Applying Lexical-Conceptual Knowledge for Multilingual Multi-Document Summarization

Ariani Di Felippo[1,2], Fabrício E. S. Tosta[1], Thiago A. S. Pardo[1,3]

[1] Interinstitutional Center for Computational Linguistics (NILC), São Carlos/SP, Brazil
[2] Language and Literature Department (DL), Federal University of São Carlos (UFSCar)
Rodovia Washington Luís, km 235 - SP 310, São Carlos, 13565-905, Brazil
[3] Institute of Mathematical and Computer Sciences (ICMC), University of São Paulo (USP)
Avenida Trabalhador São-carlense, 400, São Carlos, 13566-590, Brazil
arianidf@gmail.com, fabricio3341@hotmail.com, taspardo@icmc.usp.br

Abstract. We define Multilingual Multi-Document Summarization (MMDS) as the process of identifying the main information of a cluster with (at least) two texts, one in the user's language and one in a foreign language, and presenting it as a summary in the user's language. Although it is a relevant task due to the increasing amount of on-line information in different languages, there are only baselines for (Brazilian) Portuguese, which apply machine-translation to obtain a monolingual input and superficial features for sentence extraction. We report our investigation on the application of *conceptual frequency* measure to build a summary in Portuguese from a bilingual cluster (Portuguese and English). The methods tackle two additional challenges: using Princeton WordNet for nouns annotation and applying MT to translate selected sentences in English to Portuguese. The experiments were performed using a *corpus* of 20 clusters, and show that lexical-conceptual knowledge improves the linguistic quality and informativeness of extracts.

Keywords: multilingual, multi-document, summarization, concept, extract.

1 Introduction

As the amount of on-line news texts in different languages is growing at an exponential pace, Multilingual Multi-Document Summarization (MMDS) is a quite desirable task. It aims at identifying the main information in a cluster of (at least) two texts, one in the user's language and one in a foreign language, and presenting it as a coherent/cohesive summary in the user's languages. However, MMDS is a highly challenging task, since it requires merging content in different languages as well as dealing with the classical multi-document issues, such as capturing the most relevant content, and maintaining the coherence/cohesion of summary by treating redundancy.

The few previous methods usually consist of two steps: translation of the foreign texts and summarization [1] [2] [3] [4]. The first step is performed by some machine-translation (MT) engine, producing a monolingual multi-document cluster. Then, an extractive[1] multi-document summarization (MDS) method is used to build the summaries, which sometimes treats redundancy. About the input, Roark and Fisher [2] extract sentences from the machine-translated and the original texts in the user's language. Consequently, the summaries present ungrammatical sentences and

[1] Summarization technique that involves ranking sentences using some scoring mechanism, picking the top scoring sentences, and concatenating them in a certain order to build the summary [5].

disfluencies resulting from MT. Evans et al. [1] [3] only extract sentences from the translated texts, and replace them with similar ones from the text in the user's language. This method avoids the MT problems, but the content selection does not take into account the information from the text in the user's language. As an attempt to address both problems, Tosta et al. [4] extract sentences from machine-translated and original texts, and only replace selected sentences with MT problems by similar ranked ones from the text in the user's language. The research of Tosta et al [4] was the first on MMDS involving the (Brazilian) Portuguese language. About the summarization step, the extractive methods are predominantly superficial, based on features such as *word frequency*, *sentence position,* etc., which usually have lower cost and are more robust, but produce poor results.

We turn to the use of *conceptual knowledge* in MMDS, which has already been used in other summarization tasks in order to achieve a better content selection (e.g., [6], [7], [8], [9]). This work makes the assumption that such knowledge allows to take into account information from all source texts in their original language to perform content selection, producing better summaries both in terms of informativeness, since the selection is based on salient concepts, and linguistic quality, because only summary sentences in a foreign language require to be translated.

