	0.70	0.66	0.62	0.58	0.58	<b>0.82</b>
comercialmente	0.78	<b>0.86</b>	0.80	0.76	0.82	0.84	<b>0.86</b>	0.72	0.76	0.78
genericamente	0.68	<b>0.74</b>	<b>0.74</b>	0.72	0.68	0.72	0.72	0.72	0.72	<b>0.74</b>
brevemente	0.60	0.64	<b>0.66</b>	0.60	0.64	0.48	0.48	0.60	0.60	0.58
inversamente	0.52	<b>0.98</b>	<b>0.98</b>	<b>0.98</b>	<b>0.98</b>	0.96	0.96	<b>0.98</b>	<b>0.98</b>	<b>0.98</b>
conjuntamente	0.92	<b>0.92</b>	<b>0.92</b>	0.90	0.88	<b>0.92</b>	<b>0.92</b>	0.90	<b>0.92</b>	<b>0.92</b>
emocionalmente	0.54	0.74	<b>0.76</b>	0.66	0.68	0.70	0.70	0.66	0.66	0.72
visualmente	0.56	0.70	0.76	0.78	0.76	<b>0.86</b>	<b>0.86</b>	0.70	0.74	0.70
realmente	0.52	0.82	0.82	0.92	<b>0.94</b>	0.80	0.76	0.78	0.92	0.86
exatamente	0.76	0.76	<b>0.80</b>	0.76	<b>0.80</b>	0.70	0.70	0.76	0.76	0.78
justamente	0.92	0.90	<b>0.92</b>	<b>0.92</b>	0.88	0.90	0.90	<b>0.92</b>	<b>0.92</b>	<b>0.92</b>
naturalmente	0.58	0.72	<b>0.80</b>	0.60	0.62	0.76	0.72	0.58	0.60	0.74
particularmente	0.66	0.90	0.92	<b>0.96</b>	0.92	0.86	0.88	0.92	0.78	0.94
verdadeiramente	0.74	0.88	<b>0.90</b>	0.86	0.84	0.84	0.84	0.82	<b>0.90</b>	0.84
seriamente	0.62	0.70	0.64	0.56	0.68	<b>0.72</b>	0.70	0.60	0.56	0.52
virtualmente	0.74	0.78	0.80	0.72	<b>0.86</b>	0.82	0.76	0.74	0.72	0.82
seguramente	0.92	0.92	0.92	0.92	0.90	0.90	0.86	0.92	0.92	<b>0.96</b>
francamente	0.44	0.70	0.64	0.66	0.68	0.60	0.58	0.64	<b>0.72</b>	0.64
decididamente	0.40	0.70	0.64	<b>0.74</b>	0.64	0.60	0.60	0.72	0.70	0.62
superficialmente	0.80	0.74	<b>0.82</b>	0.80	0.80	0.80	0.80	0.80	0.80	0.74
<b>Average</b>	0.68	0.82	0.83	0.81	0.82	0.81	0.80	0.79	0.80	0.82

Table 5.1: Weka classifiers accuracy: 1st experiment

By comparing the accuracy average of each algorithm and the average of MFS

baseline, all learning algorithms presented a higher performance in comparison to the baseline. The average score of all nine classifiers is 0.81.

The BayesNet (BN) presented the highest accuracy score: 0.83. Three other classifiers, namely NaiveBayes (NB), LADTree (LT) and Sequential minimal optimization (SMO) classifiers, presented the same accuracy score (0.82), being the second best classifiers. AttributeSelectedClassifier (ASC) presented the worst performance (0.79), but still remained above the baseline.

Considering, for example, the precision, recall and f-measure average scores, Table 5.2 presents only the results of the four best classifiers previously mentioned:

<b>Classifier</b>	<b>Precision</b>	<b>Recall</b>	<b>F-measure</b>
<b>BayesNet</b>	0.80	0.82	0.81
<b>Naivebayes</b>	0.80	0.83	0.81
<b>LADTree</b>	0.81	0.82	0.81
<b>SMO</b>	0.81	0.82	0.81

Table 5.2: The four best classifiers: precision, recall and f-measure

It is worth to note that these scores obtained by each one of these four classifiers are slightly lower and different from the accuracy measure presented in table 5.1. This is explained by the fact that these three measures are more sensitive to the small cases of the training dataset, reflecting more precisely the results (Manning and Schütze, 2001, pp. 267-271). Even considering these measures, again the performance of the classifiers remain very similar.

In sum, the overall performances of all classifier are very similar (from 0.79 to 0.83 in terms of accuracy) and this makes the decision of choosing only one as the best classifier difficult.

The standard deviation<sup>1</sup> is measure that can be used to evaluate how much variation there is among the accuracy scores of the classifiers regarding each *Adv-mente*. Table 5.3 illustrates the results obtained:

<b>Adv-mente</b>	<b>Standard Deviation of Classifiers</b>
especialmente	0.041
normalmente	0.075
igualmente	0.074
oficialmente	0.046

Table 5.3: Standard Deviation of Classifiers for each *Adv-mente*

<sup>1</sup>Standard deviation is the square root of the variance (Manning and Schütze, 2001).

Table 5.3 – continued from previous page

Adv-mente	Standard Deviation of Classifiers
pessoalmente	0.025
claramente	0.048
independentemente	0.026
perfeitamente	0.077
propriamente	0.070
profundamente	0.041
fortemente	0.028
essencialmente	0.049
formalmente	0.077
paralelamente	0.081
sexualmente	0.037
curiosamente	0.020
ligeiramente	0.035
rigorosamente	0.033
levemente	0.076
duramente	0.033
sinceramente	0.057
profissionalmente	0.044
mentalmente	0.053
firmemente	<b>0.085</b>
comercialmente	0.049
genericamente	0.019
brevemente	0.066
inversamente	<b>0.009</b>
conjuntamente	0.015
emocionalmente	0.037
visualmente	0.063
realmente	0.066
exatamente	0.037
justamente	0.015
naturalmente	0.082
particularmente	0.053
verdadeiramente	0.029
seriamente	0.074
virtualmente	0.049
seguramente	0.026
francamente	0.045
decididamente	0.053
superficialmente	0.028

Table 5.3: Standard Deviation of Classifiers for each *Adv-mente*

For these 43 *Adv-mente*, the results of the standard deviation concerning the performance of the classifiers varied according to each word in a range from 0.009 to 0.085.

The lowest standard deviation result belongs to the performance of the classifiers of the adverb *inversamente* (0.009). The low results obtained show that there is not a significant difference among the performance of these algorithms for each adverb.

If the standard deviation for every classifier and also for the MFS are calculated, the following results (see Table 5.4) are obtained:

Classifiers	MFS	NB	BN	J48	LT	KStar	IBk	ASC	FC	SMO
Standard deviation	0.143	0.095	0.093	0.121	0.111	0.115	0.125	0.121	0.116	0.130

Table 5.4: Standard Deviation of the performance of each Classifier

Based on these results, it can be inferred that the classifiers have less variation than the MFS (between 0.02 and 0.05 less) and they do not show much difference (0.037) regarding the standard deviation among each other.

Moreover, the standard deviation scores in this case are higher than the standard deviation scores per each adverb seen Table 5.3. The scores presented in Table 5.4 show that there is a higher difference among the results when evaluating each classifier. This means that the adverbs used are an important factor that may influence in the algorithm's performance.

If each classifier is analysed individually, it can be inferred that most of the classifiers had a performance equal or above the baseline (of each adverb) in terms of accuracy score in most of the cases. Few of them performed below the baseline, for example, in the case of the *Adv-mente comercialmente* regarding the algorithms J48 (0.76), AttributeSelectedClassifier (ASC; 0.72) and FilteredClassifier (FC; 0.76). It is difficult to establish a pattern to indicate which algorithms usually present lower performances, as this may depend on each word (its possible constructions), the amount of instances available, the approach adopted by the classifier, the techniques of splitting the dataset for training and test or even the features used.

The *Adv-mente* mentioned before (*comercialmente*) is an adverb that can be classified as MP or MV, in other words, it has two possible constructions. From the 50 instances selected for this adverb, 11 were manually classified as being MP and the remaining 39 as MV. Comparing it with the adverb *profesionalmente*, that also presents the same ambiguity class (MP,MV) and a similar distribution

of instances per class, more precisely 16 instances of MP and 34 instances of MV (see Table 4.2), it is worth to note that in the case of this last adverb all classifiers performed above the MFS baseline of the adverb. As the only different factor in the experiment was the actual instances of these adverbs in the NSCSAMPLE, a possible explanation for the differences in the results must be related to the more or less characteristic context of each adverb in each set of sentences.

Regarding the features used, as an attempt to verify which from the 17 features proposed (see section 4.5) were the most effective ones the Weka's Attribute Selection tool was used. In this tool, some algorithms (e.g. CfsSubEval and ChiSquareAttributeEval) and search methods (e.g. Best first and ExhaustiveSearch) are available. These algorithms function as evaluators of the attributes (e.g. to evaluate an attribute set) and the methods (e.g. exhaustive search) define the way that these features will be searched. For this reason, it is necessary to combine these two information in order to identify the most effective features according to the specific combination of an evaluator and a method selected.

Among the algorithms available the CfsSubsetEval was chosen to perform this task and the method used was the one provided by default in Weka, namely Best-First. This algorithm analyse a subset of attributes (features) and return a numeric measure that guides the search method. As a result, it provides a set of attributes that are very connected with the class, but are not so interconnected among them.

For this task, only seven from the 43 *Adv-mente* were used. They were chosen according to the classes that they could be assigned to and with the aim of having examples of all the 11 classes proposed. The results are given in table 5.5.

