Proceedings of Symposium in Information and Human Language Technology. Natal, RN, Brazil, November 4–7, 2015. ©2015 Sociedade Brasileira de Computação.

Enriching entity grids and graphs with discourse relations: the impact in local coherence evaluation

Márcio de S. Dias and Thiago A. S. Pardo

Interinstitutional Center for Computational Linguistics (NILC)
Institute of Mathematical and Computer Sciences, University of São Paulo
Av. Trabalhador São-carlense, 400 - Centro
CEP: 13566-590 - São Carlos/SP, Brazil.

{marciosd,taspardo}@icmc.usp.br

Abstract. *This paper describes how discursive knowledge, given by the discursive theories RST (Rhetorical Structure Theory) and CST (Cross-document Structure Theory), may improve the automatic evaluation of local coherence in multi-document summaries. Two of the main coherence models from literature were incremented with discursive information and obtained 91.3% of accuracy, with a gain of 53% in relation to the original results.*

1. Introduction

Coherence is an important aspect that affects the quality of texts produced by textual generators such as summarizers, question/answering systems, etc. A coherent multi-document summary makes reading and understanding easier than one summary with contradictions and repetitive information.

According to Koch and Travaglia (2002), coherence means the possibility of establishing a meaning for the text. 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) of the text, inferences, intertextuality, intentionality and acceptability, among others [Kock and Travagila 2002].

Textual coherence occurs in local and global levels [Dijk and Kintsch 1983]. Local level coherence is presented by the local relationships among the parts of a text, for instance, adjacent sentences and shorter sequences. On the other hand, a text presents global coherence when this text links all its elements as a whole. Local coherence is essential in order to achieve global coherence [Mckoon and Ratcliff 1992]. Thus, many researches in computational linguistics have been developing models for dealing with local coherence ([Barzilay and Lapata 2005], [Barzilay and Lapata 2008], [Burstein et al. 2010], [Castro Jorge 2014], [Dias et al. 2014b], [Eisner and Charniak 2011], [Elsner et al. 2007], [Feng et al. 2014], [Filippova and Strube 2007], [Foltz et al. 1998], [Freitas 2013], [Guinaudeau and Strube 2013], and [Lin et al 2011]).

To illustrate the problem we have in hands, Figure 1 shows two summaries, a coherent (Summary A) and a less coherent one (Summary B). Summary B presents redundant information among the sentences: S1 with S3, and S2 with S4. These redundancies damage the quality and the informativity of the text and, consequently, its coherence.

Summary A (coherent summary)	Summary B (incoherent summary)
(S1) In the last five years, astronomers have identified a few dozen objects that are even smaller than brown dwarfs that are not bound to any star system, nicknamed the planetary mass objects, or planemos located around star-forming regions. **(S2)** By using telescopes at the European Southern Observatory (ESO), astronomers have discovered a planet that is seven times the size of Jupiter, the heaviest that revolves around the sun, and the other that is twice its size. **(S3)** The mass of these two worlds is similar to other already cataloged exoplanets but they do not revolve around a star, they revolve around each other. **(S4)** Ray Jayawardhana, from the University of Toronto, and Valentin Ivanov, from the European Southern Observatory, have published the findings in "Science Express", the "Science" magazine website.	**(S1)** By using telescopes at the European Southern Observatory (ESO), astronomers have discovered a planet that is seven times the size of Jupiter, the heaviest that revolves around the sun, and the other that is twice its size. **(S2)** The mass of these two worlds is similar to other already cataloged exoplanets but they do not revolve around a star, they revolve around each other. **(S3)** The biggest celestial body, whose size is seven times greater than Jupiter, was detected about 400 light years from our solar system. **(S4)** The extraordinary fact is that it does not revolve around a star, but around another cold body that is twice its size.

Figure 1. Examples of coherent (A) and incoherent (B) summaries

The discursive information used in this work is related to intra or inter text organization, i.e., the Rhetorical Structure Theory (RST) [Mann and Thompson 1987] and the Cross-document Structure Theory (CST) [Radev 2000], respectively. 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, whereas CST organizes multiple texts on the same topic and establishes relations among different textual segments, forming a graph.

