A Discursive Grid Approach to Model Local Coherence in Multi-document Summaries

Márcio S. Dias
Interinstitutional Center for Computational
Linguistics (NILC)
University of São Paulo, São Carlos/SP,
Brazil
`marciosd@icmc.usp.br`

Thiago A. S. Pardo
Interinstitutional Center for Computational
Linguistics (NILC)
University of São Paulo, São Carlos/SP,
Brazil
`taspardo@icmc.usp.br`

Abstract

Multi-document summarization is a very important area of Natural Language Processing (NLP) nowadays because of the huge amount of data in the web. People want more and more information and this information must be coherently organized and summarized. The main focus of this paper is to deal with the coherence of multi-document summaries. Therefore, a model that uses discursive information to automatically evaluate local coherence in multi-document summaries has been developed. This model obtains 92.69% of accuracy in distinguishing coherent from incoherent summaries, outperforming the state of the art in the area.

1 Introduction

In text generation systems (as summarizers, question-answering systems, etc.), coherence is an essential characteristic in order to produce comprehensible texts. As such, studies and theories on coherence ((Mann and Thompson, 1998), (Grosz et al., 1995)) have supported applications that involve text generation ((Seno, 2005), (Bosma, 2004), (Kibble and Power, 2004)).

According to Mani (2001), Multi-document Summarization (MDS) is the task of automatically producing a unique summary from a set of source texts on the same topic. In MDS, local coherence is as important as informativity. A summary must contain relevant information but also present it in a coherent, readable and understandable way.

Coherence is the possibility of establishing a meaning for the text (Koch and Travaglia, 2002). Coherence supposes that there are relationships among the elements of the text for it to make sense. It also involves aspects that are out of the text, for example, the shared knowledge between the producer (writer) and the receiver (reader/listener) of the text, inferences, intertextuality, intentionality and acceptability, among others (Koch and Travaglia, 2002).

Textual coherence occurs in local and global levels (Dijk and Kintsch, 1983). Local level coherence is presented by the local relationship among the parts of a text, for instance, sentences and shorter sequences. On the other hand, a text presents global coherence when this text links all its elements as a whole. Psycholinguistics consider that local coherence is essential in order to achieve global coherence (Mckoon, 1992).

The main phenomena that affect coherence in multi-document summaries are redundant, complementary and contradictory information (Jorge and Pardo, 2010). These phenomena may occur because the information contained in the summaries possibly come from different sources that narrate the same topic. Thus, a good multi-document summary should a) not contain redundant information, b) properly link and order complementary information, and c) avoid or treat contradictory information.

In this context, we present, in this paper, a discourse-based model for capturing the above properties and distinguishing coherent from incoherent (or less coherent) multi-document summaries. Cross-document Structure Theory (CST) (Radev, 2000) and Rhetorical Structure Theory (RST) (Mann and Thompson, 1998) relations are used to create the discursive model.

RST considers that each text presents an underlying rhetorical structure that allows the recovery of the writer's communicative intention. RST relations are structured in the form of a tree, where Elementary Discourse Units (EDUs) are located in the leaves of this tree. CST, in turn, organizes multiple texts on the same topic

60

and establishes relations among different textual segments.

In particular, this work is based on the following assumptions: (i) there are transition patterns of discursive relations (CST and RST) in locally coherent summaries; (ii) and coherent summaries show certain distinct intra- and inter-discursive relation organization (Lin et al., 2011), (Castro Jorge et al., 2014), (Feng et al., 2014). The model we propose aims at incorporating such issues, learning summary discourse organization preferences from corpus.

This paper is organized as follows: in Section 2, it is presented an overview of the most relevant researches related to local coherence; Section 3 details the proposed approach in this paper; Section 4 shows the experimental setup and the obtained results; finally, Section 5 presents some final remarks.

2 Related Work

Foltz et al. (1998) used Latent Semantic Analysis (LSA) (Landauer and Dumais, 1997) to compute a coherence value for texts. LSA produces a vector for each word or sentence, so that the similarity between two words or two sentences may be measured by their cosine (Salton, 1988). The coherence value of a text may be obtained by the cosine measures for all pairs of adjacent sentences. With this statistical approach, the authors obtained 81% and 87.3% of accuracy applied to the earthquakes and accidents corpus from North American News Corpus[1], respectively.