Particularly, we report our investigation on 2 methods for summarizing a bilingual cluster (Portuguese and English) to produce an extract in Portuguese. Both methods use the frequency of occurrence of the nominal concepts in the cluster to score the sentences. The scoring yields a ranking in which the sentences with the most frequent or redundant concepts are in the top positions. Given the sentence ranking, one content selection strategy is taking the top-ranked sentences in the user's language, avoiding redundancy. The other one only consists of selecting the top-ranked sentences, independently of language, also avoiding redundancy. If sentences in the foreign language are selected, they are automatically translated to the user's language.

Our experiments were performed using the CM2News *corpus*[2] [10], with 40 news texts grouped by topic in 20 clusters. Each cluster has 1 text in Portuguese and 1 in English. The concepts of CM2News were derived from Princeton WordNet[3] (WN.Pr) [11] in a semi-automatic annotation process, including (i) translation of each noun in Portuguese to English (since the *synsets* are in English), and (ii) selection of the *synset* that represents the underlying concept/sense of each noun in Portuguese and English. The experiments show that the conceptual knowledge improves summaries in terms of linguistic quality and informativeness, confirming our hypotheses.

This main contributions of this work are: being the first investigation that proposes semantic methods for MMDS of (Brazilian) Portuguese texts, outperforming a *first-sentence* baseline method [4]; providing a semantic layer of annotation to the CM2News *corpus*, and adaptation of an editor for multilingual sense annotation.

In Section 2, we describe some related works. In Section 3, we describe the lexical-conceptual methods. In Section 4, the *corpus* annotation is described. The evaluation will be discussed in Section 5. In Section 6, some final remarks will be given.

---

[2] http://www.nilc.icmc.usp.br/nilc/index.php/team?id=23#resource

[3] A semantic network of English in which the meanings of word forms and expressions of noun, verb, adjective, and adverb classes are organized into "sets of synonyms" (*synsets*). Each *synset* expresses a distinct concept/sense and the *synsets* are interlinked through conceptual-semantic (i.e., hyponymy, meronymy, entailment, and cause) and lexical (i.e., antonymy) relations [11].

2 Related Work

The closest works to ours are [1] [2] [3] [4]. Roark and Fisher [1] take as input a cluster of some translated texts to English, some English spoken language texts, and some English texts. The method ranks the sentences from all the texts based on 9 superficial features and sets a high preference for English sentences when selecting them from the ranking to compose the English extract. Of the nine features, 8 are different versions of *tf-idf, log-likelihood ratio,* and *log-odds ratio* lexical measures, and the ninth is the position of the sentence in the text. The method was trained on a subset of 80 clusters from DUC 2005 using the SVMlight machine-learning algorithm, but the authors do not provide details about evaluation.

Evans et al. [3] aim at generating an English extract from a cluster of English texts and machine translations of Arabic texts into English. The machine-translated sentences are ranked by DEMS [6], a summarizer which apply 3 main criteria of relevance: identifying importance-signaling words through an analysis of lead sentences in a large *corpus* of news, identifying high-content verbs through a separate analysis of subject-verb pairs news corpus, and finding the dominant concepts[4] in the input clusters of texts. Additionally, the sentence relevance also relies on some of the most widely superficial features, such as *position*, which increases the weight of sentences near the beginning of texts, and *length*, which penalizes sentences that are shorter or longer than a threshold, etc. The sentences selected from the rank are replaced with similar sentences from the English texts. The similarity is computed at clause or phrase level, which requires the syntactic simplification of the English sentences. Next, the similarity is performed by Simfinder [12], which uses lexical and syntactic features. For evaluation, the authors have used the DUC 2004 *corpus*, which contains 24 topics with English texts, Arabic texts, Arabic-to-English machine translations, and 4 human summaries. Using ROUGE[5] [13], the automatic evaluation shows that the similarity-based summarization approach outperforms a *first-sentence* baseline[6]. In an early work, Evans at al. [1] have developed a multilingual version of the English-based summarizer Columbia Newsblaster[7]. This version starts with machine-translated texts, and also replaces the extracted sentences with similar ones in English. However, the similarity is computed at the sentence level, not requiring any syntactic simplification of the non-English sentences.