Considering that all well-formed and coherent texts have a well-defined discursive organization, this paper shows how discursive information (RST and CST) may improve the accuracy of local coherence models in order to automatically differentiate coherent from incoherent (less coherent) summaries. Thus, local coherence models from the literature have been enriched with discursive information. In addition, the original approaches have been re-implemented to have their performances analyzed with the corpus of multi-document summaries used in this work. In particular, this work is based on the following assumptions: (i) there are regularities on the distribution of discursive relations in coherent summaries; (ii) coherent summaries show distinct organization of intra- and inter-discursive relations. We show that such assumptions hold and that we improve the original results in the area.

Section 2 presents an overview of the most relevant researches related to local coherence. In Section 3, the coherence models proposed in this work are described. Section 4 shows the experimental setup and the obtained results. Finally, Section 5 presents some final remarks.

2. Related Work

One of the most used local coherence models is the one of Barzilay and Lapata (2008), which proposed an Entity Grid Model to evaluate local coherence, i.e., to classify coherent or incoherent texts. This model is based on Centering Theory [Grosz et al. 1995]; the authors' hypothesis is that locally coherent texts present certain regularities concerning entity distribution. These regularities are calculated over a matrix (entity grid) in which the rows represent the sentences of the text, and the columns the text entities.

Barzilay and Lapata's approach used (+) or not (-) syntactical, coreference and salience information. The syntactical information uses the grammatical function of the entities. For example, in the "Department" column in the entity grid in Figure 2b, it is shown that the "Department" entity happens in the first sentence in the subject (S) position. The hyphen ('-') indicates that the entity did not happen in the corresponding sentence, (O) object position and (X) nor subject or object. Coreference occurs when words refer to the same entity and, therefore, these words may be represented by a single column in the grid. For example, when the text in Figure 2a mentions "Microsoft Corp.", "Microsoft", and "the company", such references are mapped to a single column ("Microsoft") in its entity grid in Figure 2b. Salience is related to the frequency of entities in texts, allowing to build grids with the least and/or the most frequent entities in the text.

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]$_O$.

4 [Microsoft]$_S$ claims [its tactics]$_S$ are commonplace and good economically.

5 [The government]$_S$ may file [a civil suit]$_O$ ruling that [conspiracy]$_S$ to curb [competition]$_O$ through [collusion]$_X$ is [a violation of the Sherman Act]$_O$.

6 [Microsoft]$_S$ continues to show [increased earnings]$_O$ despite [the trial]$_X$.

(a)

	Department	Trial	Microsoft	Evidence	Competitors	Markets	Products	Brands	Case	Netscape	Software	Tactics	Government	Suit	Earnings	
1	S	O	S	X	O	–	–	–	–	–	–	–	–	–	–	1
2	–	–	O	–	–	X	S	O	–	–	–	–	–	–	–	2
3	–	–	S	O	–	–	–	–	S	O	O	–	–	–	–	3
4	–	–	S	–	–	–	–	–	–	–	–	S	–	–	–	4
5	–	–	–	–	–	–	–	–	–	–	–	–	S	O	–	5
6	–	X	S	–	–	–	–	–	–	–	–	–	–	–	O	6

(b)

Figure 2. Text (a) and its Entity Grid (b) [Barzilay and Lapata, 2008]

From this grid, the number of times that each possible transition occurs in the grid is computed and, then, its probability is calculated. 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 presented in Figure 2b is 0.09, computed as the ratio between its frequency of occurrence in the grid (7 occurrences) and the total number of transitions (75 transitions).

The probabilities of transitions form a characteristic vector for each text of a corpus. The characteristic vector becomes one training instance for a machine learning process using the SVM^{light} [Joachims 2002] package.

The generated models were used in a text-ordering task (the one that interests to us in this paper). For each original text considered "coherent", a set of randomly sentence permutated versions were produced and this set was considered as "incoherent" texts. Ranking values for coherent and incoherent texts were produced by the predictive model trained in the SVM^{light} package, using a set of pairs of texts (coherent text, incoherent text). According to Barzilay and Lapata (2008), 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) using, respectively, sets of texts on earthquakes and accidents, in English.