Barzilay and Lapata (2008) proposed to deal with local coherence with an Entity Grid Model. This model is based on Centering Theory (Grosz et al., 1995), whose assumption is that locally coherent texts present certain regularities concerning entity distribution. These regularities are calculated over an Entity Grid, i.e., a matrix in which the rows represent the sentences of the text and the columns the text entities. For example, Figure 2 shows part of the Entity Grid for the text in Figure 1. For instance, the "Depart." (Department) column in the grid (Figure 2) shows that the entity "Department" only happens in the first sentence in the Subject (S) position. Analogously, the marks O and X indicate the syntactical functions "Object" and "other syntactical functions" that are neither subject nor object, respectively. The hyphen ('-') indicates that the entity did not happen in the corresponding sentence.

Probabilities of entity transitions in texts may be computed from the entity grid and they compose a feature vector. For example, the probability of transition [O -] (i.e., the entity happened in the object position in one sentence and did not happen in the following sentence) in the grid in Figure 2 is 0.12, computed as the ratio between its occurrence in the grid (3 occurrences) and the total number of transitions (24).

---

1 (The Justice Department)$_S$ is conducting an (anti-trust trial)$_O$ against (Microsoft Corp.)$_X$ with (evidence)$_X$ that (the company)$_S$ is increasingly attempting to crush (competitors)$_O$.
2 (Microsoft)$_O$ is accused of trying to forcefully buy into (markets)$_X$ where (its own products)$_S$ are not competitive enough to unseat (established brands)$_O$.
3 (The case)$_S$ revolves around (evidence)$_O$ of (Microsoft)$_S$ aggressively pressuring (Netscape)$_O$ into merging (browser software)$_O$.
…

Figure 1. Text with syntactic tags (Barzilay and Lapata, 2008)

	Depart.	Trial	Microsoft	Evidence	Compet.	Markets	Products	Brands	Case	Netscape	Software	⋮	
1	S	O	S	X	O	-	-	-	-	-	-		1
2	-	-	O	-	-	X	S	O	-	-	-		2
3	-	-	S	O	-	-	-	-	S	O	O	-	3

Figure 2. Entity Grid (Barzilay and Lapata, 2008)

The authors evaluated the generated models in a text-ordering task (the one that interests us in this paper). In this task, each original text is considered "coherent", and a set of randomly sentence-permutated versions were produced and considered "incoherent" texts. Ranking values for coherent and incoherent texts were produced by a predictive model trained in the SVMlight (Joachims, 2002) package, using a set of text pairs (coherent text, incoherent text). It is supposed that the ranking values of coherent texts are higher than the ones for incoherent texts. Barzilay and Lapata obtained 87.2% and 90.4% of accuracy (fraction of correct pairwise rankings in the test set) applied respectively to the set of texts related to earthquakes and accidents, in English. Such results were achieved by a model considering three types of information, namely, coreference, syntactical and salience information.

---

1 https://catalog.ldc.upenn.edu/LDC95T21

Using coreference, it is possible to recognize different terms that refer to the same entity in the texts (resulting, therefore, in only one column in the grid). Syntax provides the functions of the entities; if not used, the grid only indicates if an entity occurs or not in each sentence; if salience is used, different grids are produced for more frequent and less frequent entities. It is important to notice that any combination of these features may be used.

Lin et al. (2011) assumed that local coherence implicitly favors certain types of discursive relation transitions. Based on the Entity Model from Barzilay and Lapata (2008), the authors used terms instead of entities and discursive information instead of syntactic information. The terms are the stemmed forms of open class words: nouns, verbs, adjectives and adverbs. The discursive relations used in this work came from the Penn Discourse Treebank (PDTB) (Prasad et al., 2008). The authors developed the Discursive Grid, which is composed of sentences (rows) and terms (columns) with discursive relations used over their arguments. For example, part of the discursive grid (b) for a text (a) is shown in Figure 3.

(S1) Japan normally depends heavily on the Highland Valley and Cananea mines as well as the Bougainville mine in Papua New Guinea. (S2) Recently, Japan has been buying copper elsewhere.

