

5th International Conference on Corpus Linguistics (CILC2013)

Manual Typification of Source Texts and Multi-document Summaries Alignments

Renata T. Camargo^{a*}, Verônica Agostini^b, Ariani Di Felippo^a, Thiago A. S. Pardo^b

^aFederal University of São Carlos, Rodovia Washington Luís, km 235 - SP 310, São Carlos, 13565-905, Brazil

^bUniversity of São Paulo, Avenida Trabalhador São-carlense, 400, São Carlos, 13566-590, Brazil

Abstract

The Multi-document Summarization (MDS) has been focused in Natural Language Processing (NLP) and its aim is to produce automatic summaries from a collection of texts that deal with the same subject (Mani, 2001). The alignment of human-written abstracts to their source documents makes explicit the correspondences that exist in such documents/abstract pairs and create a potentially rich data source to create of rules and models to support more linguistically motivated MDS methods. In this paper we describe the typification of such alignments in the CSTNews *corpus*. This work is part of two larger projects called Sucinto and Sustento, and it supports MSD researches of Brazilian Portuguese language. Specifically, the typification process consisted of assigning labels to the alignment between a summary sentence and its corresponding source sentence which codify formal and content aspects of the alignment. In order to present this work, we outline the alignment, and detail the typification process, the results of our work and some conclusions.

© 2013 The Authors. Published by Elsevier Ltd.
Selection and peer-review under responsibility of CILC2013.

Keywords: typification; alignment; summarization

1. Introduction

With the continuing growth of available information on the web, strategies are required to efficiently present textual information. Accordingly, Multi-document Summarization (MDS) has been focused in Natural Language

* Corresponding author. Tel.: +55-16-3415-2567
E-mail address: renatatironi@hotmail.com

Processing (NLP), since the aim in MDS is to automatically produce summaries from multiple texts on the same topic (Mani, 2001).

While successful in some cases, the MSD methods are not able to adequately capture the large set of linguistic devices utilized by humans when they produce multi-document summaries. We believe that future progress in MSD will be driven by the development of more sophisticated, linguistically informed methods.

In this scenario, the alignment of a human-written summary and its source texts is a very important task, and it is seen as a type of *corpus* annotation. In the MSD, this kind of alignment allows the analysis of the human multi-document summarization process, since it makes explicit the correspondences that exist in such summary-documents pairs and create a potentially rich data source from which rules and models may be learned to support more linguistically motivated MDS methods.

Despite the relevance, the alignments are more linguistically informative when their characteristics are codified explicitly. So, in this paper, we describe the typification of the alignments in CSTNews, a *corpus* of Brazilian Portuguese (BP) language created to support MDS researches (Cardoso et al., 2011). Our typification process was completed manual and consisted of assigning labels to the alignment between a summary sentence and its corresponding source sentence. The labels or tags codify formal and content aspects of the alignment.

Moreover, the typification was performed in the context of two NLP projects, Sustento (FAPESP 2012/13246-5/CNPq 483231/2012-6) and Sucinto (FAPESP 2012/03071-3), both under development at NILC (Interinstitutional Center for Computational Linguistics), one of the biggest research groups on NLP in Brazil. The Sustento aims at generating linguistic knowledge for MDS, and the Sucinto project aims at investigating MDS strategies for providing a more feasible and intelligent access to on-line information provided by news agencies. Then, the typification of the alignment is a paradigmatic example of the collaborative work between linguists and computational scientists which has been developed at NILC since its foundation in the early 1990s.

The remainder of the paper is structured as follows: Section 2 details related works on alignment and typification process; Section 3 we present the CSTNews *corpus*; Section 4 sketches out the previous alignment of the *corpus* and details the process of typification; Section 5 shows the results; and the paper finishes with some final remarks in Section 6.

2. Related Works

The alignment consists of relating textual segments (words, sentences, paragraphs, or even documents) usually based on the content of the textual segments. The related segments must have something in common, and this depends on the application in which the alignment will be used. Several applications developed in the NLP research field may make use of the alignments: (i) Machine Translation (e.g. Gale & Church, 1991, 1993; Yamada & Knight, 2001; Caseli, 2003), Question Answering (e.g. Soricut & Brill, 2004), Text Simplification (e.g. Specia, 2010), Automatic Summarization (e.g. Marcu, 1999; Hirao, Suzuki, Isozaki, & Maeda, 2004), among others. Considering any of these applications, the alignment provides knowledge about the task to be automated and the process may be manual or automatic.