Freitas (2013) also applied Barzilay and Lapata's entity model to evaluate coherence in newspaper texts written in Brazilian Portuguese and obtained 74.4% of accuracy with syntactic and salience information applied to the corpus.

Lin et al. (2011) created one of the first models that use discursive information to evaluate local coherence. The authors' assumption is that local coherence implicitly favors certain types of discursive relation transitions. Lin et al. used four discursive relations, based on the Discourse Lexicalized Tree Adjoining Grammar (D-LTAG) [Webber 2004], to develop the Discourse Role Matrix, which is composed of sentences (rows) and terms (columns), with discursive relations used over their arguments. Terms were the stemmed forms of the open class words. For example, see the discursive grid (b) of the text (a) in Figure 3, both reproduced from Lin et al. (2011).

(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 bying copper elsewhere.

(a)

S#	Terms			
	copper	cananea	depend	…
S1	nil	Comp.Arg1	Comp.Arg1	
S2	Comp.Arg2 Comp.Arg1	nil	nil	

(b)

Figure 3. Part of the text and its discursive grid [Lin et al., 2011]

Figure 3b shows a matrix, whose columns correspond to the extracted terms of the text in Figure 3a and the rows represent the contiguous sentences. A cell $C_{Ti,Sj}$ contains the set of the discursive roles of the term T_i that appears in sentence S_j. For example, the term "depend" in S1 takes part of the Comparison (Comp) relation as argument 1 (Arg1), so the cell $C_{depend,S1}$ contains the Comp.Arg1 role. A cell may be empty (nil, as in $C_{depend,S2}$) or contain multiple discursive roles (as in $C_{copper,S2}$, since "copper" in S_2 participates in two relations). The authors obtained 89.25% and 91.64% of accuracy using the sets of texts on earthquakes and accidents, respectively.

Guinaudeau and Strube (2013) consider some disadvantages in the Entity Grid Model, such as: data sparsity, domain dependence and computational complexity. The authors then proposed to represent entities in a graph and to model local coherence by applying centrality measures to the nodes in the graph. Their main assumption is that this (bipartite) graph contains the entity transition information needed for local coherence computation, causing feature vectors and a learning phase unnecessary. Figure 4 shows part of a graph of the entity grid illustrated in Figure 2b.

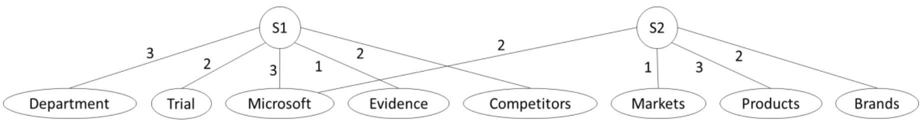


Figure 4. Bipartite Graph

According to the graph in Figure 4, an edge between a sentence node S_i and an entity node e_j is created if the corresponding cell c_{ij} in the entity grid is not equal to "-". Each edge is associated with a weight $w(e_j, S_i)$ that is dependent on the grammatical role of the entity e_j (S = 3; O = 2; X = 1) in the sentence S_i. Given the graph, the authors defined three kinds of projection: *Unweighted One-mode Projection* (*PU*), *Weighted One-mode Projection* (*PW*) and *Syntactic Projetion* (*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*, syntactic information is accounted for by integrating the edge weights in the bipartite graph. The distance between sentences S_i and S_k may also be integrated in the weight of one-mode projections in order to decrease the importance of links that exist between non-adjacent sentences. From *PU*, *PW* and *PAcc*, the local coherence of a text *T* may be measured by computing the average outdegree of a projection graph.

According to Guinaudeau and Strube (2013), coherent texts present a coherence value higher than incoherent ones. Due to this, the model obtained 84.6% and 63.5% of accuracy in the accidents and earthquakes corpora, respectively.