Tosta et al [4] have proposed 2 *baselines* using 10 clusters to build extracts in Portuguese. Each cluster is composed of 3 news texts, each one in a different language (English, Spanish and Portuguese). The methods are considered *baseline*

---

[4] The nouns are grouped into concept sets using WN.Pr *synsets*, and hyponymy relation. To build a set, the highly polysemous nouns are not disambiguated, but replaced by others that are strongly related with the same verb (e.g., "officer" is replaced by "policeman" due to the relation with "arrest"). Having the sets, the sentence ranking is based on the concepts frequency [6].

[5] ROUGE (*Recall-Oriented Understudy for Gisting Evaluation*) computes the number of common n-grams among the automatic and reference/human summaries, being able to rank automatic summaries as well as humans would do, as its author has shown [14].

[6] In this method, the first-sentence from each text in the cluster is selected until a maximum of words/bytes is reached, and, if the first sentence was already included from each text in the set, the second sentence from each text is included in the summary, and so on [3].

[7] http://newsblaster.cs.columbia.edu/

because they rely on: (i) translation of the foreign texts to Portuguese using MT[8], and (ii) selection of the relevant sentences using established superficial features, i.e., *word frequency* and *sentence position* [14]. To avoid redundancy, the traditional *word overlap* measure is calculated between each candidate sentence of the rank and the summary sentences. If an ungrammatical translated-sentence is selected, *word overlap* is also used to find a similar sentence from the Portuguese text. The methods were intrinsically evaluated according to the linguistic quality of their summaries. The authors have used the 5 criteria of DUC [15]: (i) grammaticality (i.e., no occurrence of datelines, capitalization errors or ungrammatical sentences), (ii) non-redundancy (i.e., no unnecessary repetition), (iii) referential clarity (i.e., easy identification of the pronouns and noun phrases references), (iv) focus (i.e., it should only contain information that is related to the rest of the summary), and (v) structure and coherence (i.e., it should be well-structured, not just be a heap of related information). In such evaluation, the *sentence position* method had better results.

All methods overviewed in this section first apply MT to translate the foreign texts, obtaining a monolingual cluster. However, when the content is extracted exclusively from machine-translated texts, the summary might contain sentences that are ungrammatical and difficult to understand, since MT is far from perfect. And, when the approaches use texts that were automatically translated to guide selection from the texts in the preferred language, relevant information that exclusively occurs in the preferred language is not selected to compose the summary. Thus, it would be more appropriate to take all the texts in their original language, since the goal is to detect the most relevant information of the "cluster". Moreover, the approaches are mainly based on flat text features. Thus, in this paper, we exploit deep linguistic information for MMDS, particularly the conceptual knowledge. Some works have already focused on concepts and their relationships for different summarization tasks under the assumption that they provide a richer representation of the source. For example, Wu and Liu [8] detect the main subtopics of texts by indexing the words to the concepts of a domain-related ontology[9]. The second-level concepts with higher counts codify the main subtopics. Paragraphs that are "closest" to the subtopics are selected. A similar idea but with additional structural features was proposed by Hennig et al. [9] for sentence scoring. The features they used were *tag overlap, subtree depth* and *subtree count*. Next, we describe our extractive MMDS strategies.