Automatic text alignment has a long history and many algorithms have been suggested for aligning multilingual parallel *corpus* (i.e., a collection of texts, each of which is translated into one or more other languages than the original) (Wu, 2000). Specifically, these algorithms have been used to map translated texts across distinct languages. Manual alignments, on the other hand, are usually accomplished in order to produce a “gold standard”. Gold standard alignments are an important resource for developing and evaluating automatic alignment methods.

Although the manual alignment may be a time-consuming process, many authors have been used it. For example, in Marcu, (1999) and Jing & McKeown, (1999), ten texts were manually aligned in sentence level by 14 judges based on content overlap. In Hatzivassiloglou, Klavans, & Eskin, (1999), text units (usually paragraphs) were aligned by two human annotators when these units focused on the same concept, actor, object or action. In Barzilay & Elhadad (2003), two judges aligned a pair of sentences if there was at least one clause in common that expressed the same information; the cases of disagreement were solved by a third judge.

The alignment typification, in special, may be defined as the process in which the alignments are classified according to some criteria. To express/codify the types of the alignments, labels or tags are commonly used. Regarding typification, two works of the literature illustrate different classifications. In order to develop an

automatic alignment tool, gold standard alignments were accomplished by two annotators in Daumé III & Marcu (2004, 2005). In this work, annotators were asked to perform word-to-word and phrase-to-phrase between abstracts and documents and to classify or typify each alignment as either “possible” or “sure”. In another work, was carried out by Clough, Gaizauskas, Piao, & Wilks, (2001), the authors tried to figure out the types of rewrite operations that may occur between moving from news agency sources to the newspaper version. At the document level, trained journalists should assign one of the following labels: (i) wholly-derived, (ii) partially-derived and (iii) non-derived. At the word or sentence level, they should judge among three labels: (i) verbatim, (ii) rewrite and (iii) new. One may see an example of classification at the sentence level in Figure 1.

Original (news agency)	A drink-driver who ran into the Queen Mother's official Daimler was fined £700 and banned from driving for two years.
Rewrite (tabloid)	A drunk driver who ploughed into the Queen Mother's limo was fined £700 and banned for two years yesterday.

Figure 1. Example of alignment classification (Clough, Gaizauskas, Piao, & Wilks, 2001).

Base on the related works and motivated by the necessity of linguistic knowledge on human multi-document summarization, we performed the typification of the previous alignments between documents and summaries of the CSNews corpus, which is described in the next section.

3. The CSTNews Corpus

The CSTNews (Cardoso, Maziero, Castro Jorge, Seno, Di Felippo, Rino, Nunes, & Pardo, 2011) has been widely used on researches in MDS for BP and, as far as we know, it is the only available MDS corpus for BP language.

The CSTNews is composed of 50 clusters of news texts. Specifically, each cluster contains: (i) 2 or 3 source texts on the same topic from different online newspapers (in a total of 140 documents), (ii) 1 multi-document human abstract; (iii) 1 multi-document automatic summary, and annotated version of the source texts and multi-document summaries in different linguistic levels (e.g., morphosyntactic and rhetorical/discursive levels), and according different linguistic theories or models.

The texts were manually collected from the major Brazilian online newspapers: *Folha de São Paulo*, *Estadão*, *Jornal do Brasil*, *O Globo*, and *Gazeta do Povo*. The manual collection was carried out for about 60 days, from August to September of 2007. The clusters have on average 42 sentences (10 sentences to 89 sentences) and the multi-document human summaries have on average 7 sentences (3-14). Moreover, the clusters are labeled by the “sections” of the newspapers. Thus, the corpus consists of clusters from the following categories: sports (10 clusters), world (14 clusters), money (1 cluster), politics (10 clusters), science (1 cluster), and daily news (14 clusters). The distribution of the categories may be seen in Figure 2. The unbalanced text distribution among the categories of the CSTNews corpus is resulted from the difficulty of compiling texts about the same news published by more than one online newspaper.