(a)

	Terms			
	copper	cananea	depend	…
S_1	nil	Comp.Arg1	Comp.Arg1	
S_2	Comp.Arg2 Comp.Arg1	nil	nil	

(b)

Figure 3. A text (a) and part of its grid (b)

A cell contains the set of the discursive roles of a term that appears in a sentence Sj. For example, the term "depend" in S1 is part of the Comparison (Comp) relation as argument 1 (Arg1), so the cell Cdepend,S1 contains the Comp.Arg1 role. The authors obtained 89.25% and 91.64% of accuracy applied to the set of English texts related to earthquakes and accidents, respectively.

Guinaudeau and Strube (2013) created an approach based on graph to eliminate the process of machine learning of the Entity Grid Model from Barzilay and Lapata (2008). Due to this, the

authors proposed to represent entities in a graph and then to model local coherence by applying centrality measures to the nodes in the graph. Their main assumption was that this bipartite graph contained the entity transition information needed for the computation of local coherence, thus feature vectors and a learning phase are unnecessary. Figure 4 shows part of the bipartite graph of the entity grid illustrated in Figure 2.
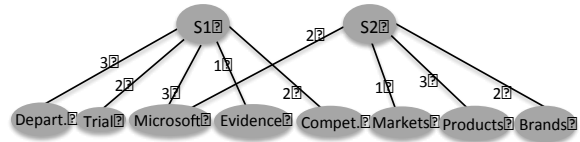

Figure 4. Bipartite graph

There is a group of nodes for the sentences and another group for the entities. Edges are stablished when the entities occur in the sentences, and their weights correspond to the syntactical function of the entities in the sentences (3 for subjects, 2 for objects and 1 for other functions).

Given the bipartite graph, the authors defined three kinds of projection graphs: *Unweighted One-mode Projection* (*PU*), *Weighted One-mode Projection* (*PW*) and *Syntactic Projection* (*PAcc*). In *PU*, weights are binary and equal to 1 when two sentences have at least one entity in common. In *PW*, edges are weighted according to the number of entities "shared" by two sentences. In *PAcc*, the syntactical weights are used. From *PU*, *PW* and *PAcc*, the local coherence of a text may be measured by computing the average outdegree of a projection graph. Distance information (*Dist*) between sentences may also be integrated in the weight of one-mode projections to decrease the importance of links that exist between non-adjacent sentences.

The approach was evaluated using the corpus from Barzilay and Lapata (2008). This model obtained 84.6% and 63.5% of accuracy in the Accidents and Earthquakes corpus, respectively.

Feng et al. (2014) is similar to Lin et al.'s (2011) work. Feng et al. (2014) created a discursive grid formed by sentences in rows and entities in columns. The cells of the grid are filled with RST relations together with nuclearity information. For example, Figure 5 shows a text fragment with 3 sentences and 7 EDUs. In Figure 6, a RST discourse tree representation of the text in Figure 5 is shown. Figure 7 shows a fragment of the RST-style discursive role grid of the text in Figure 5. This grid is based on the discursive tree representation in Figure 6. One may see in

Figure 7 that the entity "Yesterday" in sentence 1 occurs in the nuclei (N) of the Background and Temporal relations; the entity "session", in turn, is the satellite (S) of the Temporal relation.

S1: [The dollar finished lower yesterday,]e1 [after tracking another rollercoaster session on Wall Street.]e2
S2: [Concern about the volatile U.S. stock market had faded in recent sessions,]e3 [and traders appeared content to let the dollar languish in a narrow range until tomorrow,]e4 [when the preliminary report on third-quarter U.S. gross national product is released.]e5
S3: [But seesaw gyrations in the Dow Jones Industrial Average yesterday put Wall Street back in the spotlight]e6 [and inspired market participants to bid the U.S. unit lower.]e7
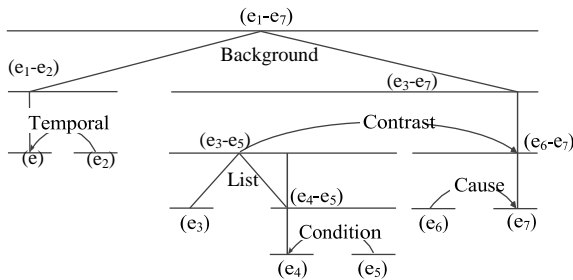
Figure 5. A text fragment (Feng et al., 2014)

Figure 6. RST discursive tree representation (Feng et al., 2014)

	dollar	Yesterday	session	…
S1	Background.N Temporal.N	Background.N Temporal.N	Temporal.S	…
S2	List.N Condition.N Contrast.N	nil	nil	…
S3	Contrast.N Background.N Cause.N	Cause.S	nil	…