3 Lexical-Conceptual Strategies for MMDS

For describing the extractive MMDS strategies, we take into account the traditional summarization phrases: *analysis, transformation* and *synthesis* [17]. The analysis corresponds to the texts understanding, producing an internal representation of their content. The transformation performs summarization operations on the internal representation, producing the summary internal representation. In the synthesis, the summary internal representation is linguistically realized into the final summary. In our methods, the analysis consists of identifying the concepts expressed by (common) nouns (words, expressions, and abbreviations), which are the most frequent word

[8] http://translate.google.com/

[9] This "ontology" consists of a generalization/specialization hierarchy of concepts (i.e., a taxonomy).

class, covering part of the main content of the texts. To identify the nominal concepts, we use WN.Pr as the conceptual repository. We acknowledge that the granularity of the concepts inventory in WN.Pr is often too fine-grained, resulting in difficulties for finding the *synset* that best represents an underlying concept. Even though, the decision of using WN.Pr was due to (i) its widespread use in the area for summarization and also for other applications, (ii) it has been manually produced, and (iii) the current partial development state of most of the similar resources for Portuguese. Since a concept in WN.Pr is codified by a set of synonyms word forms in English (i.e., a *synset*), the annotation of the nouns from texts in Portuguese has an additional challenge: the translation of nouns to English. Here, we have performed an automatic annotation with subsequent manual or human revision. In Section 4, we describe the *corpus* as well as the semi-automatic annotation procedure.

The transformation corresponds to the content selection. To select the sentences, our methods perform 4 steps: (i) computing the compression rate (i.e., the desired summary size), (ii) calculation of the frequency of each nominal concept in the cluster, (iii) scoring all the sentences according to the frequency of occurrence of their nominal concepts in the cluster, and (iv) ranking the sentences by their score. Particularly about the step (ii), the *concept frequency* measure captures the content of the multilingual cluster by counting the occurrence of the concepts underlying synonyms (i.e., different words that express the same concept) and equivalences (i.e., expressions of a concept in different languages). For example, the 2 sentences in Table 1 are from the same cluster and the concepts expressed by nouns were annotated. The numbers encoded by the symbols "< >" indicate the *synset* ID of the noun concept, and the numbers in parenthesis codify the frequency of each concept/*synset* in the cluster. The nouns "manifestante" (Portuguese) and "protester" (English), for instance, express the same concept (i.e., "*a person who dissents from some established policy*"), which is codified by the ID <10002760> ({dissenter, dissident, protester, objector, contestant}). The frequency of the concept in the cluster is 16, and this value is associated to every occurrence of a noun that lexicalizes the referred concept. Once the measure is specified for all concepts, sentences are ranked according to the sum of the frequency of their constitutive concepts. The score of the sentence in Portuguese is 51 and it occupies the first position of the rank, while the sentence in English, with a score=28, occupies the 12th position. Being composed of the most frequent concepts, the top-ranked sentences are descriptive of the main topic of the cluster. Thus, highly ranked sentences are very suitable for the summary.

Table 1. Example of sentence scoring and ranking based on concept frequency measure.

Sentences	**Score**	**Rank**
Um grupo<31264>(6) de **manifestantes<10002760>(16)** conseguiu furar o bloqueio<8376948>(2) da Polícia Militar e chegar ao estádio<4295881>(14) Mané Guarrincha neste sábado<15164570>(4), horas<15227846>(2) antes do jogo<7470671(5) de abertura<7452699>(2) da Copa das Confederações.[10]	51	1^{rst}
Brazil's <9379111>(4) opening<74522699>(2) Confederations Cup match<7470671>(5) was affected by **protesters<10002760>(16)** that left 39 people<7942152>(1) injured.	28	12th

[10] "*A group of protesters broke through the military police line and got to the Mané Guarrincha stadium on Saturday, hours before the Confederations Cup's opening match*."