It can be inferred from the features selected by the Weka's CfsSubsetEval algorithm that the feature POS in all the seven cases was selected, varying only in the position of the this part-of-speech (-3, -2, -1, 1, 2, 3). However, in most of the cases, the closest part-of speeches (POS-1 and POS+1) were the ones selected. The features ENDSSENT and BEGINSSENT were selected in three cases, showing that the mobility of the adverb is a good feature to be taken into consideration. The features related to the punctuation (FOLCOM, PRECOM) were selected in cases in which there was a sentence external-modifier class (P), confirming that probably the commas are u to separate sentence external-modifiers from the main clause. The features related to the verb, namely PREVPOS and CLOSEV, were

Adv-mente	Classes	N° of features	Selected Features
sexualmente	MP, MV	5	ENDSENT POS-1 POS+1 CLOSEV PREVPOS
essencialmente	MQ, MF	2	POS-1 POS+1
brevemente	MS, MT	4	POS-3 POS-2 POS+1 1PERSONPS
inversamente	PC, MV	4	FOLCOM POS+1 CLOSEV PREVPOS
superficialmente	PAa, MQ, MV	4	BEGSENT ENDSSENT PRECOM POS-1
naturalmente	PAm, MS, MV	6	PRECOM FOLCOM POS-3 POS-1 POS+1 PREVPOS
francamente	PS, MQ, MS	3	FOLCOM POS+1 CLOSEV

Table 5.5: Selected features: Weka's CfsSubsetEval

selected, as it was expected, in the cases in which there was an MV class (proper manner adverbs with a strong relation to the verb). The only exception for this last feature was the word *francamente* that cannot be classified as an MV, but even though used this feature. The feature 1PERSONPS was selected by only one word (*brevemente*). This was a surprising choice as this last feature was proposed based on the behavior of the PS class and *brevemente* cannot be assigned to it.

A last set of experiments was carried out to evaluate the performance of the classifiers considering only the features selected (see Table 5.5) for each *Adv-mente* by the combination of the CfsSubsetEval evaluator and BestFirst method. In this task, the same nine algorithms used in the first task were applied to a set of instances (50 instances) of each one of the seven *Adv-mente* listed in Table 5.5.

The performance of all classifiers (in terms of accuracy) for each one of the seven *Adv-mente* in this task are given in Tables 5.6, 5.7 and 5.8. In addition, they are compared to the accuracy scores obtained for the same *Adv-mente* in the first task carried out (see Table 5.1). This comparison is made with the aim of evaluating how classifiers performed when all the 17 features were used (1st task) and when only the selected features of each *Adv-mente* were considered (present task).

These tables have two columns for each one of the nine algorithms. The first column, in which the number (1) appears in the header, corresponds to the accu-



<b>Adv-mente</b>	<b>MFS</b>	<b>NB (1)</b>	<b>NB (2)</b>	<b>BN (1)</b>	<b>BN (2)</b>	<b>J48 (1)</b>	<b>J48 (2)</b>
sexualmente	0.58	0.92	0.94	0.86	0.90	0.92	0.92
essencialmente	0.58	0.86	0.94	0.88	0.94	0.88	0.88
brevemente	0.60	0.64	0.76	0.66	0.74	0.60	0.60
inversamente	0.52	0.98	0.96	0.98	0.96	0.98	0.98
superficialmente	0.80	0.74	0.88	0.82	0.86	0.80	0.80
naturalmente	0.58	0.72	0.82	0.80	0.80	0.60	0.58
francamente	0.44	0.70	0.72	0.64	0.64	0.66	0.64

Table 5.6: Classifiers' performance comparison - NB, BN and J48

<b>Adv-mente</b>	<b>MFS</b>	<b>LT (1)</b>	<b>LT (2)</b>	<b>KStar (1)</b>	<b>KStar (2)</b>	<b>IBk(1)</b>	<b>IBk(2)</b>
sexualmente	0.58	0.94	0.96	0.90	0.92	0.96	0.94
essencialmente	0.58	0.88	0.94	0.96	0.94	0.96	0.94
brevemente	0.60	0.64	0.70	0.48	0.72	0.48	0.66
inversamente	0.52	0.98	0.98	0.96	0.96	0.96	0.94
superficialmente	0.80	0.80	0.84	0.80	0.88	0.80	0.82
naturalmente	0.58	0.62	0.72	0.76	0.80	0.72	0.78
francamente	0.44	0.68	0.62	0.60	0.72	0.58	0.64

Table 5.7: Classifiers' performance comparison - LT, KStar and IBk

racy scores obtained in the first task and that are also shown in Table 5.1. The second column (it has the number (2) in the header) gives the accuracy scores of the performance of the classifiers in this task.

This contrastive analysis shows that even though most of the classifiers had a performance above the baseline (MFS) when considering all the 17 features for each word (1st task), many of them performed slightly better (or at least equal) when only the features selected for each *Adv-mente* were used. Few had a performance worse than in the first task. The *Adv-mente brevemente* and *inversamente* are good illustrative examples of different results obtained.

Regarding the adverb *brevemente*, six (NB, BN, LT, KStar, IBk and SMO) of

<b>Adv-mente</b>	<b>MFS</b>	<b>ASC (1)</b>	<b>ASC (2)</b>	<b>FC (1)</b>	<b>FC (2)</b>	<b>SMO (1)</b>	<b>SMO (2)</b>
sexualmente	0.58	0.90	0.90	0.84	0.84	0.92	0.94
essencialmente	0.58	0.88	0.88	0.80	0.88	0.90	0.94
brevemente	0.60	0.60	0.60	0.60	0.60	0.58	0.68
inversamente	0.52	0.98	0.98	0.98	0.98	0.98	0.96
superficialmente	0.80	0.80	0.80	0.80	0.80	0.74	0.80
naturalmente	0.58	0.58	0.58	0.60	0.58	0.74	0.76
francamente	0.44	0.64	0.64	0.72	0.68	0.64	0.72

Table 5.8: Classifiers' performance comparison - ASC, FC, SMO

nine algorithms performed better when only the features selected for this adverb were considered. Nevertheless, the remaining three classifiers (J48, ASC,FC) did not presented any improvement in their performances in comparison to the results of the same classifiers in the first task.

In the case of *inversamente*, none of the classifiers presented better results. Four (NB, BN, IBk and SMO) of them had exactly the same performance in comparison to the results of the first task, whereas the remainder (J48, LT, KStar, ASC and FC) had an accuracy score worse than the same algorithms used with all 17 features.

This last task shows that the most effective features depends on each *Adv-mente* and also the algorithm used.

## 5.2 Experiment 2

After training each one of the nine classifiers on all instances of the NSCSAMPLE, their performance was evaluated. Table 5.9 presents the accuracy score of each classifier in comparison with the MFS baseline (average) of the NSCSAMPLE previously established in Section 4.6.4. The statistical 10-fold cross-validation was used in the evaluation.

<b>MFS</b>	<b>NB</b>	<b>BN</b>	<b>J48</b>	<b>LT</b>	<b>KStar</b>	<b>IBk</b>	<b>ASC</b>	<b>FC</b>	<b>SMO</b>
0.74	0.47	0.47	0.49	0.48	0.48	0.45	0.48	0.49	0.50

Table 5.9: Weka classifiers accuracy: 2nd experiment *Adv-mente*

The difference among the accuracy results of the learning algorithms was marginal, varying from 0.45 to 0.50. The k-nearest-neighbor classifier (IBk: 0.45) presented the worst performance, whereas the Sequential minimal optimization (SMO: 0.59) classifier had the best one. Based on these results, it is possible to conclude that using the entire dataset to train the classifiers degrades the result in a very significant way, resulting in almost 30% less accuracy than the baseline. This seems an indication that the WSD task of *Adv-mente* must be undertaken with reference to each lexical item and not to adverbs as a whole.

By analysing further the results provided by Weka, it was observed that, in some cases, seven from the nine learning algorithms were not able to classify cor-

rectly any of the instances of an entire class. This is one of the possible explanations that could justify why the classifiers performed below the MFS baseline in terms of accuracy as it has been shown in Table 5.9.

When evaluating individually the result obtained for each class in this experiment, these seven learning algorithms obtained a precision score equal to 0.00 (zero) regarding some specific classes. Table 5.10 illustrates the classes that were not used by each classifier:

<i>Classifier</i>	<i>Classes with precision score = 0.00</i>
<b>NB</b>	MT and PAh
<b>BN</b>	MT and PAh
<b>J48</b>	PAh
<b>LT</b>	MP, MT and PAh
<b>KStar</b>	-
<b>IBk</b>	-
<b>ASC</b>	MS, MT, PAh
<b>FC</b>	MT and PAh
<b>SMO</b>	MT and PAh

Table 5.10: Classifiers: Classes with precision score equal to 0.00 *Adv-mente*

In general, the seven classifiers that presented this issue were not able to assign correctly any instance to the PAh (disjunctive adverbs of attitude: habit) class. Many of them also had the same problem with MT (time adverbs), MP (view-point adverbs) and MS (subject-oriented manner adverbs) classes. The number of instances of these classes are one of the lowest ones in the NSCSAMPLE, and consequently, may not have provided enough examples to be trained and tested by the learning algorithms in Weka.

Despite PAh, MT, MP and MS classes having few instances in the NSCSAMPLE (see Table 3.2), other classes also presented a low frequency in this sample such as PS (75 instances), but the classifiers were able to assign correctly some instances to this last class. In addition, there were two learning algorithms, namely the lazy learners KStar and IBk, that classified the instances into all 11 classes, even though they had assigned wrongly many other instances, hence performing below the MFS baseline. This last point indicates that probably the approach<sup>2</sup>

<sup>2</sup>Witten and Frank (2005) call lazy learners the classifiers known by adopting an instance-based learning approach. The authors explain that this kind of algorithms, basically, use a nearest-neighbor method and are known for performing better than other classifiers when they are combined with noisy *exemplars* pruning and attribute weighting (Witten and Frank, 2005, pp. 235-243).

used by the lazy learners is better than the one used by the other Bayesian (NB and BN), metalearning (ASC and FC) and functions (SMO) classifiers.

It is worth noting that if another measuring technique is used, rather than 10-fold cross-validation statistical, the results may be different. An attempt to use the leave-one-out cross-validation was made. In this technique,  $n$ -fold cross-validation is used and  $n$  corresponds to total number of instances in the dataset built. Each instance of the dataset is held out while the algorithm is trained in all the remaining instances. Unfortunately, it was not possible to obtain results from the leave-one-out cross-validation due to hardware limitations.

It was decided to split the dataset (NSCSAMPLE) so that 22% of its instances were used to train the learning algorithm and the remainder for testing. The nine algorithms were used to perform this task. The accuracy score of each classifier is given in Table 5.11 and it is compared to the MFS baseline (average) of all 53 *Adv-mente*.

<b>MFS</b>	<b>NB</b>	<b>BN</b>	<b>J48</b>	<b>LT</b>	<b>KStar</b>	<b>IBk</b>	<b>ASC</b>	<b>FC</b>	<b>SMO</b>
0.74	0.44	0.43	0.46	0.46	0.41	0.39	0.46	0.46	0.44

Table 5.11: Weka classifiers accuracy: 22% split for training

The difference among the performance of the classifiers varied in a range from 0.39 to 0.46, which is a small variance. Four classifiers (J48,LT, ASC,FC) obtained the same accuracy results (0.46), while one learning algorithm, namely IBk, presented the worst score 0.39.