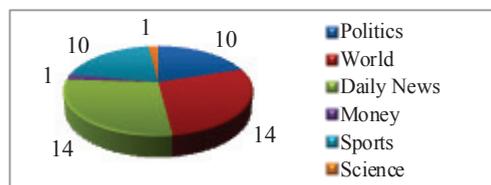


Figure 2. Distribution of sections in the corpus

Regarding the multi-document human summaries, we emphasize that they are *abstract*, i.e., summaries at least some of whose material is not present in the input. In other words, they were manually produced with rewriting operations of the source material. In addition, their production was guided by a compression rate of 70%. So, the

summaries contain a maximum of 30% of the number of words from the largest source text of the *cluster*. In the next section, we outline the previous summary-to-documents alignments whose typification is focused here.

4. The alignment process

In this section, we report the alignment of the multi-document human summaries and their source texts in the CSTNews *corpus*. First, we emphasize that the alignment was performed in the summary-to-document direction. Second, to connect a summary and its source texts of a cluster, completely manual alignments were created at the sentence level, since sentences are self-contained textual segments.

The alignment was preceded by a training phase in order to understand the phenomenon, identify cases of disagreement, and specify alignment rules, especially to address the disagreements. As the result, an annotation manual was written to be used as guideline during the annotation. The annotation manual contains 8 alignment rules, which may be seen in Figure 3.

Rules	Number	Definition
General	1	Align on content overlap
	2	Align based on main information overlap
	3	Align based on secondary information overlap
	4	Align all the overlapping content
Specific	5	Align even when there is contradictory numerical data
	6	Align even when there are different levels of generalization
	7	Align even when there are different levels of assertiveness
	8	Don't align when one means "whole" and the other "part"

Figure 3. Alignment rules

The rules were divided into generic and specific ones. To illustrate the annotation procedures, we briefly explain the generic rules. The rule 1 specifies that a summary sentence must be aligned to a document sentence based on the content overlap, not only considering the word overlap between them. The rule 2 states a central instruction to the annotators: the alignment must be firstly based on the main information which is expressed by the sentences, i.e., a summary sentence must be alignment to a document sentence if they express similar topic sentence. If this was not possible, the rule 3 establishes that two summary-document sentences may be aligned based on secondary information overlap. The rule 4 states an important alignment requirement too: a summary sentence must be connected to all sentences of the different cluster documents with which it shares content. According to the rule 4, the summary-documents alignments codify one-to-many relationships.

After training, the summary-to-documents alignments were performed by two computational linguistic annotators in daily meetings from 1 to 2 hours during the period of approximately 60 days. The maximum duration of the sessions was to 2 hours because the task is tedious, and the performance of the annotators could be compromised after this period.

Figure 4 and Figure 5 provide examples of alignments between a summary sentence (SS) and a document source sentence (DS) based on rules 1 and 6, respectively.

In Figure 4, we see that SS was aligned to the DS because both sentences have the same meaning or express the same topic. In Figure 5, we see that SS was aligned to the DS because both convey the same topic, i.e., a "traffic jam", although SS provides specific information about it, which is (i) the extension of the traffic jam in km and (ii) the time of the event. Finally, we point out that some alignments required domain specific knowledge, especially those between summaries and sources texts of the "politics" category. Following, we will describe the results of manual alignments of multi-document summary-to-documents in the CSTNews *corpus*.

Summary Sentence (SS)
<i>17 pessoas morreram após a queda de um avião na República Democrática do Congo.</i> (17 people died after a plane crash in the Democratic Republic of Congo)
Document Sentence (DS)
<i>Um acidente aéreo na localidade de Bukavu, no leste da República Democrática do Congo (RDC), matou 17 pessoas na quinta-feira à tarde, informou nesta sexta-feira um porta-voz das Nações Unidas.</i> (A crash in the town of Bukavu in the eastern Democratic Republic of Congo (DRC), killed 17 people on Thursday afternoon, said on Friday a spokesman of the United Nations.)

Figure 4. Example of an alignment involving Rule 1

Summary Sentence (SS)
<i>A Companhia de Engenharia de Tráfego (CET) anunciou que o índice de congestionamento era de 54 quilômetros às 8h, 113 km às 9h e 110 km meia hora depois, valores bem acima das médias para os horários, que eram de 36, 82 e 76 quilômetros respectivamente, mas não havia registro de acidentes graves, apesar de haver feridos.</i> (The Traffic Engineering Company (CET) announced that the extent of traffic jam was 54 km at 8am, 9am 113 km and 110 km half an hour later, values much above the averages for the times, which were 36, 82 and 76 km respectively, but there were no serious accidents, although there were injured people.)
Document Sentence (DS)
<i>Com o asfalto molhado, o trânsito ficou mais lento e o congestionamento ficou o dobro da média.</i> (With the wet pavement, the traffic got slowly and the traffic jam got the double of the average.)