Feng et al. (2014) and Dias et al. (2014b) are based on Lin. et al. (2011), however both use Rhetorical Structure Theory relations with nuclearity information (Nuclei and Satellites) instead of the D-LTAG information. The authors also use entities instead of terms to create a new Discursive Role Matrix. With these modifications, the authors created the Full RST-style Model and Feng et al. created the Shallow RST-style Model. The Full RST-style Model encodes long-distance discursive relations for the entities. The Shallow RST-style Model only considers relations that hold between text spans of the same sentence, or between two adjacent sentences. Feng et al. used a corpus formed by 735 texts of the Wall Street Journal (WSJ) and 20 permutations for each source text have been used. The Full RST-style Model from Feng et al. obtained an accuracy of 99.1%, and the Shallow RST-style Model obtained 98.5% of accuracy, in the text-ordering task. Dias et al. used a corpus of 140 news texts in Portuguese with 20 permutations for each text. The Full RST-style Model from Dias et al. obtained 79.4% of accuracy with 10-fold cross validation in the sentence ordering task.

Castro Jorge et al. (2014) combined CST relations and syntactic information to evaluate the coherence of multi-document summaries. The authors created a CST relation grid represented by sentences in rows and in columns, and the cells are filled with 1 or 0 (presence/absence of relations). Their corpus was composed of 50 multi-document summaries (considered coherent) in Brazilian Portuguese and 20 permutations for each summary have been used. The SVM^{light} was also used to create the predictive model. This approach obtained the accuracy of 81.39% in the text-ordering task.

3. Local coherence models with discursive information

In order to demonstrate the impact of discursive information on the evaluation of local coherence in multi-document summaries written in Brazilian Portuguese, the Entity Grid and the Graph Models have been re-implemented and new versions with discursive information were developed.

The Entity Grid Model was re-implemented considering syntactic information. The reference information was not used since there is not a robust tool to resolve coreference for Brazilian Portuguese. Our proposal is to combine one entity grid of syntactic information from Barzilay and Lapata, as in Figure 2b, with one grid of discursive information, as in Figure 5, that considered CST information to form the discursive grid. This grid records the CST relations that happen between two adjacent sentences. The same idea was used when RST or RST/CST information were considered to create the discursive grid. Thus, the model works with two grids, one based on syntactic information and the other with discursive information (CST, RST or both).

	S_1	S_2	S_3	S_4	S_5	S_6
S_1	–	Elaboration	–	–	–	-
S_2	–	–	–	–	–	-
S_3	–	–	–	Elaboration	–	-
S_4	–	–	–	–	Follow-up	-
S_5	–	–	–	–	–	Equivalence
S_6	-	-	-	-	-	-

Figure 5. CST Grid

The probabilities of transitions are calculated by considering the discursive information between sentences. Figure 6 shows part of a feature vector related to the grids in Figures 2 and 5.

SSElaboration	S-Follow-up	S-Equivalence	O-Elaboration	SXFollow-up	SXEquivalence	….
0.013	0.026	0.013	0.08	0	0	

Figure 6. Part of a feature vector that combines syntactic information with CST relations

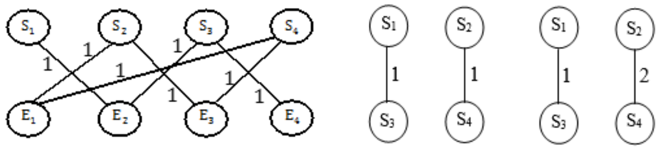
The transitions in Figures 2 and 5 are considered as features. The number of features are 160, which is the result of multiplying 16 (number of possible combinations of syntactic patterns of the entity-based model) *10 (total number of CST relations). The probability values in Figure 6 are the results of dividing the total of each pattern by 75, which is the total number of transitions for the entity grid in Figure 2b. For example, the pattern "*O-Elaboration*" is calculated by the frequency of the transition "*O-"* (obtained from the entity grid) together with the occurrence of the *Elaboration* relation (obtained from the discursive grid) in one of the sentences of the transition "*O-"*. Thus, for this pattern, the probability value 0.08 was obtained by dividing the number of times that this pattern appeared in the text by 75.