These last results show that this method is worse than the 10-fold cross-validation (see Table 5.9). However, in both cases, all classifiers had performances below the MFS baseline (0.74).

The same problem presented in the 10-fold cross-validation technique regarding the wrong classification of all instances of an entire class was also observed when adopting percentage split method (in this case 22%). The difference is that while in the former (see Table 5.10) about four classes (PAh, MT, MP and MS) were disregarded by some algorithms, in the latter method only one class, namely PAh, was disregarded by all classifiers. It is possible to conclude that if different portions of the dataset are used to train the algorithm and there are few instances to represent certain classes, different results regarding the classification of the in-

stances in the dataset may be obtained. Nevertheless, when adopting the 10-fold cross-validation, the general performance of the classifiers was better, because more instances were correctly classified, even though some classes were disregarded.

One last task was performed to determine the most effective features when dealing with all instances of the NSCSAMPLE. The algorithm (CfsSubsetEval) and method (BestFirst) used in the experiment 1 (see Section 5.1) were also adopted in this experiment. The features selected by the Weka's CfsSubsetEval algorithm were the following:

- PRECOM FOLCOM POS-1 POS+1 1PERSONPS

A set of experiments were undertaken with all nine algorithms and all instances of the NSCSAMPLE, but instead of considering all the 17 features initially proposed, only the five features previously mentioned were used. Table 5.12 shows the accuracy, precision, f-measure and recall scores of each algorithm.

<b>Classifier</b>	<b>Accuracy</b>
<b>NB</b>	0.48
<b>BN</b>	0.48
<b>J48</b>	0.49
<b>LT</b>	0.47
<b>KStar</b>	0.48
<b>IBk</b>	0.49
<b>ASC</b>	0.48
<b>FC</b>	0.49
<b>SMO</b>	0.49

Table 5.12: Selected features: performance of the classifiers

Based on these last results, it is possible to conclude that no significant improvement in the performance of the classifiers were observed when comparing the first task that was carried out with all 17 features (see Table 5.9) and this last one in which only the five features selected were considered. In the former, the accuracy scores of the classifiers varied in a range from 0.45 to 0.50, while in the latter task the results varied from 0.47 to 0.49. This last task shows that even trying to establish the most effective features to improve results of the learning algorithms, it is difficult to obtain high accuracy scores when dealing with the entire dataset to train the classifiers in order to classify instances of ambiguous *Adv-mente*.

## Chapter 6

# Conclusions and Future Work

This dissertation had presented a proposal for classifying the Brazilian Portuguese *-mente* ending adverbs based on the French *-ment* ending adverbs study undertaken by Molinier and Levrier (2000). The 532 most frequent *Adv-mente* from the NILC/São Carlos corpus were manually classified using the formal (syntactic) criteria presented by these authors. While some *Adv-mente* were non-ambiguous, 98 were classified into two or more classes, each corresponding to a different construction and, thus, to a different meaning. The aim of this dissertation is to explore machine learning methods to help disambiguating these words.

In order to automatically classify, and consequently disambiguate these adverbs, a supervised machine learning approach was adopted. To carry out the machine learning experiments, the Weka. This tool provides many learning algorithms and data preprocessing tools available that can be applied to a training dataset previously built. It was decided to analyse the adverbs with a frequency equal or higher to 50 in the NILC/São Carlos corpus. Hence 53 *Adv-mente* were chosen. To build the training dataset, 50 instances (sentences) per each one of the 53 *Adv-mente* were randomly selected, so that a total of 2650 instances (sentences) were taken from the corpus.

This data was hand-labeled to assign the classes to each adverb and afterwards a POS tagger (LxTagger) was used to assign contextually appropriate grammatical descriptors to words in the sentences. Then, a set of features and their respective values were proposed in order to be extracted for WSD. A dataset was built from

the sample of instances to be processed by the ML toolkit Weka.

For the experiments, nine classifiers (learning algorithms) and a 10-fold cross-validation technique were used. The first experiment considered 43 ambiguous *Adv-mente*. Each classifier was trained with the sample of instances of each adverb. The results were encouraging, as all classifiers have a performance above the baseline (Most Frequent Sense - MFS) established for each adverb. However, the accuracy average of the performance of these classifiers did not varied, so it was hard to say which one of them has the best performance.

In the second experiment, all the 53 *Adv-mente*, their respective instances (2650 in total) and all the eleven classes were taken into consideration. Unlike the first experiment, the results were not positive, as all the nine classifiers performed below the baseline established. In conclusion, the best approach to deal with these adverbs was the one in which a classifier was built to each word, showing that WSD of *Adv-mente* is mainly a lexical problem.

In both experiments, an attempt was made to determine the most effective features based on the ones previously proposed. In order to do this, the Weka's Attribute Selection tool was used. This tool provides different algorithms and search methods that combined evaluate all features and the dataset with the purpose of identifying the most useful features. After this task, in both experiments the classifiers were trained again with a specific amount of instances taken from the dataset, considering only the most potential features previously identified for each case. While in the first experiment the results obtained were satisfactory, in the second experiment they were not significant.

Based on these experiments, it seems that the most effective features proposed were related to the part-of-speech of the words close to the *Adv-mente*, especially the ones that come in the positions immediately before and after the adverb (POS-1 and POS+1). The features related to the position of the adverb (BEGSENT and ENDSSENT) in the sentence and the presence/absence of comma(s) (PRE-COM and FOLCOM) were also identified as being useful features. The CLOSEV and PREVPOS seemed to successfully help in the classification of proper manner adverbs (MV class). The other features must be reviewed, as they may not be sufficiently relevant to disambiguate the adverbs.

Some general conclusions can be drawn based on both experiments and per-

formance of some classifiers. It was observed that the result of these experiments could be affected by: (a) the lack of instances to represent certain classes, (b) the method used by each classifier and its respective parameters, (c) the technique used to measure the error rate of a classifier (e.g. percentage split), (d) the portions of the dataset used to train and test the learning algorithms and (d) the lack of relevant features to handle certain classes (e.g. MT - adverbs of time).

This study was a first attempt to classify and treat automatically the Brazilian Portuguese *Adv-mente* based on a systematic syntactic-semantic linguistic approach and features observed in their context. In spite of the fact that a relative small amount of instances was used and not so many features were proposed, the general results were satisfactory.

There are still much work to be done, but the results show that it is possible to treat the *Adv-mente* computationally and that this information may help, for example, Machine Translations (MT) systems and parsers, in which the disambiguation are a very important part of their systems in order to improve their performance. For example, MT systems can make use of the information regarding the *Adv-mente* presented in this work in order to improve the accuracy of the translations of these adverbs in context. Taking into consideration the parsers, the information of the *Adv-mente* can be included in these systems to provide an in-depth analysis of adverbs and of their relation with other words in a sentence.

Finally, the analysis here provided can also be extended to study the European Portuguese *-mente* ending adverbs. If only the linguistic field is considered, the classification of *Adv-mente* proposed in the present work could be used to improve the treatment given to these words in dictionaries, grammars and textbooks.

## 6.1 Future Work

Much work still needs to be done in order to continue this study and to deepen the information about *Adv-mente*.

This study selected instances only from one corpus, namely NILC/São Carlos. This corpus is mostly composed of texts from a Brazilian newspaper known as Folha de São Paulo. Therefore, the *Adv-mente* selected from this corpus may reflect more the stylistic way of writing used in this newspaper rather than the



standard way that it was expected these adverbs to be used. In order to check this premise, it would be necessary to analyse more *Adv-mente* from different sources.

For this reason, it seems important to analyse a larger number of *Adv-mente* and also a varied and larger corpus may. By doing this, more instances of *Adv-mente* will be selected, sparse data will be avoided and a varied type of texts from different sources will be considered.

Regarding the analysis of *Adv-mente* and annotation of more instances, it is intended to adopt an inter-annotation agreement with the purpose of measuring how well native speakers can agree on the meaning of and adverb in context. However, the syntactic manipulation of real instances in view of its formal classification poses a real challenge and should be undertaken carefully (Gross, 1988), preferably by professional linguists.

More features need to be tested in order to improve the performance of the classifiers, especially when dealing with all adverbs and classes at the same time. In this study, for example, strictly semantic features were not considered. The semantic roles may be one feature to be considered, as they may help to identify some classes of adverbs. For instance, the semantic feature HUMAN attached to subject noun phrases of certain verbs may help to identify the MS class adverbs, since most of them modify verbs requiring human subjects. In this sense, attention should be given to the semantic role as well, for human arguments of a predicate can display the same role (AGENT) in different syntactic functions, for example, a subject of an active verb construction may become a prepositional complement of the corresponding passive sentence.

# References

- Afonso, S. (2002). Clara e sucintamente: Um estudo em corpus sobre a coordenação em -mente. In *Actas do XVIII Encontro da Associação Portuguesa de Linguística (APL2002)*, pages 27–36, Porto.
- Ait-Mokhtar, S., Chanod, J., and Roux, C. (2002). Robustness beyond shallowness: incremental deep parsing. *Natural Language Engineering*, 8(2-3):121–144.
- Almeida Costa, J. and Sampaio e Melo, A. (1999). *Dicionário da Língua Portuguesa*. Porto Editora, Porto, 8th edition.
- Aluísio, S. M. and Pinheiro, G. M. (2003). Corpus NILC : descrição e análise crítica com vistas ao projeto Lacio-Web. NILC-TR-03-03. Technical report, Universidade de São Paulo USP, São Paulo.
- Araujo, B. S., editor (2006). *Aprendizaje Automático: conceptos básicos y avanzados*. Pearson Prentice Hall, Madrid.
- Baptista, J. and Català, D. (2002). Compound Temporal Adverbs in Portuguese and in Spanish. In Ranchhod, E. M. and Mamede, N., editors, *Third International Conference, PorTAL 2002, LNAI 2389*, pages 133–136, Berlin, Heidelberg. Springer-Verlag.
- Baptista, J. and Català, D. (2009a). Adverbes focalisateurs et analyse syntaxique automatique de groupes nominaux. In *8es Journées scientifiques du Réseau LTT*, pages 97–110, Lisbonne.