Given the rank, one of our selection strategies, called CF (*concept frequency*), performs the sentence selection exclusively based on the rank, independently of the source language. Specifically, CF starts selecting the best-ranked sentence to compose the summary (in Portuguese), and, if it happens that this sentence (as any other along the content selection) is in English, it is automatically translated to Portuguese. After the first selection, if the compression rate is not reached, the 2nd best-ranked sentence is a candidate to compose the summary. Since the input is a multi-document cluster, checking for redundancy between the candidate sentence and the previously selected one is necessary, because the summary should reflect the diverse topics of the cluster without redundancy. In order to avoid redundancy, we assume a threshold (i.e., a pre-established limit) that the new selected sentence may have in relation to any of the previously selected sentences. Thus, if this limit is reached, the new sentence is considered redundant and ignored, and the summarization process goes to the next candidate sentence; otherwise, the sentence is included in the summary. In case of ties (i.e., sentences with the same relevance score in the rank) between a machine-translated sentence and an original sentence in Portuguese, the CF method picks the shortest one. This whole process is repeated until the desired summary length is achieved. The CF method was proposed under the assumption: the application of a late-translation strategy, in which the MT is only used to translate the selected sentences in English to Portuguese, minimizes the problems in the summaries that are caused by the full MT of the source texts.

The other strategy, called CFUL (*concept frequency + user language*), is driven by the user's language. It exclusively selects the top-ranked sentences from the text written in Portuguese language to compose the summary, also avoiding redundancy. In case of ties between two original sentences in Portuguese, the CFUL method uses the same criterion applied by CF, i.e., picking the shortest one. Consequently, the final summary only contains sentences in such preferred language. This approach relies on the assumption that a summary built exclusively with original sentences in Portuguese reflects the most relevant information of the cluster, since the concepts that occur in the English text are also taken into account for sentence ranking.

Finally, in the synthesis stage, the methods produce the extracts, as the vast majority of the works in automatic summarization today. So, the CF and CFUL methods simply juxtapose the sentences selected from the rank, ordering them according to their position in their corresponding source texts.

4 The CM2News *corpus*

For testing the MMDS methods, we have used the CM2News *corpus* [10]. It has 40 original news texts (in a total of 19,984 words) grouped by topic in 20 clusters. Each cluster is composed of 2 news texts, 1 in English and 1 in (Brazilian) Portuguese, both on the same topic, and 1 human summary in Portuguese (abstract[11]), which corresponds to the 30% of the size of the biggest text of the cluster (i.e., 70% compression rate). The clusters cover different domains: world, politics, health, science, entertainment, and environment. Since the *corpus* was not semantically annotated, we have carried out the annotation of the nominal concepts as follows.

[11] Summaries that contain some degree of paraphrase of the input.

Each cluster was semi-automatically annotated by groups of 2 or 3 experts with the support of an easy-to-use annotation tool adapted for this task. For each new cluster under analysis, the groups were mixed, trying to avoid any annotation bias. The task was carried out by 12 computational linguists in daily meetings of 90 or 120 minutes, during 15 consecutive days. The annotation training took 1 day.

The mentioned annotation tool/editor is called MulSen[12] (*Multilingual Sense Estimator*), an adaptation of NASP[13] [19]. Given a cluster, the editor firstly performs an automatic pre-processing task over the source texts, which is the morphosyntactic annotation. To address this task, it incorporates two part-of-speech (POS) taggers, one for each language [16] [19]. Once the nouns are tagged, MulSen translates the nouns from the text in Portuguese to English, which is necessary considering that WN.Pr is our conceptual repository. The translation is done using the online bilingual dictionary WordReference®[14,15]. When the text in English is under annotation, MulSen just skips the MT stage. Finally, the editor suggests the *synsets* that better represent the concepts. The suggestions result from the application of *word sense desambiguation* (WSD) algorithms for English and Portuguese languages [19]. Thus, the WSD methods generate a pre-annotation of the nouns, which should be validated (or not) by the experts to complete the process. The tool allows the manual revision of the POS tagging, MT, and conceptual annotation (or *synset* selection) outputs.

To annotate the nouns, the experts have followed 4 generic and 4 specific rules.