Figure 5. Example of an alignment involving Rule 6

4.1. The alignment results

Approximately 78% of the summary sentences were aligned to more than one sentence of the source texts. This result was expected, since a multi-document summary could be potentially connected to 2 or 3 related source texts of its cluster. All the alignment types is given in Table 1.

Table 1. Alignment types in the corpus

Alignment Types												
1-0	1-1	1-2	1-3	1-4	1-5	1-6	1-7	1-8	1-9	1-	1-	1-
										10	11	12
2	71	91	72	33	37	13	6	6	1	1	2	1
Amount of Alignments												

According to the annotation results in Table 2, we may see that: (i) 2 summary sentences were not aligned; these cases of one-to-zero alignment (1-0) mean that the content in summary was inferred from the source documents of the cluster; (ii) 71 summary sentences were aligned to only one sentence of the source texts (1-1); (iii) 91 summary sentences were aligned to only 2 sentences of the source texts (1-2), and so on.

From the 2067 sentences in the source texts, 877 (42.43%) were aligned to some summary sentence, but it does not mean that the sentences were aligned only once. A sentence of a summary may be aligned to more than one sentence of the source text, and the sentences of the source texts may be redundant or even identical. From 336 summary sentences, 334 were aligned (99.4%) to some source sentence.

To ensure the reliability of the annotation, we computed the agreement between the annotators by using the kappa (k) measure or statistic (Cohen, 1960; Carletta, 1996). It is generally thought to be a more robust measure than simple percent agreement calculation since k takes into account the agreement occurring by chance. This measure is applied in many NLP tasks. The value of the measure ranges from 0 to 1, where k=0 indicates no

agreement between annotators and $k=1$ indicates complete agreement between them. The annotator agreement was computed once a week. To do so, the annotators individually aligned the same cluster and compared the results of their alignment to verify the agreement. Five clusters were used in this task and the kappa result was 0.831. This value indicates that the task, although subjective, it was well-defined.

The alignment presented here is available in an XML format (Extensible Markup Language), because it is a format that is widely used in *corpus* annotation. In the next section, we describe the typification of the alignments.

5. The typification process

This task was performed by two computational linguists over a period of 2 months with daily sections of about 1 hour. The process consisted of assigning labels or tags to each pair of aligned sentences of the CSTNews *corpus*. The pairs of aligned sentences were classified according to the word-form and content overlap between them, since the goal of the typification was to verify how much of the superficial linguistic material and content of the source texts were presented in the summaries. So, the labels were divided into two types: form and content.

Given a SD-SS pair, the annotators could indicate the overlap between them based on common word-forms by the labels: (i) identical, when the two sentences are the same, (ii) partial, when they are similar, i.e., they have a few or several word-forms in common, and (iii) different, when they have few words in common.

To indicate the content overlap, the labels are: (i) specification, when the SS contains some specific information related to the original content of the aligned SD; (ii) generalization, when the SS generalizes the content of the SD; (iii) contradiction, when SS and SD present some contradictory information; (iv) inference, when the SS expresses information that was inferred from the corresponding SD; (v) neutral, when the SS contains some information that results from an unknown process of the original SD content, and (vi) other, when the annotators do not agree with the prior alignment, since the alignment is a subjective task. The generalization and specification, as a matter of fact, are two import cross-document fusion operations commonly used by humans to condensate the content of source texts and produce a multi-document summary (Mani 2001).

Annotators also classified the alignments based on the occurrence of onomastic elements, i.e., proper nouns. The onomastic aspects were divided in two types: (i) toponomastics, when names of places occurred in the aligned sentences, and (ii) anthroponomastics, when names of persons occurred in the aligned sentences. Figure 6 shows all the labels used to characterize the alignments.