- Baptista, J. and Català, D. (2009b). Disambiguation of Focus Adverbs in Portuguese and Spanish. In *International Symposium on Data and Sense Mining, Machine Translation and Controlled Languages, and their applications to emergencies and safety critical domains - ISMTCL*, pages 31–37, Besançon. Presses Universitaires de Franche-Comté.
- Baptista, J., Hagège, C., and Mamede, N. J. (2008). *Identificação, classificação e normalização de expressões temporais do português: A experiência do Segundo HAREM e o futuro*, chapter 2, pages 33–54. Linguateca.
- Bechara, E. (2001). *Moderna Gramática do Português Contemporâneo*. Lucerna, Rio de Janeiro, 37th edition.
- Bick, E. (2000). *The Parsing System "Palavras". Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*. University Press, Århus.
- Branco, A. and Silva, J. a. (2004). Solutions for the Rapid Development of State-of-the-Art POS Taggers for Portuguese. *Proceedings of the 4th International Conference on Language Resources and Evaluation (LREC2004)*, pages 507–510.
- Brito, A. M. (2003). *Categorias sintáticas (chapter 11)*. In *Gramática da Língua Portuguesa*. Caminho, Lisboa, 7th edition.
- Buarque de Holanda Ferreira, A. (2004). *Novo Dicionário Eletrônico Aurélio versão 5.0*. Positivo Informática Ltda.
- Català, D. and Baptista, J. (2007a). Limits to pattern-matching, finite-state techniques on lexical analysis for multiword expressions: the case for compound adverbs in Spanish. In *26th International Colloquium on Compared and Lexicon Grammar*, pages 17–24, Bonifacio(Corse du Sud).
- Català, D. and Baptista, J. (2007b). Spanish Adverbial Frozen Expressions. In *Proceedings of the Workshop on a Broader Perspective on Multiword Expressions*, pages 33–40, Prague, Czech Republic. Association for Computational Linguistics.

- Conlon, S. and Evens, M. (1992). Can computers handle adverbs? In *Proceedings of the 14th conference on Computational linguistics-Volume 4*, pages 1192–1196. Association for Computational Linguistics.
- Costa, J. (2008). *O advérbio em português europeu*. Colibri, Lisboa.
- Cunha, C. and Cintra, L. (1984). *Nova Gramática do Português Contemporâneo*. Edições João Sá da Costa, Lisboa, 16th edition.
- Defrise, C. (1989). Lexical description for NLP: The case of the French adverb *presque*. *Machine Translation*, 4(3):195–232.
- Gross, M. (1975). *Méthodes en syntaxe. Régime des constructions complétives*. Hermann, Paris.
- Gross, M. (1977). *Grammaire transformationnelle du français. Syntaxe du nom*. Larousse, Paris.
- Gross, M. (1986). *Grammaire transformationnelle du français. 3 - Syntaxe de l'adverbe*. ASSTRIL, Paris.
- Gross, M. (1988). Methods and Tactics in the Construction of a Lexicon-Grammar. In *Linguistics in the Morning Calm 2. Select papers from SICOL (1986)*, pages 177–197, Seoul. Hanshin Publishing Company.
- Guimier, C. (1996). *Les adverbes du français: le cas des adverbes en -ment*. Ophrys, Paris.
- Hagège, C., Baptista, J., and Mamede, N. J. (2009). Portuguese Temporal Expressions Recognition: from TE characterization to an effective TER module implementation. *The 7th Brazilian Symposium in Information and Human Language Technology*.
- Hagège, C., Baptista, J., and Mamede, N. J. (2010). Caracterização e Processamento de Expressões Temporais em Português. *Linguamática*, 2(1):63–76.
- Hall, M., National, H., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., and Witten, I. H. (2009). The WEKA Data Mining Software : An Update. *SIGKDD Explorations*, 11(1):10–18.

- Harris, Z. S. (1976). *Notes du cours de syntaxe*. Le Seuil, Paris.
- Harris, Z. S. (1991). *A Theory of Language and Information. A Mathematical Approach*. Clarendon Press, Oxford.
- Houaiss, A. and Mauro, V. (2009). *Dicionário eletrônico Houaiss da língua portuguesa 3.0 [CD-ROM]*. Instituto Antônio Houaiss. Objetiva, Rio de Janeiro.
- Houaiss, A. and Villar, M. (2001). *Dicionário Eletrônico Houaiss da Língua Portuguesa 1.0 [CD-ROM]*. Instituto Antônio Houaiss. Objetiva, Rio de Janeiro.
- Ide, N. and Véronis, J. (1998). Word Sense Disambiguation: The State of the Art. *Computational Linguistics*, 24(1).
- Iraolak, D. M. (2004). *Supervised Word Sense Disambiguation : Facing Current Challenges*. PhD thesis, Universidad del País Vasco, Donostia.
- Jurafsky, D. and Martin, J. H. (2009). *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics, and Speech Recognition*. Pearson/Prentice Hall, New Jersey, 2nd edition.
- Lenci, A., Bel, N., Busa, F., Calzolari, N., Gola, E., Monachini, M., Ogonowski, A., Peters, I., Peters, W., Ruimy, N., Villegas, M., and Zampolli, A. (2000). SIMPLE: A General Framework for Multilingual Lexicons. *International Journal of Lexicography*, 13(4):249–263.
- Mamede, N., Baptista, J., and Hagège, C. (2007). Nomenclature of chunks and dependencies in Portuguese XIP grammar 2.1. (Internal Report). Technical report, L2F/INESC ID Lisboa, Lisboa.
- Manning, C. and Schütze, H. (2001). *Statistical foundations of natural language processing*. The MIT Press, London.
- Molinier, C. and Levrier, F. (2000). *Grammaire des Adverbes. Description des formes en -ment*. Librairie Droz, Genève-Paris.
- Nimb, S. (2002). Adverbs in Semantic Lexica for NLP - The Extension of the Danish SIMPLE Lexicon with Time Adverbs. In *Proceedings of the Third*

- International Conference on Language Resources and Evaluation*, pages 800–806, Gran Canaria.
- Palma, C. (2009). Expressões fixas adverbiais: descrição léxico-sintáctica e subsídios para um estudo contrastivo Português-Espanhol (MA Thesis). Faro, Univ. Algarve - FCHS, Faro, Universidade do Algarve- FCHS.
- Santos, D. (2009). Caminhos percorridos no mapa da portuguesificação: A Linguateca em perspectiva. *Linguamática*, 1:25–59.
- Santos, D. and Bick, E. (2000). Providing Internet access to Portuguese corpora: the AC/DC project. In Gavrilidou, M., Carayannis, G., Markantonatou, S., Piperidis, S., and Stainhauer, G., editors, *Proceedings of the Second International Conference on Language Resources and Evaluation (LREC 2000)*, number September, pages 205–210, Atenas, Greece.
- Santos, D. and Sarmiento, L. (2002). O projecto AC/DC: acesso a corpora/disponibilização de corpora. In Mendes, A. and Freitas, T., editors, *Actas do XVIII Encontro Nacional da Associação Portuguesa de Linguística (APL 2002)*, pages 705–717, Lisboa. APL.
- Sparck-Jones, K. and Galliers, J. R. (1996). *Evaluating Natural Language Processing Systems: an Analysis and Review*. Springer-Verlag.
- Specia, L. (2007). *Uma abordagem híbrida relacional para a desambiguação lexical de sentido na tradução automática (A hybrid relational approach for Word Sense Disambiguation)*. PhD Thesis. PhD thesis, ICMC-USP, São Carlos, Brazil.
- Stevenson, M. (2003). *Word sense disambiguation: The Case for Combinations of Knowledge Sources*, volume 30. CSLI Publications, Standford.
- Stevenson, M. and Wilks, Y. (2003). *Word-Sense Disambiguation*, chapter 13, pages 249–265. Oxford University Press, Oxford, UK.
- Weaver, W. (1955). *Translation. In Machine Translation of Languages: Fourteen Essays*. The MIT Press, Cambridge, reprinted edition.

Witten, I. and Frank, E. (2005). *Data mining: practical machine learning tools and techniques with Java implementations*, volume 31. Morgan Kaufmann Publishers, San Francisco, 2nd edition.

Zan, H., Zhang, J., Zhu, X., and Yu, S. (2010). Studies on Automatic Recognition of Common Chinese Adverb's Usages Based on Statistical Methods. In *CIPS-SIGHAN Joint Conference on Chinese Language Processing*.





# Appendices



# Appendix A

## Classification of Brazilian Portuguese *Adv-mente*

List of the 532 most frequent *Adv-mente* in the NILC/São Carlos corpus, classified according to Molinier and Levrier (2000) syntactic and semantic criteria.

<i>Number</i>	<i>Adv-mente</i>	<i>Frequency</i>	<i>Classification</i>
1	principalmente	4478	MF
2	somente	4111	MF
3	atualmente	2639	MT
4	especialmente	2490	MQ, MF, MV
5	realmente	2255	PS, MV, MQ
6	exatamente	2200	MF, MQ, MS
7	totalmente	2028	MQ
8	diretamente	2009	MV
9	novamente	1899	MT
10	simplesmente	1802	PC, PS, MF, MS
11	finalmente	1676	PC
12	certamente	1619	PAm
13	normalmente	1546	PAh, MV
14	recentemente	1478	MT
15	justamente	1453	PC, MF, MS
16	provavelmente	1418	PAm
17	imediatamente	1332	MV
18	aproximadamente	1289	MF
19	respectivamente	1223	MF
20	geralmente	1216	PAh
21	diariamente	1175	MT
22	completamente	1154	MQ
23	exclusivamente	1095	MF, MS
24	igualmente	1092	PC, MV
25	inicialmente	1044	MT

Appendix A. Classification of Brazilian Portuguese *Adv-mente*

---

<i>Number</i>	<i>Adv-mente</i>	<i>Frequency</i>	<i>Classification</i>
26	oficialmente	1009	MP, MV
27	aparentemente	934	PAm
28	extremamente	924	MQ
29	absolutamente	853	MQ
30	naturalmente	812	PAm, MS, MV
31	necessariamente	796	PAm
32	difícilmente	750	MV
33	relativamente	747	MQ
34	pessoalmente	738	PS, MF
35	claramente	717	MS, MV
36	infelizmente	715	PAa, MS
37	facilmente	707	MV
38	inteiramente	688	MQ
39	particularmente	677	MQ, MF, MV
40	independentemente	675	PC, MS
41	efetivamente	670	PAm
42	definitivamente	633	MT
43	perfeitamente	600	MQ, MV
44	evidentemente	600	PAm
45	supostamente	594	PAm
46	anteriormente	583	MT
47	propriamente	578	MQ, MV
48	precisamente	540	PC, MF, MS, MV
49	posteriormente	524	MT
50	obviamente	516	PAm
51	profundamente	492	MQ, MV
52	eventualmente	475	PAm
53	parcialmente	471	PAm
54	fortemente	463	MQ, MV
55	automaticamente	458	MV
56	altamente	436	MQ, MV
57	frequentemente	416	MT
58	juntamente	406	PC, MV
59	essencialmente	418	MQ, MF
60	anualmente	394	MT
61	devidamente	391	MV
62	corretamente	389	MV
63	previamente	388	MT
64	possivelmente	386	PAm
65	mensalmente	383	MT
66	constantemente	381	MT
67	literalmente	353	PS
68	originalmente	352	MT, MV
69	raramente	346	MT
70	tradicionalmente	337	PAh
71	plenamente	331	MQ
72	amplamente	313	MQ
73	suficientemente	311	MQ
74	puramente	308	MQ

