The Mitkov algorithm for anaphora resolution in Portuguese

Abstract This paper reports on the use of the Mitkov´s algorithm for pronoun resolution in texts written in Brazilian Portuguese. Third person pronouns are the only ones focused upon here, with noun phrases as antecedents. A system for anaphora resolution in Brazilian Portuguese texts was built that embeds most of the Mitkov's features. Some of his resolution factors were directly incorporated into the system; others had to be slightly modified for language adequacy. The resulting approach was intrinsically evaluated on hand-annotated corpora. It was also compared to Lappin & Leass's algorithm for pronoun resolution, also customized to Portuguese. Success rate was the evaluation measure used in both experiments. The results of both evaluations are discussed here.

Keywords: Pronoun resolution, anaphora resolution.

1 Introduction

A major problem in Natural Language Processing (NLP) is to recognize or build text segments that convey coherent information. Amongst the linguistic devices for that, referential cohesion is one of the most significant ones for acknowledging, and guaranteeing, coherence. In this paper we address such a phenomenon in the interpretation scenario: we aim at identifying cohesive mechanisms that help automatically resolving referential links. Anaphoricity is the only linguistic construction under focus here, and so is pronoun resolution (PR). An anaphoric pronoun signals a relationship between two or more text components that share with each other their meanings. When anaphoric, it comes after its antecedent referent in the text, which is usually a noun phrase (NP). An example follows[1]:

(1) ***O parlamentar*$_i$**, porém, é alvo de acusação em outro escândalo. ***Ele*$_i$** será investigado sobre as denúncias de corrupção (...).

[**The member of the parliament**]$_i$, however, is enrolled as guilty in other scandal. **He$_i$** will be investigated on the corruption accusations (…).

Above, the pronoun ***Ele*** (He) is an anaphor whose antecedent is the NP ***O parlamentar*** *(*The member of the parliament). This conveys, thus, a full meaning, whilst the pronoun itself is issued the NP meaning. Differently from this, other types of anaphors also may be used, e.g., the generalization introduced by the NP 'the corruption accusations'. This type of anaphor involves ontological associations and other referential links (e.g., 'other scandal' enables the plural 'accusations'). Usually, ontological referential links do not entail the same meaning, but similar meanings that make them connect to each other. Contrarily, anaphoric pronouns convey no meaning, but their antecedent ones. So, to resolve the anaphoric ***Ele*** above, retrieving the former NP is of utmost importance.

---

[1] In this paper, all the examples have been extracted from a Brazilian Portuguese corpus of authored texts. Their literal English translations are supplied for readability.

PR may be very complex when there are many antecedent candidates. To resolve that, varied linguistic features may be needed, involving morphologic, syntactic, semantic, or pragmatic processing. In this line, several computational approaches have been undertaken for English or Spanish. Examples of the former are those by (Hobbs, 1978), (Lappin & Leass, 1994), (Mitkov, 1998), and (Bergsma & Lin, 2006); an example of the latter is the work by (Palomar *et al*., 2001). There are few approaches to Brazilian Portuguese (BP), namely, modified versions of Lappin & Leass' algorithm (Coelho, 2005) and Hobbs' model (Santos & Carvalho, 2007), and another that uses heterogeneous knowledge for PR (Paraboni, 1997).

We adopted Mitkov's model for the following reasons: (1) it has been explored for several languages and it has proved to be both language-independent and portable; (2) it is heuristics-based and does not depend upon deep knowledge. Instead, it applies surface or empirical information to determine candidate antecedents of an anaphor. It also adopts usual parsing and morphological preprocessing tools, which are largely available for most languages currently explored in NLP, and so they are for BP. In our implementation, Mitkov's original algorithm has been modified to handle only 3rd person pronouns that convey NPs as antecedents. An antecedent NP may be in the very same sentence as the anaphor (intra-sentential PR) or in a different sentence (inter-sentential PR).