Form types	Identical Partial Differ
Content types	Specification Generalization Contradiction Inference Neutral Other
Onomastics	Toponomastics Anthroponomastics

Figure 6. Labels used to characterize the alignments

Based on the set of labels in Figure 6, the typification was performed as follows. First of all, we chose one tag for the alignment concerning its form (i.e., identical, partial or different), considering the whole sentences of the pair to decide between the tags. Thus, we compared a summary sentence with the document sentence of the pair to decide if they were exactly the same, similar or completely different. For the content tags, we considered n-grams to decide what to choose, that is why more than one tag was allowed. Finally, if there was the presence of anthroponyms or toponyms, we put either the “toponomastics” tag or “athroponomastics” one indicating the existence of such relation.

In the example of Figure 7, one may notice a partial alignment, since the two sentences have some words in common but are not identical. Moreover, we identified one content transformation between the sentences. In this case, there was a generalization in evidence considering the content overlapping pair: the n-gram “many states” is more general than “Amazonas, Distrito Federal, Mato Grosso, Acre and Rondônia”, which are names of some Brazilian states. Furthermore, we also identified the presence of some names of places (states), which was labeled with the tag “toponomastics”. Therefore, the alignment in Figure 7 received the tags: (i) partial, (ii) neutral, (iii) generalization and (iv) toponomastics.

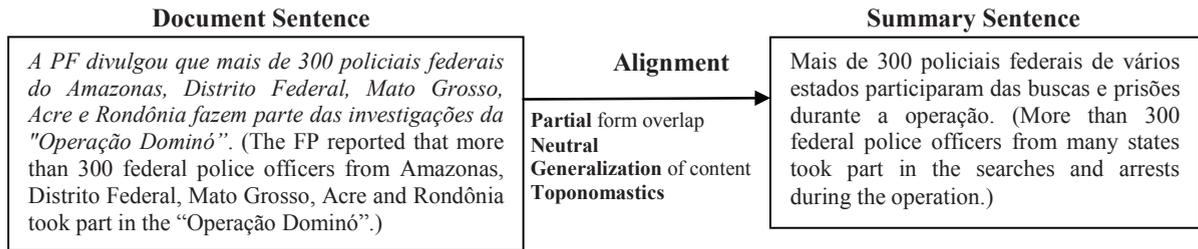


Figure 7. Example 1 of typification

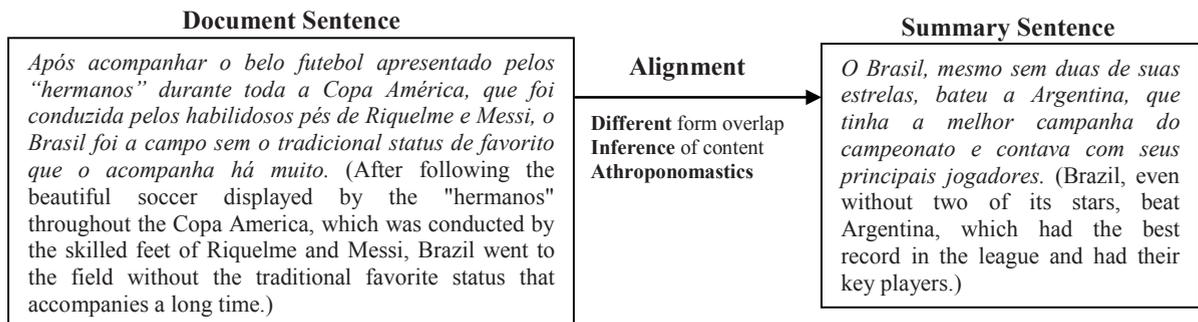


Figure 8. Example 2 of typification

In the example showed in Figure 8, the two sentences received the form type “different”, because the two don’t have words in common. The fact that the players Messi and Riquelme were the main players was inferred by the human who summarized. Besides, we identified the presence of the names of the players, which we labeled with the athroponomastics tag as well. Therefore, the alignment in Figure 8 received the tags: (i) different, (ii) inference and (ii) athroponomastics.