In the Graph Model with discourse, a discursive incidence grid (see Figure 7a) was created, where the rows represent the sentences (S_i) and the columns the entities (E_j) of the summary. The cells $C_{Si,Ej}$ in this grid are filled with the occurrence of discursive information (RST and/or CST), i.e., $C_{Si,Ej} = 1$ when an entity is part of a sentence that participates in a discursive relation. For instance, entity 1 (E_1) occurs in sentences S_2 and S_4, both related to another sentence by RST and/or CST relations.

The Bipartite Graph is generated from the discursive incidence grid (see Figure 7a). Figure 7b shows this graph, whose edges are associated with a weight $w(E_i, S_j) = 1$ when there is a discursive relation in the sentence (S_j) that entity (E_i) belongs to. Figure 7c e 8d show the *PU* and *PW* projection graphs, respectively, which were generated from the bipartite graph (Figure 7b). Therefore, local coherence was calculated in the same way that the original model.

	E_1	E_2	E_3	E_4
S_1	-	1	-	-
S_2	1	-	1	-
S_3	-	1	-	1
S_4	1	-	1	-

(a)



(b) (c) (d)

Figure 7. (a) Discursive Incidence Grid, (b) Bipartite Graph, (c) PU Graph, and (d) PW Graph

4. Experiments and Results

In order to show that the use of discursive relations may improve the evaluation of local coherence in multi-document summaries, the text-ordering task from Barzilay and Lapata (2008) and the following models, which use (+) or not (-) syntactic and salience information, have been used: (Syntactic+Salience+), (Syntactic-Salience-), (Syntactic-Salience+) and (Syntactic+Salience-) from Barzilay and Lapata, the models from Guinaudeau and Strube (2013) – the PU Project Graph Model without distance information (PU-DI), the PW Project Graph Model without distance information (PW-DI), the PU Project Graph Model with distance information (PU+DI) and the PW Project Graph Model with distance information (PW+DI) – considering the discursive versions developed in this work. Syntactic Projetion (PAcc) was not used in the experiments because of the low accuracy in its original version.

The experiments were conducted over the CSTNews corpus, which is a set of CST and RST manually annotated texts in Brazilian Portuguese [Cardoso et al. 2011]. The corpus in its original version is composed of 140 texts distributed in 50 sets of news texts from various domains. Each cluster contains 2 or 3 texts, with CST and RST annotations, and their correspondent multi-document summary, which is an extract. Due to the need of more multi-document summaries for the corpus, Dias et al. (2014a) used a methodology to create human multi-document summaries for the corpus. Today, the corpus has 5 more extractive and 5 more abstractive summaries for each cluster.

For the experiments, 251 extractive multi-document summaries (considered coherent) were used and, for each of these, 20 permutations (considered incoherent) have been generated, totalizing 5020 pairs of summaries. They compose the instances for the learning process with SVM^{light}. 10-fold cross-validation was used to train and test the models. Table 1 shows the accuracy achieved by the original models and by the modified ones (with discursive information).

Table 1. Results of the evaluation, where diacritics * (p < .01) indicate whether there is a significant statistical difference in accuracy compared to the best result (in bold) of each approach (using T-test)

Entity grids	Acc (%)	Graphs	Acc (%)
Syntactic+Salience+	64.78*	*PU-DI*	57.69*
Syntactic-Salience-	68.40*	*PW-DI*	54.98*
Syntactic-Salience+	61.90*	*PU+DI*	52.71*
Syntactic+Salience-	60.21*	*PW+DI*	51.21*
Syntactic-Salience- with RST	84.47*	*PU-DI with RST and CST*	**80.22**
Syntactic-Salience- with CST	91.13	*PW-DI with RST and CST*	79.66*
Syntactic-Salience- with RST and CST	76.80*	*PU+DI with RST and CST*	78.50*
Syntactic+Salience- with RST	81.85*	*PW+DI with RST and CST*	78.43*
Syntactic+Salience- with CST [Castro Jorge et al. 2014]	**91.31**	-	-
Syntactic+Salience- wiht RST and CST	75.14*	-	-

The t-test has been used for pointing out whether differences in accuracy are statistically significant, by comparing the best discursive model of each approach (bold values in Table 1) with other models of the same approach (Table 1).