Appendix A. Classification of Brazilian Portuguese *Adv-mente*

---

<i>Number</i>	<i>Adv-mente</i>	<i>Frequency</i>	<i>Classification</i>
75	formalmente	307	MP, MV
76	gratuitamente	302	MV
77	meramente	291	MF
78	verdadeiramente	290	PS, MQ, MV
79	felizmente	281	PAa, MS
80	antigamente	279	MT
81	livremente	276	MV
82	lentamente	274	MV
83	integralmente	271	MQ
84	paralelamente	270	PC, MV
85	indiretamente	270	MV
86	excessivamente	269	MQ
87	legalmente	266	MP
88	radicalmente	263	MQ
89	abertamente	257	MV
90	economicamente	251	MP
91	sexualmente	246	MP, MV
92	regularmente	246	MT
93	seriamente	230	PS, MQ, MS
94	diferentemente	227	MV
95	frequentemente	224	MT
96	curiosamente	221	PAa, MV
97	individualmente	220	MP
98	antecipadamente	218	MT
99	comumente	214	MV
100	internamente	211	MV
101	fundamentalmente	203	MF
102	ligeiramente	202	MQ, MV
103	permanentemente	199	MT
104	rigorosamente	198	MQ, MV
105	estritamente	223	MQ
106	internacionalmente	197	MP
107	gravemente	192	MQ
108	virtualmente	189	PAm, MQ, MV
109	ultimamente	183	MT
110	cuidadosamente	183	MS
111	ilegalmente	182	MV
112	separadamente	179	MV
113	paradoxalmente	179	PC
114	expressamente	179	MV
115	excepcionalmente	179	PAh
116	consequentemente	179	PC
117	significativamente	178	MQ
118	seguramente	175	PAm, MS, MV
119	adequadamente	175	MV
120	levemente	173	MQ, MV
121	judicialmente	172	MP
122	semanalmente	161	MT
123	irregularmente	158	MV

Appendix A. Classification of Brazilian Portuguese *Adv-mente*

---

<i>Number</i>	<i>Adv-mente</i>	<i>Frequency</i>	<i>Classification</i>
124	predominantemente	155	MQ
125	sucessivamente	154	MV
126	socialmente	154	MP
127	duramente	154	MQ, MV
128	consequentemente	153	PC, MS(raro)
129	francamente	151	PS, MQ, MS
130	progressivamente	149	MV
131	isoladamente	149	MV
132	infinitamente	147	MQ
133	gradualmente	146	MV
134	proporcionalmente	145	MV
135	substancialmente	143	MQ
136	intensamente	143	MQ
137	mundialmente	141	MV
138	indevidamente	138	MV
139	ativamente	138	MS
140	habitualmente	133	PAh
141	espontaneamente	147	MS
142	notadamente	130	MF
143	gradativamente	128	MV
144	discretamente	127	MS
145	sinceramente	125	PS, MS
146	financeiramente	125	MP
147	primeiramente	120	PC
148	largamente	120	MQ
149	usualmente	119	PAh
150	potencialmente	110	PAm
151	positivamente	110	MV
152	artificialmente	109	MV
153	surpreendentemente	108	PAa
154	informalmente	108	MS
155	deliberadamente	108	MV
156	reservadamente	105	MV
157	decididamente	103	PS, MQ, MS
158	negativamente	102	MV
159	injustamente	98	MV
160	eleitoralmente	98	MP
161	moralmente	93	MP
162	demasiadamente	93	MQ
163	indefinidamente	92	MV
164	violentamente	91	MS
165	eminentemente	89	MQ
166	desesperadamente	89	MS
167	longamente	88	MT
168	preferencialmente	87	MF
169	tranquilamente	85	MS
170	profissionalmente	85	MP, MV
171	eternamente	85	MT
172	estranhamente	85	PAa

Appendix A. Classification of Brazilian Portuguese *Adv-mente*

---

<i>Number</i>	<i>Adv-mente</i>	<i>Frequency</i>	<i>Classification</i>
173	prontamente	83	MV
174	nacionalmente	82	MP
175	mentalmente	82	MP, MV
176	firmemente	81	MS, MV
177	comercialmente	80	MP, MV
178	fatalmente	79	MQ
179	duplamente	78	MQ
180	genericamente	76	MP, MV
181	atentamente	76	MS
182	repentinamente	75	MT
183	liminarmente	75	PC
184	detalhadamente	74	MV
185	brevemente	74	MS, MT
186	inversamente	73	PC, MV
187	conjuntamente	72	PC, MV
188	vagamente	71	MV
189	ostensivamente	71	MS
190	extraordinariamente	70	MQ
191	emocionalmente	70	MP, MV
192	secretamente	69	MV
193	clandestinamente	69	MV
194	visualmente	67	MP, MV
195	decisivamente	67	MV
196	perigosamente	66	MV
197	inconscientemente	65	MS
198	culturalmente	65	MP
199	tão-somente	64	MF
200	objetivamente	64	PS
201	superficialmente	63	PAa, MQ, MV
202	oportunamente	62	MT
203	interinamente	62	MV
204	concretamente	61	PS
205	insistentemente	60	MS
206	sumariamente	59	PC
207	criminalmente	59	MP, MV
208	unilateralmente	58	MV
209	passivamente	58	MS
210	exaustivamente	58	MV
211	bruscamente	58	MV
212	urgentemente	57	MV
213	nominalmente	57	MV
214	minimamente	57	MQ
215	ideologicamente	57	MP
216	coincidentemente	57	PAm
217	severamente	56	MS
218	momentaneamente	56	MT
219	estritamente	56	MV
220	ininterruptamente	55	MT
221	fielmente	55	MS

Appendix A. Classification of Brazilian Portuguese *Adv-mente*

---

<i>Number</i>	<i>Adv-mente</i>	<i>Frequency</i>	<i>Classification</i>
222	precocemente	54	MV
223	minuciosamente	54	MS
224	manualmente	54	MV
225	inesperadamente	54	MV
226	conscientemente	54	MS
227	enormemente	52	MQ
228	abusivamente	52	MV
229	verbalmente	51	MV
230	vagarosamente	51	MS
231	tranquilamente	51	MS
232	inutilmente	51	MV
233	friamente	51	MS
234	docemente	51	MS
235	sabidamente	50	MS
236	comprovadamente	50	PAm
237	brutalmente	50	MS
238	modernamente	49	MT, MV
239	exageradamente	49	MQ
240	diversamente	49	MV
241	repetidamente	48	MF
242	ocasionalmente	48	MT
243	acidentalmente	48	MV
244	tristemente	47	PAa, MS
245	manifestamente	47	PAm
246	incessantemente	47	MV
247	constitucionalmente	47	MP
248	reconhecidamente	46	PAm, MS
249	militarmente	46	MP
250	incrivelmente	46	PAa, MQ
251	tardiamente	45	MT
252	propositalmente	45	MV
253	pacientemente	45	MS
254	impunemente	45	MV
255	honestamente	45	PS, MS
256	futuramente	45	MT
257	intencionalmente	44	MV
258	forçosamente	44	PAm
259	convenientemente	44	PAa
260	aproxidamente	44	MQ
261	imensamente	43	MQ
262	externamente	43	MV
263	indiscriminadamente	42	MV
264	frontalmente	42	MS
265	alternadamente	42	MV
266	abruptamente	42	MV
267	universalmente	41	MP
268	modestamente	41	PS, MS
269	misteriosamente	41	MS
270	marcadamente	41	MQ



Appendix A. Classification of Brazilian Portuguese *Adv-mente*

---

<i>Number</i>	<i>Adv-mente</i>	<i>Frequency</i>	<i>Classification</i>
271	carinhosamente	41	MS
272	pretensamente	40	MS
273	localmente	40	MP
274	casualmente	40	MT
275	vivamente	39	MV
276	musicalmente	39	MP
277	intelectualmente	39	MP, MV
278	instintivamente	39	MV
279	empiricamente	39	MP
280	assumidamente	39	PA <sub>s</sub> , MS
281	seguidamente	38	PC
282	inadvertidamente	38	PA <sub>m</sub>
283	grandemente	38	MQ
284	fartamente	38	MQ
285	crescentemente	38	MQ
286	concomitantemente	38	PC
287	coletivamente	38	MV
288	alternativamente	38	MV
289	administrativamente	38	MP, MV
290	terrivelmente	37	MQ
291	solenemente	37	MS
292	prematuramente	37	MT
293	paulatinamente	37	MQ
294	calmamente	37	MS
295	tremendamente	36	MQ
296	suavemente	36	MV
297	oralmente	36	MV
298	desnecessariamente	36	PA <sub>a</sub>
299	adicionalmente	36	PC
300	silenciosamente	35	MS
301	pontualmente	35	MT, MS
302	ofensivamente	35	MS
303	exemplarmente	35	MV
304	comparativamente	35	PC
305	brilhantemente	35	MS
306	religiosamente	34	MP, MV
307	propositadamente	34	MV
308	pesadamente	34	MV
309	absurdamente	33	PA <sub>a</sub> , MQ
310	verticalmente	32	MV
311	tragicamente	32	PA <sub>a</sub> , MV
312	preventivamente	32	MV
313	mormente	32	MF
314	horizontalmente	32	MV
315	delicadamente	32	MS
316	alegremente	32	MS
317	vigorosamente	31	MS
318	inerentemente	31	MV
319	ansiosamente	31	MS