The 4 generic instructions are: (i) firstly annotate the text in English of a cluster, since its vocabulary can provide appropriate translations for the annotation of the nouns in Portuguese, (ii) annotate the POS silence, i.e., nouns that were not automatically detected, (iii) ignore the POS noise, i.e., words that were wrongly annotated as nouns, and (vi) annotate all the different occurrences of a concept (i.e, synonyms and equivalences) in the cluster with the same (and more adequate) *synset*.

The first specific rule establishes the annotation of the multiword expressions head with a *synset* that codify the concept of the whole expression, since the taggers do not detect multiword expressions. For instance, in the Portuguese sentence "*Um dos manifestantes levou gás de pimenta no rosto*" ("*One of the protesters was hit in the face by pepper spray*"), "gás" was annotated with the *synset* {pepper spray} ("*a nonlethal aerosol spray made with the pepper derivative oleoresin capiscum*") because it is part of the expression "gás de pimenta", which is not detected by the taggers. The second rule determines that the annotators should analyze all the possible translations provided by MulSen as well as their respective *synsets* before completing the process. It is important because the adequate translation may not be the first in the list of alternatives provided by the editor. The third rule is for the cases where translations have to be manually inserted in the editor, because the editor could not (i) find any translation in WordReference or (ii) provide an appropriate one among the suggested list. For inserting a translation, the third rule establishes that the annotators should test all the possible equivalences found in others resources before finally adding the more appropriate in MulSen. The forth rule determine that, if there is not a

[12] http://www.icmc.usp.br/pessoas/taspardo/sucinto/resources.html

[13] We thank Fernando A. A. Nóbrega for helping adapting the tool.

[14] http://www.wordreference.com/

[15] We have excluded others resources (e.g., *Google Translation*) because of use/license limitations.

proper *synset* to codify a concept of a noun, it should be selected a more generic one. This means that, if any of the *synsets* activated by the chosen translation is not adequate, the annotators should look for a satisfactory hypernym *synset*.

In the next section we report our experiments and the results that we obtained.

5 Evaluation and Results

The evaluation was carried out over the CM2News *corpus*. For each cluster, we manually built 1 extract based on CF and 1 based on CFUL. We have applied a 70% compression rate (in relation to the longest text), and *word overlap* to avoid redundancy, such as [4]. Regarding the CF method, we used Microsoft Bing® for translating the summary sentences in English to Portuguese. The strategies were analyzed based on the informativeness and linguistic quality of the extracts. Our methods were compared to the best *baseline* of Tosta et al [4], i.e., *sentence position* method with redundancy treatment.

To analyze the quality of the extracts, we used the 5 criteria of DUC [15]. The criteria were manually analyzed by 15 computational linguists. The 20 clusters of CM2News were divided in 5 groups of 4 clusters. Each group was composed of the summaries generated by CF and CFUL, totalizing 8 extracts. The analysis of each group was performed by 3 different judges. Given a summary, the judges scored each of the 5 textual properties through an online form. For all properties, judges had a scale from 1 to 5 points, being 1=very poor, 2=poor, 3=barely acceptable, 4=good, and 5=very good. Looking to the average values (Table 2), one may see that the CFUL method outperforms the CF strategy and the *baseline* in all the criteria, indicating that the content selection based on the combination of conceptual knowledge and user's language is better at dealing with textually factors in the summaries. This performance is not surprising, since the sentences come exclusively from one of the source texts. It is interesting to comment that this simulates a usual behavior in human summarization, which is choosing a source as basis for MDS [5]. One may also see that CF outperforms the *baseline* in 4 (except for "structure and coherence") from the 5 criteria, which confirms the hypothesis that the late-translation approach produces fewer textual problems. Even for structure and coherence, the *baseline* performance was not significantly higher than CF (2,8 and 2,6, respectively).