In particular, the results showed that the use of discursive information of CST and RST relations significantly increased the accuracy. In all the enriched variations with

RST and/or CST relations in the *Syntactic+Salience-* and *Syntactic-Salience-* models, the accuracy was better than the ones obtained by the original models from Barzilay and Lapata. This probably happened due to the addition of discursive information, which defined better the patterns of coherent and incoherent summaries, and thus improved the evaluation of the methods. The *Syntactic+Salience- with CST* model from Castro et al. presented the best accuracy among all the evaluated models. In this case, the CST relations improved the accuracy of the original model in 51.65%, which is considered the best gain for this approach.

The reference summaries (considered coherent) presented transition patterns found by the models incremented with discursive information. In our experiments, the highest occurrence pattern was "--*Elaboration*": it happened 176 times in 976 valid transition patterns on the reference summaries. After this one, the transition patterns "--*Follow-up*" and "--*Overlap*" had 139 and 114 occurrences, respectively.

All the Graph Models from Guinaudeau and Strube (2013) enriched with RST and CST relations obtained better accuracy than the original Graph Models. Within the Graph Models with discursive information, the *PU-DI with RST and CST* model presented the best accuracy and it obtained 39.05% of gain. However, for this approach, the "*PW+DI with RST and CST"* model obtained the best gain in accuracy – 53.15%.

Models with CST information obtained better results, which may be justified by the availability of more CST relations than RST relations in multi-document summaries.

5. Final Remarks

According to the results obtained in the text-ordering task, the discursive information substantially improved the evaluation of local coherence in multi-document summaries in the two approaches of the literature. Although the discursive information is considered "expensive", due to its subjectivity, it is a powerful knowledge and should be further computationally explored (with robust discursive parsers for Brazilian Portuguese). Thus, this approach proved to be promising and it may be used for other languages, such as English, as long as there is a corpus with CST and RST annotations and a syntactic parser.

As future work, the same methodology used in this work will be used on new methods to improve the local coherence evaluation of multi-document summaries.

Acknowledgements

The authors are grateful to FAPESP and the University of Goiás for supporting this work.

References

Barzilay, R. and Lapata, M. (2005). Modeling local coherence: An Entity-based Approach. In the Proceedings of the 43rd Annual Meeting on Association for Computational Linguistics, p. 141-148, Stroudsburg, PA, USA.

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.

Burstein, J., Tetreault, J. and Andreyev, S. (2010). Using entity-based features to model coherence in student essays. Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, HLT ’10, p. 681–684, Stroudsburg, PA, USA.

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 Criacao de Novos Sumarios 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.

Eisner, M. and Charniak, E. (2011). Extending the entity grid with entity-specific features. In the Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies: short papers - Volume 2, HLT ’11, p. 125–129, Stroudsburg, PA, USA.

Elsner, M., Austerweil, J. and Charniak, E. (2007). A unified local and global model for discourse coherence. Human Language Technologies 2007: The Conference of the North American Chapter of the Association for Computational Linguistics. Rochester, New York, USA.

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 (COLING-2014), p. 940-949, Dublin, Ireland.

Filippova, K. and Strube, M. (2007). Extending the entity-grid coherence model to semantically related entities. In the Proceedings of the Eleventh European Workshop on Natural Language Generation, ENLG ’07, p. 139–142, Stroudsburg, PA, USA.

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.

Freitas, A. R. P. (2013). Análise automática de coerência usando o modelo grade de entidades para o português. Dissertação (Mestrado), Universidade Estadual de Maringá – Centro de Tecnologia, Departamento de Informática, Programa de Pós-Graduação em Ciência da Computação.

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.

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

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.

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.

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.

Webber, B. (2004). D-ltag: extending lexicalized tag to discourse. Cognitive Science, vol. 28, n. 5, p. 751-779.