Appendix A. Classification of Brazilian Portuguese *Adv-mente*

---

<i>Number</i>	<i>Adv-mente</i>	<i>Frequency</i>	<i>Classification</i>
320	acertadamente	31	MV
321	uniformemente	30	MV
322	racionalmente	30	MS
323	maravilhosamente	30	MS
324	interiormente	30	MV
325	erradamente	30	MV
326	primordialmente	29	MF
327	ordinariamente	29	MV
328	maciçamente	29	MV
329	indistintamente	29	MV
330	grosseiramente	29	MS
331	espiritualmente	29	MV
332	textualmente	28	MP
333	popularmente	28	MP
334	obsessivamente	28	MV
335	gentilmente	28	MS
336	generosamente	28	MS
337	defensivamente	28	MS
338	atenciosamente	28	MS
339	moderadamente	27	MQ
340	conceitualmente	27	MP
341	cegamente	27	MV
342	assustadoramente	27	MS
343	apropriadamente	27	PAa, MV
344	terminantemente	26	MV
345	sutilmente	26	MS
346	reiteradamente	26	MT
347	cumulativamente	26	MV
348	cotidianamente	26	PAh
349	acentuadamente	26	MQ
350	remotamente	25	MQ
351	incorretamente	25	MV
352	eficientemente	25	MS
353	cautelosamente	25	MS
354	sabiamente	24	PAa, MS
355	qualitativamente	24	MP
356	idealmente	24	PAm, MV
357	falsamente	24	PS, MQ, MV
358	extensivamente	24	MV
359	experimentalmente	24	MP, MV
360	distintamente	24	MV
361	artesanamente	24	MV
362	abundantemente	24	MQ
363	resumidamente	23	PC, MV
364	intuitivamente	23	MV
365	humildemente	23	MS
366	equivocadamente	23	MV
367	eficazmente	23	MV
368	demoradamente	23	MV

Appendix A. Classification of Brazilian Portuguese *Adv-mente*

---

<i>Number</i>	<i>Adv-mente</i>	<i>Frequency</i>	<i>Classification</i>
369	apressadamente	23	MS
370	apaixonadamente	23	MV
371	velozmente	22	MV
372	rotineiramente	22	MT
373	quinzenalmente	22	MT
374	merecidamente	22	PAa
375	declaradamente	22	PAm, MV
376	retrospectivamente	21	MV
377	retroativamente	21	MV
378	preliminarmente	21	MV
379	milagrosamente	21	MV
380	meticulosamente	21	MS
381	materialmente	21	MP, MV
382	magistralmente	21	MV
383	fluentemente	21	MV
384	fixamente	21	MV
385	detidamente	21	MV
386	veementemente	20	PS, MS
387	secamente	20	MS
388	respeitosamente	20	MS
389	regionalmente	20	MV
390	quantitativamente	20	MP
391	precipitadamente	20	MS
392	institucionalmente	20	MP
393	extra-oficialmente	20	MP, MV
394	disfarçadamente	20	MV
395	diametralmente	20	MV
396	civilmente	20	MP, MV
397	impiedosamente	19	MS
398	globalmente	19	MP
399	ferozmente	19	MQ, MS
400	extensamente	19	MV
401	elegantemente	19	MS
402	descaradamente	19	MS
403	cruelmente	19	PAs, MQ, MS, MV
404	convincentemente	19	PS, MS
405	animadamente	19	MV
406	vulgarmente	18	MS
407	vertiginosamente	18	MV
408	semestralmente	18	MT
409	orgulhosamente	18	MS
410	operacionalmente	18	MV
411	maliciosamente	18	MS
412	incidentalmente	18	MV
413	harmoniosamente	18	MV
414	digitalmente	18	MP, MV
415	convencionalmente	18	MV
416	ardentemente	18	MV
417	singularmente	17	PAa, MQ, MV

Appendix A. Classification of Brazilian Portuguese *Adv-mente*

---

<i>Number</i>	<i>Adv-mente</i>	<i>Frequency</i>	<i>Classification</i>
418	serenamente	17	MS
419	ricamente	17	MV
420	inteligentemente	17	PAs, MS
421	inocentemente	17	MS
422	estruturalmente	17	MP, MV
423	dolorosamente	17	MV
424	dignamente	17	MS
425	desgraçadamente	17	PAa
426	densamente	17	MV
427	corajosamente	17	MS
428	consistentemente	17	MV
429	confusamente	17	MS
430	cognitivamente	17	MP
431	teimosamente	16	MS
432	sumamente	16	MQ
433	sorrateiramente	16	MS
434	similarmente	16	MV
435	seletivamente	16	MV
436	ruidosamente	16	MS
437	furiosamente	16	MS
438	familiarmente	16	MP, MV
439	escandalosamente	16	PAa
440	efusivamente	16	MS
441	dominantemente	16	MQ
442	deliciosamente	16	MV
443	cordialmente	16	MS
444	ulteriormente	15	MT
445	surdamente	15	MV
446	primitivamente	15	MT
447	medianamente	15	MQ
448	loucamente	15	MS
449	incondicionalmente	15	MV
450	especificamente	15	MQ, MV
451	epistemologicamente	15	MP, MV
452	desordenadamente	15	MS
453	criminosamente	15	MV
454	produtivamente	14	MV
455	presentemente	14	MT
456	inadequadamente	14	MV
457	furtivamente	14	MV
458	decentemente	14	MS
459	convulsivamente	14	MV
460	compulsivamente	14	MV
461	coerentemente	14	PC, MV
462	cabalmente	14	MQ
463	bravamente	14	MS
464	abstratamente	14	MS
465	tempestivamente	13	MV
466	sigilosamente	13	MS

Appendix A. Classification of Brazilian Portuguese *Adv-mente*

---

<i>Number</i>	<i>Adv-mente</i>	<i>Frequency</i>	<i>Classification</i>
467	redondamente	13	MQ, MV
468	perversamente	13	PAAs, MS
469	persistentemente	13	MS
470	milimetricamente	13	MV
471	indiferentemente	13	MV
472	graciosamente	13	MS
473	flagrantemente	13	MV
474	espantosamente	13	PAa, MQ
475	criativamente	13	MS
476	correntemente	13	PAh, MT
477	amargamente	13	MS
478	afirmativamente	13	MV
479	visceralmente	12	MQ
480	trimestralmente	12	MT
481	sucintamente	12	MS
482	resolutamente	12	MQ, MS
483	pejorativamente	12	MV
484	nervosamente	12	MS
485	mortalmente	12	MQ, MV
486	maquinalmente	12	MV
487	mansamente	12	MS
488	humanamente	12	MS
489	fervorosamente	12	MS
490	extrajudicialmente	12	MP, MV
491	distraidamente	12	MS
492	desigualmente	12	MV
493	covardemente	12	MS
494	atabalhoadamente	12	MS
495	ardorosamente	12	MS
496	aceleradamente	12	MV
497	veladamente	11	MV
498	toscamente	11	MS
499	tipicamente	11	PAh, MV
500	piamente	11	MS
501	percentualmente	11	MQ
502	penalmente	11	MP
503	opcionalmente	11	MV
504	obstinadamente	11	MS
505	marginalmente	11	MV
506	levianamente	11	MS
507	lateralmente	11	MV
508	insuficientemente	11	MQ
509	fracamente	11	MQ
510	finamente	11	MV
511	expressivamente	11	MV
512	criteriosamente	11	MS
513	ambientalmente	11	MP
514	subsequentemente	10	PC
515	regressivamente	10	MV

**Appendix A. Classification of Brazilian Portuguese *Adv-mente***

---

<i>Number</i>	<i>Adv-mente</i>	<i>Frequency</i>	<i>Classification</i>
516	preguiçosamente	10	MS
517	perdidamente	10	MQ
518	penosamente	10	MV
519	parceladamente	10	MV
520	irracionalmente	10	MS
521	injustificadamente	10	MV
522	industrialmente	10	MP, MV
523	incialmente	10	MT
524	frouxamente	10	MS
525	desbragadamente	10	MS
526	cruamente	10	MS
527	copiosamente	10	MQ
528	contratualmente	10	MP, MV
529	caprichosamente	10	MS
530	calorosamente	10	MS
531	alegadamente	10	PAm
532	afetivamente	10	MS

# Appendix B

## Comparison with AEDic and HEDic

List of the 532 most frequent *Adv-mente* in the NILC/São Carlos corpus and their presence or absence in Aurélio Electronic Dictionary of Brazilian Portuguese (Buarque de Holanda Ferreira, 2004) - AEDic - and Houaiss Electronic Dictionary (Houaiss and Mauro, 2009) - HEDic.

<i>Number</i>	<i>Adv-mente</i>	<i>Frequency</i>	<i>Dictionary</i>
1	principalmente	4478	HEDic
2	somente	4111	AEDic and HEDic
3	atualmente	2639	AEDic and HEDic
4	especialmente	2490	HEDic
5	realmente	2255	HEDic
6	exatamente	2200	HEDic
7	totalmente	2028	AEDic
8	diretamente	2009	HEDic
9	novamente	1899	AEDic and HEDic
10	simplesmente	1802	AEDic and HEDic
11	finalmente	1676	HEDic
12	certamente	1619	AEDic and HEDic
13	normalmente	1546	HEDic
14	recentemente	1478	AEDic and HEDic
15	justamente	1453	HEDic
16	provavelmente	1418	HEDic
17	imediatamente	1332	HEDic
18	aproximadamente	1289	HEDic
19	respectivamente	1223	HEDic
20	geralmente	1216	AEDic and HEDic
21	diariamente	1175	HEDic
22	completamente	1154	HEDic
23	exclusivamente	1095	HEDic
24	igualmente	1092	Not appearing
25	inicialmente	1044	HEDic