Figure 7. Part of the RST-style discursive role grid for the example text (Feng et al., 2014)

Feng et al. (2014) developed two models: the Full RST Model and the Shallow RST Model. The Full RST Model uses long-distance RST relations for the most relevant entities in the RST tree representation of the text. For example, considering the RST discursive tree representation in Figure 6, the Background relation was encoded for the entities "dollar" and "Yesterday" in S1, as well as the entity "dollar" in S3, but not for the remaining entities in the text, even though the

Background relation covers the whole text. The corresponding full RST-style discursive role matrix for the example text is shown in Figure 7. The shallow RST Model only considers relations that hold between text spans of the same sentence, or between two adjacent sentences. The Full RST Model obtained an accuracy of 99.1% and the Shallow RST Model obtained 98.5% of accuracy in the text-ordering task.

Dias et al. (2014b) also implemented a coherence model that uses RST relations. The authors created a grid composed by sentences in rows and entities in columns. The cells were filled with RST relation. This model was applied to a corpus of news texts written in Brazilian Portuguese. This model had the accuracy of 79.4% with 10-fold cross validation in the text-ordering task. This model is similar to the Full RST Model. These models were created in parallel and used in corpora of different languages. Besides the corpus and the language, the Shallow RST Model only uses the RST relations of a sentence and/or adjacent sentences, while Dias et al. capture all the possible relations among sentences.

Regarding the model of Lin et al. (2011), the discursive information used by Lin et al. and Dias et al. is the main difference between these models, i.e., Dias et al. use RST relations and Lin et al. use PDTB-style discursive relations.

Castro Jorge et al. (2014) combined CST relations and syntactic information in order to evaluate the coherence of multi-document summaries. The authors created a CST relation grid represented by sentences in the rows and in the columns, and the cells were filled with 1 or 0 (presence/absence of CST relations – called Entity-based Model with CST bool). This model was applied to a corpus of news summaries written in Brazilian Portuguese and it obtained 81.39% of accuracy in the text-ordering task. Castro Jorge et al.'s model differs from the previous models since it uses CST information and a summarization corpus (instead of full texts).

3   The Discursive Model

The model proposed in this paper considers that all coherent multi-document summaries have patterns of discursive relation (RST and CST) that distinguish them from the incoherent (less coherent) multi-document summaries.

The model is based on a grid of RST and CST relations. Then, a predictive model that uses the probabilities of relations between two sen-

tences as features was trained by the SVM[light] package and evaluated in the text-ordering task.

As an illustration, Figure 8 shows a multi-document summary. The CST relation "Follow-up" relates the sentences S2 and S3. Between the sentences S1 and S3, there is the RST relation "elaboration". The RST relation "sequence" happens between S1 and S4. After the identification of the relations in the summary, a grid of discursive relations is created. Figure 9 shows the discursive grid for the summary in Figure 8. In this grid, the sentences of the summary are represented in the rows and in the columns. The cells are filled with RST and/or CST relations that happen in the transition between the sentences (the CST relations have their first letters capitalized, whereas RST relations do not).

> (S1) Ended the rebellion of prisoners in the Justice Prisoners Custody Center (CCPJ) in São Luís, in the early afternoon of Wednesday (17).
> (S2) After the prisoners handed the gun used to start the riot, the Military Police Shock troops entered the prison and freed 30 hostages - including 16 children.
> (S3) The riot began during the Children's Day party, held on Tuesday (16).
> (S4) According to the police, the leader of the rebellion was transferred to the prison of Pedrinhas, in the capital of Maranhão.

Figure 8. Summary with discursive information from the CSTNews corpus (Cardoso et al., 2011)

	S1	S2	S3	S4
S1		-	elaboration	Sequence
S2			Follow-up	-
S3				-
S4				

Figure 9. Discursive grid for Figure 8

Consider two sentences S_i and S_j (where i and j indicate the place of the sentence in the summary): if $i < j$, it is a valid transition and 1 is added to the total of possible relationships. Considering that the transitions are visualized from the left to the right in the discursive grid in Figure 9, the cells in gray do not characterize a valid transition (since only the superior diagonal of the grid is necessary in this model).