Mitkov introduces the so-called anaphora resolution factors, which are of two types. *Restrictive factors* signal properties that must be presented by the antecedent candidates, in order for them to resolve the anaphor. Those candidates that do not present such properties are thus discarded. Oppositely, *preferential factors* do not discard candidates; just classify them according to their likelihood of resolving the anaphor the best. Usually, the latter factors are applied in conjunction with the former: classification takes place only after filtering those potential candidates through factors that are restrictive. The highest the probability of a candidate, the more likely it is to be indicated as the antecedent of the anaphor under focus.

In what follows, we first present Mitkov's proposal for anaphora resolution (Section 2), then we introduce our approach for PR in BP (Section 3). In Section 4 we describe our PR assessment. Final remarks are presented in Section 5.

2 The Mitkov's algorithm

In pursuing a knowledge-poor approach, Mitkov avoids complex syntactic and semantic constructions by adopting a set of heuristics – the antecedent indicators – that are capable of pinpointing potential antecedents of an anaphor based on surface indicators. First, the text under focus is preprocessed through parsing and extracting NPs. Then, the following steps take place: (1) at most two previous sentences to the anaphor are examined, as the referential context for its antecedent NP. The result is a set of NPs. (2) to narrow the number of candidate NPs, gender and number concordance with the anaphor is verified and potential NPs are excluded, resulting in a smaller candidate set, or the *set of potential antecedents*. (3) Each potential NP is thus scored for the likelihood of being the antecedent of the anaphor. First, each antecedent indicator allows issuing the NP an integer value ranging from -1 to 2. Then such values are summed up for the final NP score. The highest scored NP in the set of potential candidates is finally chosen as the antecedent. In a tie case, the closest candidate to the anaphor is chosen instead.

3 Adapting Mitkov's algorithm for PR in Brazilian Portuguese

In adapting Mitkov's algorithm for BP, we named our system RAPM, for ***R**esolução **A**nafórica do **P**ortuguês baseada no algoritmo de **M**itkov* (Anaphora Resolution for Portuguese based on Mitkov's algorithm). RAPM differs from the original algorithm in that it has been specifically drawn to BP. Also, and most importantly, its input texts are only automatically annotated, which is not the case in Mitkov's approach, because its morphosyntactic annotations are manually corrected before going into anaphora resolution itself. A third distinction concerns the way preprocessing is undertaken in RAPM: currently it has not yet been integrated in RAPM, for practical reasons, as we shall describe in Section 4. Finally, to resolve morphological dependencies RAPM looks up an XML file that conveys correct information on gender and number of proper nouns and the antecedent search scope is of three sentences, instead of two. The XML file includes every proper noun found in the input corpus of texts and aims at minimizing pre-processing problems: in the absence of such information, they would be assigned both genders and numbers.

RAPM proceeds, thus, in the following way: it identifies the NPs that appear previously to the pronouns using the three-sentence window, then it produces the set of potential NP candidates in the same way as Mitkov's model does. As already mentioned, antecedent indicators may endorse or prevent an NP candidate of being the antecedent of a pronoun. In the former case, the sum total is positive; in the latter, it is negative. Only five out of eleven antecedent indicators by Mitkov were incorporated into RAPM, along with three others that we found interesting to add, as follows the last three are the novel ones:

First NP (FNP). A +1 score is issued to the first NP of each sentence. This heuristic may be either justified on syntactic, or on communication terms, in that human beings usually express meanings through distinct language levels. According to Mitkov, for example, in declarative sentences the FNP occupies the subject position. In the absence of a parse tree, theories of both communication and discourse organization might help determining which should be such FNP, provided that they signaled the underlying communicative or discourse structure of the focused text. If Centering Theory (Grosz *et al*., 1995) were considered, e.g., the FNP would be the actual center of the sentence. Conversely, if (Ventura & Lima-Lopes, 2002) or (Firbas, 1992) were used, the corresponding given-new or thematic-rhematic information should signal that. In either case, given-new or theme-rheme units could provide coreferential links as do a first NP and a pronoun, at the text surface. It is worthwhile to notice that, since declarative sentences convey a default discourse organization, the theme/given unit can be the very NP occupying the subject position, as pinpointed before.