Regarding informativeness evaluation, we used the traditional automatic ROUGE measure [13], which is mandatory in the area. Particularly, we used ROUGE-1, which measures the amount of unigram overlap between reference summaries and automatic summaries, and ROUGE-2, which measures the amount of bigram overlap. We have chosen these two measures because unigrams and bigrams are the most frequent *n*-grams in language. The average results for ROUGE-1 and ROUGE-2 in terms of *recall*, *precision* and *f-measure* are shows in Table 3. Basically, *recall* computes the amount of common n-grams in relation to the number of n-grams in the reference summaries, *precision* computes the number of common n-grams in relation to the n-grams in the automatic summary, and the *f-measure* is the harmonic mean of the previous 2 measures, being a unique indicator of the system performance. According to Table 3, one may see that CFUL method outperforms the CF strategy and the *baseline* in the 2 measures. To statistically determine if the differences in performance were significant, we have performed a Wilcoxon signed-rank test with 95%

confidence, which confirmed the difference. These results indicate that our hypothesis – that the summaries built exclusively with original sentences in Portuguese reflects the most relevant information of the cluster, since the concepts of the English text are also taken into account for sentence ranking – hold. It is important to say, however, that such results are only indicative of what we may expect from the CF and CFUL methods, since our *corpus* for quality and ROUGE evaluation was small (20 clusters). For a more reliable result, we would need to apply the methods for a bigger *corpus*, which remains as future work.

Table 2. Linguistic quality evaluation of summaries with DUC criteria.

Criteria	CF	CFUL	Baseline
Grammaticality	3,5	**4,3**	3
Non-redundancy	3,4	**4,3**	3
Referential clarity	3,3	**3,7**	3,2
Focus	3,5	**4,1**	4
Structure and coherence	2,6	**3,4**	2,8

Table 3. Informativeness evaluation of summaries with ROUGE.

Method	Avg. ROUGE-1			Avg. ROUGE-2		
	Recall	Precision	F-measure	Recall	Precision	F-measure
CF	0,355	0,328	0,341	0,155	0,144	0,149
CFUL	**0,373**	**0,369**	**0,371**	**0,174**	**0,175**	**0,174**
Baseline	0,313	0,271	0,285	0,038	0,032	0,034

Acknowledgments. We thank the Brazilian National Council for Scientific and Technological Development (CNPq) (#483231/2012-6), the State of São Paulo Research Foundation (FAPESP) (#2012/13246-5, #2015/17841-3), and Coordination for the Improvement of Higher Level or Education Personnel (CAPES) for the financial support.

6 Final remarks

As far as we know, this is the first investigation on deep methods for MMDS involving Portuguese as the user's language. We showed that concept-based methods tend to produce extracts with better informativeness and linguistic quality level than a *sentence position baseline*. Other contributions of this work are the annotation of a corpus with noun concepts and the adaptation of an annotation tool, which are freely available for use. However, it is important to recognize that the methods suffer from well-known drawbacks, which are the dependence of linguistic knowledge and the effective lack of scalability. In this line, it is possible to consider to use, for instance, automatic tools for WSD. For Portuguese, one might consider the use of the general purpose methods of Nóbrega and Pardo [20]; for English, several tools are available, as the one of Pedersen and Kolhatkar [21]. The overall performance of the MMDS methods will certainly drop, but their benefits would still be valuable. Some other future works include (i) exploring the construction of automatic and reference summaries with different compression rates, under the assumption that smaller extracts have fewer language problems, and (ii) investigating the impact on redundancy treatment of using a *concept overlap* strategy for redundancy identification (instead of word overlap, as we have done in this paper).