Appendix B. Comparison with AEDic and HEDic

---

<i>Number</i>	<i>Adv-mente</i>	<i>Frequency</i>	<i>Dictionary</i>
26	oficialmente	1009	Not appearing
27	aparentemente	934	HEDic
28	extremamente	924	AEDic and HEDic
29	absolutamente	853	AEDic and HEDic
30	naturalmente	812	AEDic and HEDic
31	necessariamente	796	Not appearing
32	difícilmente	750	HEDic
33	relativamente	747	HEDic
34	pessoalmente	738	HEDic
35	claramente	717	Not appearing
36	infelizmente	715	AEDic and HEDic
37	facilmente	707	AEDic and HEDic
38	inteiramente	688	HEDic
39	particularmente	677	HEDic
40	independentemente	675	Not appearing
41	efetivamente	670	Not appearing
42	definitivamente	633	AEDic
43	perfeitamente	600	HEDic
44	evidentemente	600	AEDic and HEDic
45	supostamente	594	AEDic
46	anteriormente	583	HEDic
47	propriamente	578	HEDic
48	precisamente	540	Not appearing
49	posteriormente	524	HEDic
50	obviamente	516	AEDic
51	profundamente	492	Not appearing
52	eventualmente	475	HEDic
53	parcialmente	471	Not appearing
54	fortemente	463	Not appearing
55	automaticamente	458	HEDic
56	altamente	436	HEDic
57	frequentemente	416	AEDic
58	juntamente	406	Not appearing
59	essencialmente	418	HEDic
60	anualmente	394	Not appearing
61	devidamente	391	HEDic
62	corretamente	389	Not appearing
63	previamente	388	Not appearing
64	possivelmente	386	HEDic
65	mensalmente	383	Not appearing
66	constantemente	381	HEDic
67	literalmente	353	HEDic
68	originalmente	352	Not appearing
69	raramente	346	HEDic
70	tradicionalmente	337	Not appearing
71	plenamente	331	AEDic and HEDic
72	amplamente	313	Not appearing
73	suficientemente	311	Not appearing
74	puramente	308	HEDic



Appendix B. Comparison with AEDic and HEDic

---

<i>Number</i>	<i>Adv-mente</i>	<i>Frequency</i>	<i>Dictionary</i>
75	formalmente	307	Not appearing
76	gratuitamente	302	Not appearing
77	meramente	291	Not appearing
78	verdadeiramente	290	Not appearing
79	felizmente	281	AEDic and HEDic
80	antigamente	279	HEDic
81	livremente	276	Not appearing
82	lentamente	274	Not appearing
83	integralmente	271	Not appearing
84	paralelamente	270	Not appearing
85	indiretamente	270	Not appearing
86	excessivamente	269	Not appearing
87	legalmente	266	Not appearing
88	radicalmente	263	HEDic
89	abertamente	257	HEDic
90	economicamente	251	HEDic
91	sexualmente	246	Not appearing
92	regularmente	246	Not appearing
93	seriamente	230	Not appearing
94	diferentemente	227	Not appearing
95	frequentemente	224	AEDic
96	curiosamente	221	Not appearing
97	individualmente	220	Not appearing
98	antecipadamente	218	Not appearing
99	comumente	214	AEDic and HEDic
100	internamente	211	HEDic
101	fundamentalmente	203	AEDic
102	ligeiramente	202	Not appearing
103	permanentemente	199	Not appearing
104	rigorosamente	198	Not appearing
105	estritamente	223	AEDic and HEDic
106	internacionalmente	197	Not appearing
107	gravemente	192	Not appearing
108	virtualmente	189	Not appearing
109	ultimamente	183	AEDic and HEDic
110	cuidadosamente	183	Not appearing
111	ilegalmente	182	Not appearing
112	separadamente	179	Not appearing
113	paradoxalmente	179	Not appearing
114	expressamente	179	HEDic
115	excepcionalmente	179	Not appearing
116	consequentemente	179	HEDic
117	significativamente	178	Not appearing
118	seguramente	175	HEDic
119	adequadamente	175	AEDic
120	levemente	173	AEDic
121	judicialmente	172	Not appearing
122	semanalmente	161	Not appearing
123	irregularmente	158	Not appearing

Appendix B. Comparison with AEDic and HEDic

---

<i>Number</i>	<i>Adv-mente</i>	<i>Frequency</i>	<i>Dictionary</i>
124	predominantemente	155	Not appearing
125	sucessivamente	154	Not appearing
126	socialmente	154	HEDic
127	duramente	154	Not appearing
128	consequentemente	153	HEDic
129	francamente	151	HEDic
130	progressivamente	149	Not appearing
131	isoladamente	149	Not appearing
132	infinitamente	147	Not appearing
133	gradualmente	146	Not appearing
134	proporcionalmente	145	Not appearing
135	substancialmente	143	Not appearing
136	intensamente	143	Not appearing
137	mundialmente	141	Not appearing
138	indevidamente	138	Not appearing
139	ativamente	138	Not appearing
140	habitualmente	133	Not appearing
141	espontaneamente	147	AEDic
142	notadamente	130	AEDic and HEDic
143	gradativamente	128	Not appearing
144	discretamente	127	AEDic
145	sinceramente	125	HEDic
146	financeiramente	125	Not appearing
147	primeiramente	120	Not appearing
148	largamente	120	Not appearing
149	usualmente	119	Not appearing
150	potencialmente	110	Not appearing
151	positivamente	110	Not appearing
152	artificialmente	109	Not appearing
153	surpreendentemente	108	Not appearing
154	informalmente	108	Not appearing
155	deliberadamente	108	Not appearing
156	reservadamente	105	Not appearing
157	decididamente	103	Not appearing
158	negativamente	102	Not appearing
159	injustamente	98	Not appearing
160	eleitoralmente	98	Not appearing
161	moralmente	93	Not appearing
162	demasiadamente	93	Not appearing
163	indefinidamente	92	Not appearing
164	violentamente	91	AEDic
165	eminentemente	89	AEDic and HEDic
166	desesperadamente	89	Not appearing
167	longamente	88	Not appearing
168	preferencialmente	87	Not appearing
169	tranquilamente	85	AEDic and HEDic
170	profissionalmente	85	Not appearing
171	eternamente	85	AEDic
172	estranhamente	85	Not appearing

Appendix B. Comparison with AEDic and HEDic

---

<i>Number</i>	<i>Adv-mente</i>	<i>Frequency</i>	<i>Dictionary</i>
173	prontamente	83	Not appearing
174	nacionalmente	82	Not appearing
175	mentalmente	82	Not appearing
176	firmemente	81	Not appearing
177	comercialmente	80	Not appearing
178	fatalmente	79	AEDic
179	duplamente	78	Not appearing
180	genericamente	76	HEDic
181	atentamente	76	HEDic
182	repentinamente	75	AEDic
183	liminarmente	75	Not appearing
184	detalhadamente	74	Not appearing
185	brevemente	74	Not appearing
186	inversamente	73	Not appearing
187	conjuntamente	72	Not appearing
188	vagamente	71	Not appearing
189	ostensivamente	71	Not appearing
190	extraordinariamente	70	AEDic
191	emocionalmente	70	Not appearing
192	secretamente	69	Not appearing
193	clandestinamente	69	Not appearing
194	visualmente	67	Not appearing
195	decisivamente	67	Not appearing
196	perigosamente	66	Not appearing
197	inconscientemente	65	Not appearing
198	culturalmente	65	Not appearing
199	tão-somente	64	AEDic
200	objetivamente	64	Not appearing
201	superficialmente	63	AEDic
202	oportunamente	62	Not appearing
203	interinamente	62	Not appearing
204	concretamente	61	Not appearing
205	insistentemente	60	Not appearing
206	sumariamente	59	Not appearing
207	criminalmente	59	Not appearing
208	unilateralmente	58	Not appearing
209	passivamente	58	Not appearing
210	exaustivamente	58	Not appearing
211	bruscamente	58	Not appearing
212	urgentemente	57	Not appearing
213	nominalmente	57	HEDic
214	minimamente	57	Not appearing
215	ideologicamente	57	AEDic
216	coincidentemente	57	Not appearing
217	severamente	56	Not appearing
218	momentaneamente	56	Not appearing
219	estritamente	56	Not appearing
220	ininterruptamente	55	Not appearing
221	fielmente	55	Not appearing

Appendix B. Comparison with AEDic and HEDic

---

<i>Number</i>	<i>Adv-mente</i>	<i>Frequency</i>	<i>Dictionary</i>
222	precocemente	54	Not appearing
223	minuciosamente	54	Not appearing
224	manualmente	54	Not appearing
225	inesperadamente	54	Not appearing
226	conscientemente	54	Not appearing
227	enormemente	52	Not appearing
228	abusivamente	52	HEDic
229	verbalmente	51	Not appearing
230	vagarosamente	51	AEDic
231	tranquilamente	51	AEDic and HEDic
232	inutilmente	51	Not appearing
233	friamente	51	AEDic and HEDic
234	docemente	51	AEDic
235	sabidamente	50	Not appearing
236	comprovadamente	50	Not appearing
237	brutalmente	50	Not appearing
238	modernamente	49	Not appearing
239	exageradamente	49	Not appearing
240	diversamente	49	Not appearing
241	repetidamente	48	AEDic
242	ocasionalmente	48	Not appearing
243	accidentalmente	48	Not appearing
244	tristemente	47	Not appearing
245	manifestamente	47	Not appearing
246	incessantemente	47	Not appearing
247	constitucionalmente	47	Not appearing
248	reconhecidamente	46	Not appearing
249	militarmente	46	Not appearing
250	incrivelmente	46	HEDic
251	tardiamente	45	AEDic
252	propositalmente	45	Not appearing
253	pacientemente	45	Not appearing
254	impunemente	45	Not appearing
255	honestamente	45	HEDic
256	futuramente	45	AEDic and HEDic
257	intencionalmente	44	Not appearing
258	forçosamente	44	Not appearing
259	convenientemente	44	Not appearing
260	aproxidamente	44	Not appearing
261	imensamente	43	HEDic
262	externamente	43	HEDic
263	indiscriminadamente	42	Not appearing
264	frontalmente	42	Not appearing
265	alternadamente	42	Not appearing
266	abruptamente	42	Not appearing
267	universalmente	41	Not appearing
268	modestamente	41	Not appearing
269	misteriosamente	41	Not appearing
270	marcadamente	41	Not appearing