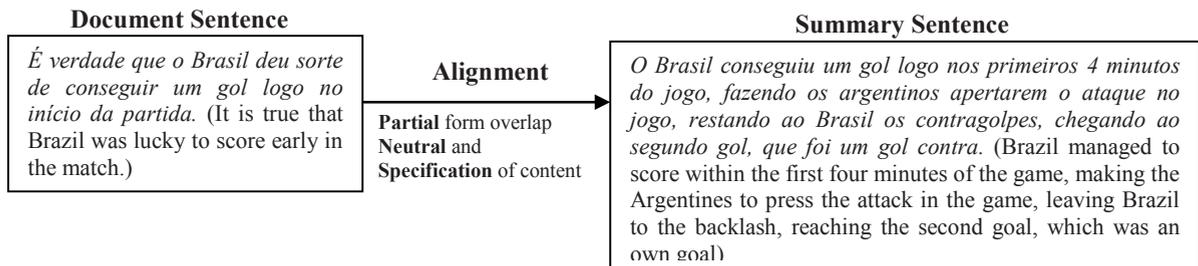


Figure 9. Example 3 of typification

In Figure 9, the two sentences received the form type partial, because the two have some words in common, received the neutral type because the two convey the same information (the fact that Brazil scored, and for that it does not matter the exact point in the game), and received the specification type because “4 minutes” is more specific than “beginning of the game”. Therefore, the alignment in Figure 9 received the tags: (i) partial, (ii) neutral and (ii) specification.

5.1. The typification results

As a result, from a total of 1007 alignments, we identified 867 partial alignments (86%), 58 identical alignments (5.7%) and 82 different alignments (8.1%). In Figure 10, one may notice the distribution of the form categories in the *corpus*.

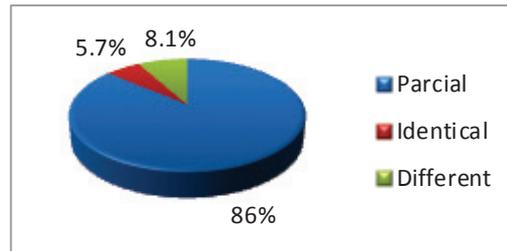


Figure 10. Distribution of the form categories in the *corpus*

Giving the nature of the sentence pairs that were aligned (sentences from texts that talk about the same content), it was expected that the sentences would be at least partial. Besides that, the partial type corresponds to two sentences that are not identical or not totally different, i.e., they may have only a few words in common, or have a huge amount of words in common, that is why we have a large number (86%) of partial alignments.

Regarding the content, we identified 949 neutral alignments (94.2%), 37 contradiction alignments (3.6%), 82 generalization alignments (8.1%), 48 specification alignments (4.7%), 33 inference alignments (3.2%) and 6 other alignments (0.5%). As it was expected, taking into account that the alignments are from document sentences to abstract sentences, there are more generalization alignments than specification ones. Further, we emphasize that from the 867 partial alignments 714 were classified as neutral (70.9%), without another content tag.

As in the alignment process, we calculated the kappa measure considering five clusters. We obtained 0.452 of kappa measure regarding all the tags. The kappa result regarding the form types was 0.717 and concerning the content tags was 0.318, because the alignments could receive more than one content tag and it is a very subjective task. We show all the kappa results in Table 2.

Table 2. Kappa results

Agreement Measure	Form Alignments	Content Alignments	Form and Content
Kappa (Carletta, 1996)	0.717	0.318	0.452

It was expected that the agreement considering the content alignments would be not so high, giving the subjectivity of the task, as we said before. Considering the typification task, it was necessary to review it more than one time and it was more difficult than we expected at the beginning.

6. Final Remarks

It is noteworthy that the typification may be used in at least two studies that fit the scenario of MDS. One tries to investigate the human process of multi-document summarization and the other intends to generate an automatic aligner for source texts and multi-document summaries.

It is still remarkable to mention that there were difficulties concerning the definition of the task due to its subjectivity. Also, it was difficult to verify which n-grams were taken into account to perform the summarization task.

In general, we may assume that humans do not conduct many transformations that are covered by our alignment types, since 818 alignments (81.2%) received only the neutral content tag. On the other hand, there are not too much extractive sentences, since only 58 of 1007 alignments (5.7%) were annotated as identical. The high number of neutral alignments is due to the possibility of n-grams overlap and the likelihood of co-occurrences of this tag with other content ones.

We may also infer that humans seem to prefer the generalizations over the specifications, since there are 82 generalizations (8.1%) and 48 specifications (4.7%) in the results. We may explain this difference considering that generalizing a piece of information is a way to remove unnecessary details and reduce content in order to create a summary.