The probabilities of relations present in the transitions are calculated as the ratio between the frequency of a specific relation in the grid and the total number of valid transitions between two sentences. For instance, the probability of the RST relation "elaboration" (i.e., the relation "elaboration" to happen in a valid transition) in the grid in Figure 9 is 0.16, i.e., one occurrence of "*elaboration*" in 6 possible transitions.

The probabilities of all relations present in the summary (both RST and CST relations) form a feature vector. The feature vectors for all the summaries become training instances for a machine learning process. In Figure 10, part of the feature vector for the grid in Figure 9 is shown.

Follow-up	elaboration	sequence	…
0.16	0.16	0.16	…

Figure 10. Part of the feature vector for Figure 9

4 Experiments and Results

The text-ordering task from Barzilay and Lapata (2008) was used to evaluate the performance of the proposed model and to compare it with other methods in literature.

The corpus used was the CSTNews[2] from Cardoso et al. (2011). This corpus has been created for multi-document summarization. It is composed of 140 texts distributed in 50 sets of news texts written in Brazilian Portuguese from various domains. Each set has 2 or 3 texts from different sources that address the same topic. Besides the original texts, the corpus has several annotation layers: (i) CST and RST manual annotations; (ii) the identification of temporal expressions; (iii) automatic syntactical analyses; (iv) noun and verb senses; (v) text-summary alignments; and (vi) the semantic annotation of informative aspects in summaries; among others. For this work, the CST and RST annotations have been used.

Originally, the CSTNews corpus had one extractive multi-document summary for each set of texts. However, Dias et al (2014a) produced 5 more extractive multi-document summaries for each set of texts. Now, the corpus has 6 reference extractive multi-document summaries for each set of texts. In this work, 251 reference multi-document extracts (with average size of 6.5 sentences) and 20 permutations for each one (totalizing 5020 summaries) were used in the experiments.

Besides the proposed model, some other methods from the literature have also been re-implemented in order to compare our results to the current state of the art. The following methods were chosen based on their importance and on the techniques used to evaluate local coher-

---

[2] www.icmc.usp.br/~taspardo/sucinto/cstnews.html

ence: the *LSA* method of Foltz et al. (1998), the *Entity Grid Model* of Barzilay and Lapata (2008), the *Graph Model* of Guinaudeau and Strube (2013), the *Shallow RST Model* of Feng et al (2014), the *RST Model* of Dias et al. (2014b) and the *Entity-based Model with CST bool* of Castro Jorge et al. (2014). The *LSA* method, *Entity Grid*, *Graph* and *Shallow RST* Models were adapted to Brazilian Portuguese, using the appropriate available tools and resources for this language, as the PALAVRAS parser (Bick, 2000) that was used to identify the summary entities, which are all nouns and proper nouns. The implementation of these methods carefully followed each step of the original ones.

Barzilay and Lapata's method has been implemented without coreference information, since, to the best of our knowledge, there is no robust coreference resolution system available for Brazilian Portuguese, and the CSTNews corpus still does not have referential information in its annotation layers. Furthermore, the implementation of Barzilay and Lapata's approach produced 4 models: with syntax and salience information (referred by Syntactic+Salience+), with syntax but without salience information (Syntactic+Salience-), with salience information but without syntax (Syntactic-Salience+), and without syntax and salience information (Syntactic-Salience-), in which salience distinguishes entities with frequency higher or equal to 2.

The Full RST Approach is similar to Dias et al.'s model (2014b), and then it was not used in these experiments.

Lin et al.'s model (2011) was not used in the experiments, since the CSTNews corpus does not have the PDTB-style discursive relations annotated. However, according to Feng et al. (2014), the PDTB-style discursive relations encode only very shallow discursive structures, i.e., the relations are mostly local, e.g., within a single sentence or between two adjacent sentences. Due to this, the Shallow RST Model from Feng et al. (2014), which behaves as Lin et al.'s (2001), was used in these experiments.

Table 1 shows the accuracy of our approach compared to the other methods, ordered by accuracy.