Lexical Reiteration (LR). A +2 score is issued to NPs that occur twice or more within the search scope; a +1 score is issued to an NP otherwise. LR assumes that a greater score signals NPs that are more salient and, thus, more likely to be the anaphor antecedent than those that score less. In RAPM reiterated lexical items are identified through direct string matching.

Indefinite NP (INP). Indefinite NPs are assigned a -1 score because very often they are supposed to be less likely to be antecedents of pronominal anaphors than definite ones (Mitkov, 1997). RAPM regards an NP as definite if its nucleus is modified by a definite article or by demonstrative or possessive pronouns.

Prepositional NP (PNP). A -1 score is issued to those NPs that occur in a prepositional phrase. Such a demoting score may be justified by the Centering Theory (Grosz *et al*., 1995): the sentence main constituents are classified according to their salience (e.g., subject, direct, and indirect objects in this decreasing salience order) and the most salient units provide the

center of a text segment. Moreover, a sentence center is more likely to be a candidate antecedent of an anaphor than an NP occurring in a prepositional phrase sentences.

Referential Distance (RD). This antecedent indicator may promote or demote a candidate according to its distance from the anaphor: NPs in the immediate antecedent clause, in the very same sentence as the anaphor, are scored +2; NPs in the previous sentence are scored +1; NPs in a sentence that is two sentences apart from that of the anaphor are scored 0; NPs still farther than those are scored -1.

Syntactic Parallelism (SP). A +1 score is issued to an NP that conveys the same syntactic function as the corresponding anaphor.

Nearest NP (NNP). A positive +1 score is issued to the nearest NP to the anaphor. This indicator is used as a baseline by Mitkov, and so it is in RAPM, in which case it corresponds to the so-called 'Baseline_NP'[2].

Proper Noun (PN). Proper nouns are scored +1 in RAPM because they occurred with relative frequency as anaphors antecedents in our corpus. The assumption behind such a score was that promoting PNs could improve PR performance.

Using the above antecedent indicators to resolve pronouns in BP aims only at predicting the language behavior concerning PR, and not at constraining automatic PR to them. For this reason, Mitkov called them *preference factors*: they are not intended to be definite, but deductible. Although such indicators may punctuate anaphor antecedents incorrectly, usually PR is improved when they are used altogether, as we shall show when we apply them to BP, in Section 4.

Adding the indicators SP, NNP, and PN to RAPM resulted from a corpus analysis that aimed at filtering out those Mitkov's antecedent factors that didn't apply to BP. Thus, they were chosen for the following reasons: (a) since the input texts to RAPM are already morphosyntactically annotated, syntactic parallelism (SP) could be readily verified; (b) a nearest NP (NNP) to the anaphoric pronoun tended to be its antecedent; (c) proper nouns (PN) were highly frequent in the corpus as anaphor antecedents. So, using them should be advantageous for PR in RAPM.

Excluding the remaining six Mitkov's indicators from RAPM was due to their inadequacy to the corpus under focus in our work. For example, the original indicator 'Section heading preference' did not apply to our corpus, which conveys only non-structured or non-titled texts. Amongst RAPM eight indicators, two are restrictive (INP and PNP) and one (RD) may be either restrictive our preferential, for its possibility of assigning negative or positive values to a candidate. The remaining indicators are preferential. As illustration of the RAPM processing, consider the journalistic text segment (2).

(2) O flúor fortifica o esmalte, uma espécie de capa protetora dos dentes. Com a difusão de seu uso, outro problema surgiu: a fluorose, o excesso de flúor no organismo. Afinal, ***a substância*$_{\mathbf{i}}$** não se encontra apenas em cremes dentais: ***ela*$_{\mathbf{i}}$** também está presente em diversos alimentos (...).

---

[2] Baselines are used for assessment and will be described in Section 4.

The fluorine fortifies the enamel, a sort of protective cape of the teeth. With the diffusion of his use, another problem appeared: the fluorosis, the excess of fluor in the organism. At last, ***the substance$_i$*** is not only in toothpastes: ***she$_i$*** also is present in several foods (...).