As future work, we may try to investigate and include other types of alignments which may occur in multi-document summaries. Thereby, our typification would be more accurate and our research would be more focused on the human task of summarization.

7. Acknowledgments

The authors are grateful to FAPESP and CAPES.

References

- Agostini, V., Camargo, R. T., Di Felippo, A. & Pardo, T. A. S. (Forthcoming 2013). *Manual Alignment of News Texts and their Multi-document Human Summaries*. Cambridge: Cambridge Scholars Publishing.
- Barzilay, R. & Elhadad, N. (2003). Sentence Alignment for Monolingual Comparable Corpora. In the *Proceedings of the Empirical Methods for Natural Language*, 25-32.
- Cardoso, P. C. F., Maziero, E. G., Castro Jorge, M. L. C., Seno, E. M. R., Di Felippo, A., Rino, L. H. M., Nunes, M. G. V. & Pardo, T. A. S. (2011). CSTNews - A Discourse-Annotated Corpus for Single and Multi-Document Summarization of News Texts in Brazilian Portuguese. In the *Proceedings of the 3rd RST Brazilian Meeting*, 88-105.
- Carletta, J. (1996). Assessing Agreement on Classification Tasks: The Kappa Statistic. *Computational Linguistics*, v. 22, n. 2, 249-254.
- Caseli, H. M. (2003). *Alinhamento sentencial de textos paralelos português-inglês*. School: Instituto de Ciências Matemáticas e de Computação (ICMC), Universidade de São Paulo (USP), Fevereiro, 101.
- Clough, P., Gaizauskas, R., Piao, S. & Wilks, Y. (2001) *METER: MEasuringText Reuse*. Department of Computer Science Research Memorandum CS-01-03, University of Sheffield.
- Cohen, J. (1960). A Coefficient of Agreement for Nominal Scales. *Educational and Psychological Measurement*, v. 20, n. 1, 37-46.
- Daumé III, H. & Marcu, D. (2004). A Phrase-Based HMM Approach to Document/Abstract Alignment. In the *Empirical Methods in Natural Language Processing (EMNLP)*, 8.
- Daumé III, H. & Marcu, D. (2005). Induction of Word and Phrase Alignments for Automatic Document Summarization. *Computational Linguistics*, v. 31, n. 4, 505-530.
- Gale, W. A. & Church, K. W. (1991). A program for aligning sentences in bilingual corpora. In the *Proceedings of the 29th Annual Meeting of the Association for Computational Linguistics (ACL)*, Berkley, pp. 177-184.
- Gale, W. A. & Church, K. W. (1993). A program for aligning sentences in bilingual corpora. *Computational Linguistics*, v. 19, n. 3, 75-102.
- Hatzivassiloglou, V., Klavans, J. L. & Eskin, E. (1999). Detecting Text Similarity over Short Passages: Exploring Linguistic Feature Combinations via Machine Learning. In the *Proceedings of the Empirical Methods for Natural Language Processing*, 203-212.
- Hirao, T., Suzuki, J., Isozaki, H. & Maeda, E. (2004). Dependency-based Sentence Alignment for Multiple Document Summarization. In the *COLING '04 Proceedings of the 20th international conference on Computational Linguistics*, 446-452.
- Jing, H. & McKeown, K. (1999). The Decomposition of Human-Written Summary Sentences. In the *Proceedings of the 22nd Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 129-136.
- Mani, I. (2001). Automatic Summarization. *Natural Language Processing Vol. 3*. Amsterdam/Philadelphia: John Benjamins Publishing Company. Centro de Linguística da Universidade do Porto, Cota: N/35, 285.
- Marcu, D. (1999). The automatic construction of large-scale corpora for summarization research. In the *Proceedings of the 22nd Conference on Research and Development in Information Retrieval*, 137-144.
- Soricut, R. & Brill, E. (2004). Automatic Question Answering: Beyond the Factoid. In the *Proceedings of HLT-NAACL*, 57-64.
- Specia, L. (2010). Translating from Complex to Simplified Sentences. In the *Proceedings of PROPOR*, 30-39.
- Yamada, K. & Knight, K. (2001). A syntax-based statistical translation model. In the *Proceedings of the 39th Annual Meeting of the Association for Computational Linguistics (ACL)*, 523-530.