References

1. Evans, D.K., Klavans, J.L., Mckeown, K.R.: Columbia NewsBlaster: multilingual news summarization on the web. In: North American Chapter of The Association for Computacional Linguistics: Human Language Technologies, p.1-4. Boston (2004)
2. Roark, B., Fisher, S.: OGI OHSU baseline multilingual multi-document summarization system. In: Multilingual Summarization Evaluation (MSE). Michigan, USA (2005)
3. Evans, D.K., Klavans, J.L. Mckeown, K.R.: Similarity-based multilingual multi-document summarization. Technical Report CUCS-014-05. New York: Columbia University (2005)
4. Tosta, F.E.S., Di-Felippo, A., Pardo, T.A.S.: Estudo de métodos clássicos de sumarização automática no cenário multidocumento multilíngue. In: 4th Workshop de IC em Tecnologia da Informação e da Linguagem Humana. p. 34-36. Fortaleza, Brazil (2013)
5. Mani, I.: Automatic Summarization. John Benjamins Publishing Co., Amsterdam. (2004)
6. Schiffman, B., Nenkova, A., Mckeown, A.: Experiments in multi-document summarization. In: 2nd International Conference on HLT Research. p.52-8. San Francisco (2002)
7. S. Ye, T.-S. Chua, M.-Y. Kan, and L. Qiu.: Document concept lattice for text understanding and summarization. Information Processing and Management, v.43, n.6, pp.1643–62 (2007)
8. Wu, C-W., Liu, C-L.: Ontology-based Text Summarization for Business News Articles.Computers and Their Applications, v. 2003, p. 389-392, 2003.
9. Hennig, L., Umbrath, W., Wetzker, R.: An ontology-based approach to text summarization. In: 3th Workshop On Natural Language Processing And Ontology Engineering (NLPOE), p. 291-294. Toronto, Canada (2008)
10. Tosta, F.E.S.: Aplicação de conhecimento léxico-conceitual na Sumarização Multidocumento Multilíngue. 2013. Dissertação (Mestrado em Linguística)–Departamento de Letras, Universidade Federal de São Carlos (2014)
11. Fellbaum, C. (Ed.): Wordnet: an electronic lexical database (Language, speech and communication). Massachusetts: MIT Press (1998)
12. Hatzivassiloglou, J.L., Klavans J.L., Holcombe, M.: Simfinder: a flexible clustering tool for summarization. In: NAACL Automatic Summarization Workshop. p.9. Pittsburgh (2001)
13. Lin, C-Y.: ROUGE: a Package for Automatic Evaluation of Summaries. In: Workshop on Text Summarization Branches Out. (2004)
14. Kumar, Y.J., Salim, N.: Automatic Multi-Document Summarization Approaches. Journal of Computer Science 8 (1): 133-140. ISSN 1549-3636 (2012)
15. Dang, H.T.: Overview of DUC 2005. In: Document Understanding Conference (2005)
16. Ratnaparkhi, A.: A maximum entropy model for part-of-speech tagging. In: Conference on Empirical Methods in Natural Language Processing. Philadelphia, PA. (1996)
17. Sparck-Jones, K.: Automatic summarizing: factors and directions. In: Mani, I.; Maybury, M.T. (Eds). Advances in automatic text summarization. MA: MIT Press, p.1-14 (1999)
18. Schmid, H.: Probablistic part-of-speech tagging using decision trees. In: International Conference on New Methods in Language Processing, p. 44-9. Manchester, UK (1994)
19. Nóbrega, F.A.A.: Desambiguação lexical de sentidos para o português por meio de uma abordagem multilíngue mono e multidocumento. Dissertação (Mestrado em Ciências de Computação e Matemática Computacional) - ICMC, USP, São Carlos (2013)
20. Nóbrega, F.A.A., Pardo, T.A.S.: General Purpose Word Sense Disambiguation Methods for Nouns in Portuguese. In: PROPOR 2014 PhD and MSc/MA Dissertation Contest / 11st International Conference on Computational Processing of Portuguese, p. 94-101. São Carlos/SP, Brazil (2014)
21. Pedersen, T., Kolhatkar, V.: WordNet::SenseRelate::AllWords - A Broad Coverage Word Sense Tagger that Maximizes Semantic Relatedness. In: North American Chapter of the Association for Computational Linguistics / Human Language Technologies Conference, p. 17-20. Boulder, Colorado (2009).