Appendix B. Comparison with AEDic and HEDic

---

<i>Number</i>	<i>Adv-mente</i>	<i>Frequency</i>	<i>Dictionary</i>
271	carinhosamente	41	Not appearing
272	pretensamente	40	Not appearing
273	localmente	40	Not appearing
274	casualmente	40	Not appearing
275	vivamente	39	Not appearing
276	musicalmente	39	Not appearing
277	intelectualmente	39	Not appearing
278	instintivamente	39	Not appearing
279	empiricamente	39	Not appearing
280	assumidamente	39	Not appearing
281	seguidamente	38	Not appearing
282	inadvertidamente	38	AEDic
283	grandemente	38	Not appearing
284	fartamente	38	Not appearing
285	crecentemente	38	Not appearing
286	concomitantemente	38	Not appearing
287	coletivamente	38	Not appearing
288	alternativamente	38	Not appearing
289	administrativamente	38	Not appearing
290	terrivelmente	37	Not appearing
291	solenemente	37	Not appearing
292	prematuramente	37	Not appearing
293	paulatinamente	37	Not appearing
294	calmamente	37	Not appearing
295	tremendamente	36	Not appearing
296	suavemente	36	Not appearing
297	oralmente	36	Not appearing
298	desnecessariamente	36	Not appearing
299	adicionalmente	36	Not appearing
300	silenciosamente	35	Not appearing
301	pontualmente	35	Not appearing
302	ofensivamente	35	Not appearing
303	exemplarmente	35	Not appearing
304	comparativamente	35	Not appearing
305	brilhantemente	35	Not appearing
306	religiosamente	34	AEDic and HEDic
307	propositadamente	34	Not appearing
308	pesadamente	34	Not appearing
309	absurdamente	33	Not appearing
310	verticalmente	32	Not appearing
311	tragicamente	32	Not appearing
312	preventivamente	32	Not appearing
313	mormente	32	AEDic and HEDic
314	horizontalmente	32	Not appearing
315	delicadamente	32	Not appearing
316	alegremente	32	Not appearing
317	vigorosamente	31	Not appearing
318	inerentemente	31	Not appearing
319	ansiosamente	31	HEDic

Appendix B. Comparison with AEDic and HEDic

---

<i>Number</i>	<i>Adv-mente</i>	<i>Frequency</i>	<i>Dictionary</i>
320	acertadamente	31	Not appearing
321	uniformemente	30	Not appearing
322	racionalmente	30	Not appearing
323	maravilhosamente	30	Not appearing
324	interiormente	30	Not appearing
325	erradamente	30	Not appearing
326	primordialmente	29	Not appearing
327	ordinariamente	29	AEDic
328	maciçamente	29	Not appearing
329	indistintamente	29	Not appearing
330	grosseiramente	29	Not appearing
331	espiritualmente	29	Not appearing
332	textualmente	28	AEDic and HEDic
333	popularmente	28	Not appearing
334	obsessivamente	28	Not appearing
335	gentilmente	28	AEDic
336	generosamente	28	Not appearing
337	defensivamente	28	Not appearing
338	atenciosamente	28	HEDic
339	moderadamente	27	Not appearing
340	conceitualmente	27	Not appearing
341	cegamente	27	HEDic
342	assustadoramente	27	Not appearing
343	apropriadamente	27	Not appearing
344	terminantemente	26	Not appearing
345	sutilmente	26	Not appearing
346	reiteradamente	26	Not appearing
347	cumulativamente	26	Not appearing
348	cotidianamente	26	Not appearing
349	acentuadamente	26	Not appearing
350	remotamente	25	Not appearing
351	incorretamente	25	AEDic
352	eficientemente	25	Not appearing
353	cautelosamente	25	Not appearing
354	sabiamente	24	Not appearing
355	qualitativamente	24	AEDic
356	idealmente	24	Not appearing
357	falsamente	24	Not appearing
358	extensivamente	24	Not appearing
359	experimentalmente	24	Not appearing
360	distintamente	24	Not appearing
361	artesanalmente	24	Not appearing
362	abundantemente	24	Not appearing
363	resumidamente	23	Not appearing
364	intuitivamente	23	Not appearing
365	humildemente	23	Not appearing
366	equivocadamente	23	Not appearing
367	eficazmente	23	Not appearing
368	demoradamente	23	Not appearing

Appendix B. Comparison with AEDic and HEDic

---

<i>Number</i>	<i>Adv-mente</i>	<i>Frequency</i>	<i>Dictionary</i>
369	apressadamente	23	Not appearing
370	apaixonadamente	23	AEDic and HEDic
371	velozmente	22	AEDic
372	rotineiramente	22	Not appearing
373	quinzenalmente	22	Not appearing
374	merecidamente	22	Not appearing
375	declaradamente	22	Not appearing
376	retrospectivamente	21	AEDic
377	retroativamente	21	AEDic
378	preliminarmente	21	Not appearing
379	milagrosamente	21	Not appearing
380	meticulosamente	21	Not appearing
381	materialmente	21	Not appearing
382	magistralmente	21	Not appearing
383	fluentemente	21	Not appearing
384	fixamente	21	Not appearing
385	detidamente	21	Not appearing
386	veementemente	20	AEDic
387	secamente	20	Not appearing
388	respeitosamente	20	Not appearing
389	regionalmente	20	Not appearing
390	quantitativamente	20	AEDic
391	precipitadamente	20	Not appearing
392	institucionalmente	20	Not appearing
393	extra-oficialmente	20	AEDic
394	disfarçadamente	20	Not appearing
395	diametralmente	20	AEDic and HEDic
396	civilmente	20	Not appearing
397	impiedosamente	19	Not appearing
398	globalmente	19	Not appearing
399	ferozmente	19	Not appearing
400	extensamente	19	Not appearing
401	elegantemente	19	Not appearing
402	descaradamente	19	Not appearing
403	cruelmente	19	Not appearing
404	convincientemente	19	Not appearing
405	animadamente	19	Not appearing
406	vulgarmente	18	HEDic
407	vertiginosamente	18	Not appearing
408	semestralmente	18	AEDic
409	orgulhosamente	18	Not appearing
410	operacionalmente	18	Not appearing
411	maliciosamente	18	Not appearing
412	incidentalmente	18	Not appearing
413	harmoniosamente	18	Not appearing
414	digitalmente	18	Not appearing
415	convencionalmente	18	Not appearing
416	ardentemente	18	Not appearing
417	singularmente	17	Not appearing

Appendix B. Comparison with AEDic and HEDic

---

<i>Number</i>	<i>Adv-mente</i>	<i>Frequency</i>	<i>Dictionary</i>
418	serenamente	17	Not appearing
419	ricamente	17	Not appearing
420	inteligentemente	17	Not appearing
421	inocentemente	17	Not appearing
422	estruturalmente	17	Not appearing
423	dolorosamente	17	AEDic
424	dignamente	17	Not appearing
425	desgraçadamente	17	AEDic and HEDic
426	densamente	17	Not appearing
427	corajosamente	17	Not appearing
428	consistentemente	17	Not appearing
429	confusamente	17	Not appearing
430	cognitivamente	17	Not appearing
431	teimosamente	16	Not appearing
432	sumamente	16	Not appearing
433	sorrateiramente	16	Not appearing
434	similarmente	16	Not appearing
435	seletivamente	16	Not appearing
436	ruidosamente	16	Not appearing
437	furiosamente	16	Not appearing
438	familiarmente	16	Not appearing
439	escandalosamente	16	Not appearing
440	efusivamente	16	Not appearing
441	dominantemente	16	Not appearing
442	deliciosamente	16	Not appearing
443	cordialmente	16	Not appearing
444	ulteriormente	15	Not appearing
445	surdamente	15	Not appearing
446	primitivamente	15	Not appearing
447	medianamente	15	Not appearing
448	loucamente	15	Not appearing
449	incondicionalmente	15	Not appearing
450	especificamente	15	Not appearing
451	epistemologicamente	15	Not appearing
452	desordenadamente	15	AEDic
453	criminosamente	15	Not appearing
454	produtivamente	14	Not appearing
455	presentemente	14	Not appearing
456	inadequadamente	14	Not appearing
457	furtivamente	14	Not appearing
458	decentemente	14	AEDic
459	convulsivamente	14	AEDic
460	compulsivamente	14	Not appearing
461	coerentemente	14	Not appearing
462	cabalmente	14	Not appearing
463	bravamente	14	Not appearing
464	abstratamente	14	Not appearing
465	tempestivamente	13	AEDic
466	sigilosamente	13	Not appearing



Appendix B. Comparison with AEDic and HEDic

---

<i>Number</i>	<i>Adv-mente</i>	<i>Frequency</i>	<i>Dictionary</i>
467	redondamente	13	AEDic and HEDic
468	perversamente	13	Not appearing
469	persistentemente	13	Not appearing
470	milimetricamente	13	Not appearing
471	indiferentemente	13	Not appearing
472	graciosamente	13	Not appearing
473	flagrantemente	13	AEDic
474	espantosamente	13	Not appearing
475	criativamente	13	Not appearing
476	correntemente	13	Not appearing
477	amargamente	13	Not appearing
478	afirmativamente	13	AEDic
479	visceralmente	12	Not appearing
480	trimestralmente	12	Not appearing
481	sucintamente	12	Not appearing
482	resolutamente	12	Not appearing
483	pejorativamente	12	Not appearing
484	nervosamente	12	Not appearing
485	mortalmente	12	Not appearing
486	maquinalmente	12	Not appearing
487	mansamente	12	Not appearing
488	humanamente	12	Not appearing
489	fervorosamente	12	Not appearing
490	extrajudicialmente	12	Not appearing
491	distraidamente	12	Not appearing
492	desigualmente	12	Not appearing
493	covardemente	12	Not appearing
494	atabalhoadamente	12	Not appearing
495	ardorosamente	12	Not appearing
496	aceleradamente	12	Not appearing
497	veladamente	11	AEDic
498	toscamente	11	Not appearing
499	tipicamente	11	Not appearing
500	piamente	11	HEDic
501	percentualmente	11	Not appearing
502	penalmente	11	Not appearing
503	opcionalmente	11	Not appearing
504	obstinadamente	11	Not appearing
505	marginalmente	11	Not appearing
506	levianamente	11	Not appearing
507	lateralmente	11	Not appearing
508	insuficientemente	11	Not appearing
509	fracamente	11	Not appearing
510	finamente	11	Not appearing
511	expressivamente	11	Not appearing
512	criteriosamente	11	Not appearing
513	ambientalmente	11	Not appearing
514	subsequentemente	10	Not appearing
515	regressivamente	10	Not appearing

<i>Number</i>	<i>Adv-mente</i>	<i>Frequency</i>	<i>Dictionary</i>
516	preguiçosamente	10	Not appearing
517	perdidamente	10	HEDic
518	penosamente	10	Not appearing
519	parceladamente	10	Not appearing
520	irracionalmente	10	Not appearing
521	injustificadamente	10	Not appearing
522	industrialmente	10	Not appearing
523	incialmente	10	Not appearing
524	frouxamente	10	Not appearing
525	desbragadamente	10	AEDic
526	cruamente	10	Not appearing
527	copiosamente	10	Not appearing
528	contratualmente	10	Not appearing
529	caprichosamente	10	Not appearing
530	calorosamente	10	Not appearing
531	alegadamente	10	Not appearing
532	afetivamente	10	Not appearing