Models	Acc. (%)
Our approach	**92.69**
Syntactic-Salience- of Barzilay and Lapata	68.40*
Syntactic+Salience+ of Barzilay and Lapata	64.78*
Syntactic-Salience+ of Barzilay and Lapata	61.99*
Syntactic+Salience- of Barzilay and Lapata	60.21*
Graph Model of Guinaudeau and Strube	57.69*
LSA of Foltz et al.	55.18*
RST Model of Dias et al.	51.32*
Shallow RST Model of Feng et al.	48.92*
Entity-based Model with CST bool of Castro Jorge et al.	32.53*

Table 1. Results of the evaluation, where diacritics * (p < .01) indicates whether there is a significant statistical difference in accuracy compared to our approach (using t-test)

The t-test has been used for pointing out whether differences in accuracy are statistically significant or not. Comparing our approach with the other methods, one may observe that the use of all the RST and CST relations obtained better results for evaluating the local coherence of multi-document summaries.

These results show that the combination of RST and CST relations with a machine learning process has a high discriminatory power. This is due to discursive relation patterns that are present in the transitions between two sentences in the reference summaries. The "elaboration" RST relation was the one that presented the highest frequency, 237 out of the 603 possible ones in the reference summaries. The transition between S1 and S2 in the reference summaries was the transition in which the "elaboration" relation more frequently occurred, 61 out of 237. After this one, the RST relation "list" had 115 occurrences, and the transition between S3 and S4 was the more frequent to happen with the "list" relation (17 times out of 115 occurrences).

The *Shallow RST Model* from Feng et al. (2014) and the *Entity-based Model with CST bool* from Castro Jorge et al. (2014), that also use discursive information, obtained the lowest accuracy in the experiments. The low accuracy may have been caused for the following reasons: (i) the discursive information used was not sufficient for capturing the discursive patterns of the reference summaries; (ii) the quantity of features used by these models negatively influenced in the learning process; and (iii) the type of text used in this work was not appropriate, because the *RST Model* of Dias et al. (2014b) and the *Shallow RST Model* of Feng et al. (2014) had better results with full/source texts. Besides this,

the quantity of summaries may have influenced the performance of the *Entity-based Model with CST bool* of Castro Jorge et al. (2014), since their model was originally applied in 50 multi-document summaries, while 251 summaries were used in this work

The best result of the *Graph Model* of Guinaudeau and Strube (2013) (given in Table 1) used the *Syntactic Projection (PAcc)*, without distance information (*Dist*).

Overall, our approach highly exceeded the results of the other methods, since we obtained a minimum gain of 35.5% in accuracy.

5   Final remarks

According to the results obtained in the text-ordering task, the use of RST and CST relations to evaluate local coherence in multi-document summaries obtained the best accuracy in relation to other tested models. We believe that such discourse information may be equally useful for dealing with full texts too, since it is known that discourse organization highly correlates with (global an local) coherence.

It is important to notice that the discursive information used in our model is considered as "subjective" knowledge and that automatically parsing texts to achieve it is an expensive task, with results still far from ideal. However, the obtained gain in comparison with the other approaches suggests that it is a challenge worthy of following.

Acknowledgements

References

Aleixo, P. and Pardo, T.A.S. 2008. CSTNews: Um Córpus de Textos Jornalísticos Anotados Segundo a Teoria Discursiva Multidocumento CST (Cross-Document Structure Theory). Technical Report Interinstitutional Center for Computational Linguistics, University of São Paulo, n. 326. p.12. São Carlos-SP, Brazil.

Barzilay, R. and Lapata, M. 2008. Modeling local coherence: An entity-based approach. Computational Linguistics, v. 34, n. 1, p. 1-34, Cambridge, MA, USA.

Bosma, W. 2004. Query-Based Summarization using Rhetorical Structure Theory. In Proceedings of the15th Meetings of CLIN, LOT, Utrecht, pp. 29-44.

Bick, E. 2000. The Parsing System Palavras, Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework, Aarhus University Press.

Cardoso, P., Maziero, E., Jorge, M., Seno, E., di Felippo, A., Rino, L., Nunes, M. and Pardo, T. 2011. CSTNews - a discourse-annotated corpus for single and multi-document summarization of news texts in brazilian portuguese. In Proceedings of the 3rd RST Brazilian Meeting. p. 88-105.

Castro Jorge, M.L.R., Dias, M.S. and Pardo, T.A.S. 2014. Building a Language Model for Local Coherence in Multi-document Summaries using a Discourse-enriched Entity-based Model. In the Proceedings of the Brazilian Conference on Intelligent Systems - BRACIS, p. 44-49. São Carlos-SP/Brazil.