For the text (2), the described algorithm generate the set of potential candidates (NP) with agree in gender and number {G,N}[3] with the anaphor ***ela*** (she){F,S} as (NP_1: [uma espécie de capa protetora dos dentes]{F, S}, NP_2: [capa protetora dos dentes]{F,S }, NP_3: [a difusão de seu uso]{F,S}, NP_4: [a fluorose] {F, S} NP_5: [a substância]{F, S}.

The last stage of PR assigns the scores positive or negative to the NP candidates. In Table 1 are presented the scores associated to the 5 NPs previous, organized them of descending form by his scores.

NP candidate	FNP	LR	SP	NNP	PN	INP	PNP	RD	∑
NP_5	0	0	1	0	0	0	0	1	2
NP_4	0	0	0	0	0	0	0	0	0
NP_3	0	0	0	0	0	0	-1	0	-1
NP_1	0	0	0	0	0	-1	0	-1	-2
NP_2	0	0	0	0	0	-1	-1	-1	-3

Table 1 : Assigned scores by antecedent indicators in PR process

The noun phrase (NP_5), ***a substância*** (the substance), is selected as antecedent for anaphor ***ela*** (she) due to his biggest total score indicated in the column ($\sum = 2$).

4 Assessing RAPM

We adopted the success rate measure (Mitkov, 2002) to assess RAPM. It is defined as the ratio between the total number of correctly resolved anaphors and the total number of anaphors that are present in the whole corpus of texts. Having the input texts previously annotated for anaphors by human experts was, thus, of utmost importance here: such a corpus is a reference corpus.

According to Mitkov, the success rate should mirror exactly the performance of the anaphora resolver itself, with no interference of any problem resulting from preprocessing. He emphasizes that the real success rate of a system may only be achieved if the input data are correct. So, he corrects by hand any possible annotation of the input texts that have been wrongly morphosyntactically tagged, as previously mentioned. However reasonable his arguments may be, although the current version of RAPM does not have an integrated preprocessing module (both working in pipeline instead), we consider such pre-editing unrealistic. So, we did not apply any correction procedure to our input data, aiming at a more realistic black box approach in the future. More importantly, we already consider PR to be fully automated, as if we had just plain texts as input. A possible drawback of this is that miss-annotated data may contribute negatively to the PR performance, as we shall discuss below.

---

[3] {Gender, Number}: F-Femme, M-Male; S-Singular, P-Plural.

In assessing RAPM, we used three corpora of distinct genres – the ones also adopted by (Coelho, 2005): a law, a literary, and a newswire one. The law corpus is composed of 16 Portuguese Attorney General documents (c.a. 110,610 words; 260 3rd person pronominal anaphors) and most texts convey long and complex sentences. The literary corpus, also conveying complex sentences, consists of the whole book 'O alienista', by the Brazilian author Machado de Assis (c.a. 16,530 words; 573 3rd person pronominal anaphors). The newswire corpus is composed by 14 texts of the Veja magazine (c.a. 13,217 words; 222 3rd person anaphoric pronouns) and it conveys simpler sentences than the others.

Our comparison with Coelho's approach (Coelho, 2005) was due to the lack of comprehensive assessments on BP for anaphora resolution: his was the only one that used success rate as the evaluation measure in a more comprehensive way and, apart from the law corpus, his corpora aimed at a general readership, oppositely to the others, which were more domain-dependent, such as (Paraboni, 1997). The extra gain in adopting his corpora was that we did not have to preprocess our data, since Coelho had it all done. We even did not have to re-run Coelho's system: we entirely reproduced his experimental setting. This allowed us decide to keep preprocessing apart from RAPM. Actually, we just used the files that were formerly automatically annotated by Coelho as input to RAPM. His tags provide morphosyntactic information resulting from the PALAVRAS parser (Bick, 2000) and co-referential annotation. This has been manually carried out with the support of the MMAX tool (Müller & Strube, 2001). Additionally, the Xtractor tool (Gasperin *et al*., 2003) was used to convert the output by PALAVRAS in XML files.

RAPM assessment consisted of comparing the file of each text of the reference corpus with the corresponding file generated automatically by the system. This file, also in XML, contains all the tags that identify both anaphoric pronouns and their corresponding antecedents. Figures 1 and 2 depict a record of each file concerning example (1) given in Section 1. For assessment purposes, an anaphor is considered correctly resolved either if the automatic solution is identical to the manual one or if it is an NP that is the nucleus or part of the nucleus of the NP manually annotated. Acknowledging either case has been manual.

```
<markable id="markable_406" span="word_147..word_148" form="antecedent" />
<markable id="markable_229" span="word_160" form="pronome" pointer="markable_406" />

<pron chunk_id="chunk_229" gender="M" number="3S" reci="no" refl="no" span="word_160" />

<np chunk_id="chunk_212" gender="M" head_span="word_148" number="S"
    span="word_147..word_148" />
```

Figure 1 : Snapshot of the reference file conveying example (1) annotation

```
<RAPM>
  <IdAnafora>chunk_229</IdAnafora>  <Anafora>Ele</Anafora>
  <IdAntecedente>chunk_212</IdAntecedente>  <Antecedente>O parlamentar</Antecedente>
</RAPM>
```

Figure 2 : Snapshot of the file conveying example (1) RAPM annotation

Aiming at a broad assessment, we derived several versions of RAPM by combining the antecedent indicators in an ad-hoc manner. The resulting indicator sets varied in size and were configured amongst those that seemed to be most promising for PR in BP. Hereafter, each version is identified by "RAPM_n", n signaling the amount of antecedent indicators

considered. Overall, eight distinct versions were provided, as follows (IS stands for the Indicators Set considered):

- **RAPM_2**: IS = {INP, RD}
- **RAPM_3**: IS = {INP, PNP, RD}
- **RAPM_4**: IS = {INP, PNP, RD, NNP}
- **RAPM_5**: IS = {FNP, LR, INP, PNP, RD}
- **RAPM_6_SP**: IS = {FNP, LR, INP, PNP, RD, SP}
- **RAPM_6_NNP**: IS = {FNP, LR, INP, PNP, RD, NNP}
- **RAPM_6_PN**: IS = {FNP, LR, INP, PNP, RD, PN}
- **RAPM_8**: IS = {FNP, LR, INP, PNP, RD, SP, NNP, PN}

Combining indicators as in the above versions was merely based upon a previous corpus analysis to find out which, amongst all Mitkov's factors, would fit RAPM. The analysis consisted of comparing automatically generated sums for the potential sets of antecedent candidates with reference annotations. Initially, the indicators with the most promising correction scores were combined, yielding the first 3 versions of RAPM. Those indicators were INP, RD, PNP, and NNP classified INP>NNP>RD>PNP (X>Y indicating X score greater than Y score). Differently from these and RAPM_8, which conveys all the indicators, RAPM_5 considers only those also managed by Mitkov. Each of the 3 RAPM_6 versions was built adding to the RAPM_5 set each new indicator we introduced, one at a time.

RAPM assessment was undertaken in three different ways: firstly, we measured the average success rate of each system depicted above, when running on the newswire corpus (Table 2). The strategy with the best success rate (RAPM_8) was then used in two other experiments: we compared its performance with the results by (Coelho, 2005) (Table 3) and finally we used again RAPM_8 results on the newswire corpus, but to compare it with two distinct baselines, namely, 'Baseline-NP' and 'Baseline_Subj' (Table 4). In this last case, we used the same baselines as did Mitkov in (Mitkov, 2002). Baseline-NP pinpoints as the antecedent the closest NP to the pronoun, provided that the NP agrees in gender and number with the pronoun. Baseline_Subj adds to the Baseline-NP a third constraint: the antecedent NP must occupy the subject position in the sentence it occurs. The results of each assessment follow, in a decreasing success rate order.