Dias, M.S.; Bokan Garay, A.Y.; Chuman, C.; Barros, C.D.; Maziero, E.G.; Nobrega, F.A.A.; Souza, J.W.C.; Sobrevilla Cabezudo, M.A.; Delege, M.; Castro Jorge, M.L.R.; Silva, N.L.; Cardoso, P.C.F.; Balage Filho, P.P.; Lopez Condori, R.E.; Marcasso, V.; Di Felippo, A.; Nunes, M.G.V.; Pardo, T.A.S. 2014a. Enriquecendo o Corpus CSTNews - a Criação de Novos Sumários Multidocumento. In the (on-line) Proceedings of the I Workshop on Tools and Resources for Automatically Processing Portuguese and Spanish - ToRPorEsp, p. 1-8. São Carlos-SP/Brazil.

Dias, M.S.; Feltrim, V.D.; Pardo, T.A.S. 2014b. Using Rhetorical Structure Theory and Entity Grids to Automatically Evaluate Local Coherence in Texts. In the Proceedings of the 11st International Conference on Computational Processing of Portuguese - PROPOR (LNAI 8775), p. 232-243. October 6-9. São Carlos-SP/Brazil.

Dijk, T.V. and Kintsch, W. 1983. Strategics in discourse comprehension. Academic Press. New York.

Feng, V. W., Lin, Z. and Hirst G. 2014. The Impact of Deep Hierarchical Discourse Structures in the Evaluation of Text Coherence. In the Proceedings of the 25th International Conference on Computational Linguistics, p. 940-949, Dublin, Ireland.

Foltz, P. W., Foltz, P. W., Kintsch, W. and Landauer, T. K. 1998. The measurement of textual coherence with latent semantic analysis. Discourse Processes, v. 25, n. 2 & 3, p. 285-307.

Grosz, B., Aravind, K. J. and Scott, W. 1995. Centering: A framework for modeling the local coherence of discourse. Computational Linguistics, vol. 21, p. 203-225. MIT Press Cambridge, MA, USA.

Guinaudeau, C. and Strube, M. 2013. Graph-based Local Coherence Modeling. In the Proceedings of the 51st Annual Meeting of the Association for

Computational Linguistics. v. 1. p. 93-103, Sofia, Bulgaria.

Joachims T. 2002. Optimizing search engines using clickthrough data. In the Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, p. 133–142. New York, NY, USA.

Jorge, M.L.C., Pardo, T.A.S. 2010. Experiments with CST-based Multidocument Summarization. In the Proceedings of the ACL Workshop TextGraphs-5: Graph-based Methods for Natural Language Processing, pp. 74-82, Uppsala/Sweden.

Kibble, R., Power, R. 2004. Optimising referential coherence in text generation. Computational Linguistic, vol. 30 n. 4, pp. 401-416.

Koch, I. G. V. and Travaglia, L. C. 2002. A coerência textual. 14rd edn. Editora Contexto.

Landauer, T. K., Dumais, S. T. 1997. A solution to Plato's problem: The latent semantic analysis theory of acquisition, induction and representation to coreference resolution. In Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics, pp. 104 -111, Philadelphia, PA.

Lin, Z., Ng, H. T. and Kan, M.-Y. 2011. Automatically evaluating text coherence using discourse relations. In the Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies – v. 1, p. 997–1006, Stroudsburg, PA, USA.

Mani, I. (2001). Automatic Summarization. John Benjamins Publishing Co., Amsterdam.

Mann, W. C. and Thompson, S. A. 1987. Rhetorical Structure Theory: A theory of text organization. Technical Report, ISI/RS-87-190.

Mckoon, G. and Ratcliff, R. 1992. Inference during reading. Psychological Review, p. 440-446.

Prasad, R., Dinesh, N., Lee, A., Miltsakaki, E., Robaldo, L., Joshi, A. and Webber, B. 2008. The penn discourse treebank 2.0. In the Proceedings of the 6th Internacional Conference on Language Resources an Evaluation.

Radev, D.R. 2000. A common theory of information fusion from multiple text sources, step one: Cross-document structure. In the Proceedings of the 1st ACL SIGDIAL Workshop on Discourse and Dialogue, Hong Kong.

Salton, G. 1988. Term-Weighting Approaches in Automatic Text Retrieval. Information Processing and Management, p. 513-23.

Seno, E. R. M. 2005. Rhesumarst: Um sumarizador automático de estruturas RST. Master Thesis. University of São Carlos. São Carlos/SP.