RAPM version	**Success rate (%)**
RAPM_8	67,01
RAPM_3	66,02
RAPM_6_NNP	64,94
RAPM_6_PN	63,40
RAPM_2	62,50
RAPM_5	61,45
RAPM_4	61,21
RAPM_6_SP	60,26

Table 2 : Overall assessment

Corpus	RAPM_8	Coelho (2005)
Newswire	67,01	43,56
Literary	38	31,32
Law	54	35,15

Table 3 : Comparison between RAPM_8 and Coelho's av. success rates

PR systems	Success rate (%)
RAPM_8	67,01
Baseline-NP	55,49
Baseline_Subj	42,27

Table 4 : Comparison between RAPM_8 and baseline strategies

As shows Table 2, RAPM_8 was the system that had the best average success rate (67%). So, it was the only system used in the other two experiments. Although RAPM_8 performed better in the overall assessment, its use may be questionable because the system that was classified the second, RAPM_3, presented a close success rate (66%) using much fewer antecedent indicators. This result suggests that using impeditive indicators, i.e., INP and PNP, may well help resolving pronouns in BP, when newswire texts are considered, and is less costly. Even RAPM_6_NNP, which reached the 3rd best av. success rate, also performs closely to RAPM_8, and demands less indicators. Comparing the 3 versions RAPM_6, adding NNP to the original 5 indicators seems to be the only one that may slightly improve the success rate. Still, it does improve on RAPM_5 in 4 perceptual points. Comparing now the success rates of our eight systems with those by Coelho, ours were consistently superior: 43,56% by Coelho, against our worst case for RAPM_6_SP (60,26%), for the newswire corpus. If we consider only RAPM_8, the outperforming is even more expressive: for the same corpus, RAPM_8 scored an average of 67,01% success rate. Even for the other corpora RAPM_8 expressively beat Coelho's system. The third comparison, between RAPM_8 and the two baselines, also confirms RAPM_8 superiority.

5 Final Remarks

Given the results presented above, it is clear that adapting Mitkov's algorithm for pronoun resolution in Brazilian Portuguese is promising, even having it modified for suitability due to language differences. Actually, our RAPM_8 version conveys not only the 5 indicators inherited from that approach, but the ones we introduced. This may be a novel and contributing feature for resolving anaphoric pronouns in BP. Moreover, having 8 distinct versions of RAPM shows that any of the proposed combinations considerably outperforms the baseline procedures, when dealing with BP. All our systems are even more suitable than the one based on Lappin & Leass model, at least in what concerns the corpora explored here. Most importantly, our assessment was quite comprehensive, if we consider the varied combinations of antecedent indicators.

Concerning the original approach by Mitkov (89,7% av. success rate), we could crudely say that RAPM_8 still has a significant room for improvement. However, it is important to notice that achieving that high rate has been possible because human expertise and hard man-work has been provided to feed the system with correct input data. This contrasts very deeply with

our strategy, which aims at getting bare texts and resolving its pronominal anaphors entirely automatically. RAPM_8 performance is even more understandable when we notice that there was an expressive noise introduced by automatic morphosyntactic annotation. The main errors generated in that phase included wrong morphological annotations of both NPs and pronouns, which generated information that was certainly crucial for the low success rate for PR. During the pronoun resolution itself, an expressive problem was that of accusing pronoun antecedents that were not NPs. RAPM was not tailored to deal with such cases, even when such antecedents were correct. So, it is very likely that overcoming such obstacles will improve RAPM performance. A deeper evaluation of the impact of the errors presented by RAPM_8 as a result of pre-processing shall be carried out in that direction. Also, verifying the adequacy of the values assigned to each case and each antecedent indicator is relevant. However, this is not straightforward to accomplish, for it involves scaling up the same method of linguistic analysis that we carried out, which was entirely dependent upon human expertise. It also involves considering other corpus-based means to verify the indicators adequacy. Moreover, having RAPM_8, which conveys all the antecedent indicators considered, as the best system for PR in BP, does not entitle us to say that RAPM_8 will work well when other data are used. There are many other ways of exploring further the present results, including verifying which combination of the values the indicators can assign to an NP would be more profitable. Aiming at this seems quite reasonable, since our approach is entirely empirical. However, it is not less complex: we could have too many combinations of values and features to investigate. So, considering other, statistical methods to pinpoint a more reliable feature combination should be also applicable.

In tuning the system, we could also consider other features, such as other syntactic filters, c-command constraints (Reinhart, 1983), or even the use of the Center Theory. Finally, we could also carry out a previous hand-editing of the input data, as did Mitkov, to certify that the input to our explored RAPM versions would not be significantly responsible for the PR problems. In such a case, we should solve first the preprocessing problems and turn back to reengineering the system. Overall, RAPM may be useful for several NLP applications, including Automatic Summarization and Information Retrieval, which are the ones focused more closely in our research. By resolving anaphoric pronouns, automatically generated summaries may be more cohesive and, thus, more coherent. Using heuristics especially driven to deal with BP phenomena will certainly bring about interesting contributions to such applications.

References

BERGSMA, S. & LIN, D. (2006). Bootstrapping path-based pronoun resolution. *COLING-ACL 2006*, 33-40.

BICK, E. (2000). *The parsing system PALAVRAS: automatic grammatical analysis of Portuguese in a constraint grammar framework*. Ph.D. thesis. Århus University, Århus.

COELHO, T.T. (2005). *Resolução de anáfora pronominal em português utilizando o algoritmo de Lappin e Leass*. Master`s thesis. University of Campinas.

FIRBAS, J. (1992). *Functional sentence perspective in written and spoken communication*. Cambridge: Cambridge University Press.

GASPERIN, C.V., VIEIRA, R., GOULART, R.R.V. & QUARESMA, P. (2003). Extracting XML chunks from portuguese corpora. *Proceedings of the workshop on traitement automaticque dês langues ninoritaries*. Batz-sur-Mer.

GROSZ, B.J., JOSHI, A.K. & WEINSTEIN, S. (1995). Centering: a framework for modeling the local coherence of discourse. *Computational linguistics* 21, 203-225.

HOBBS, J.R. (1978). Resolving pronoun references. *Lingua* 44, 311-338.

LAPPIN, S. & LEASS, H.J. (1994). An algorithm for pronominal anaphora resolution. *Computational linguistics* 20, 535-561.

MITKOV, R. (2002). *Anaphora resolution*. Longman.

MITKOV, R. (1998). Robust pronoun resolution with limited knowledge. *COLING-ACL 1998*, 869-875.

MITKOV, R. (1997). Factors in anaphora resolution: they are not the only things that matter. A case study based on two different approaches. *ACL-EACL 1997 Workshop on operational factors in practical, Robust anaphora resolution*, 14-21.

MÜLLER, C. & STRUBE, M. (2001). MMAX: a tool for the annotation of multi-modal corpora.*The 2nd IJCAI Workshop on knowledge and reasoning in practical dialogue systems*, 45-50.

PALOMAR, M., MORENO, L., PERAL, J.,MUÑOZ, R., FERNÁNDEZ, A., MARTÍNEZ-BARCO, P. & SAIZ-NOEDA, M. (2001). An algorithm for anaphora resolution in spanish texts. *Computational linguistics* 27, 545-567.

PARABONI, I. (1997). *Uma arquitetura para a resolução de referências pronominais possessivas no processamento de textos em língua portuguesa*. Master's thesis, PUC, Rio Grande do Sul.

REINHART, T. (1983). *Anaphora and semantic interpretation*. London: Croom Helm.

SANTOS, D.N.A. & CARVALHO, A.M.B.R. (2007). Hobbs' algorithm for pronoun resolution in portuguese. *6th. MICAI 2007: Advances in artificial intelligence* 4827, 966-974, New York: Springer-Verlag Berlin Heidelberg.

VENTURA, C.S.M. & LIMA-LOPES, R.E. (2002). O tema: caracterização e realização em Português. *DIRECT Papers* 47, 1-18.