Building a Language Model for Local Coherence in Multi-document Summaries using a Discourse-enriched Entity-based Model

Maria Lucía R. Castro Jorge, Márcio S. Dias, Thiago A. S. Pardo
Interinstitutional Center for Computational Linguistics (NILC)
Institute of Mathematical and Computer Sciences, University of São Paulo
{mluciacj, marciosd, taspardo}@icmc.usp.br

***Abstract*— Local Coherence is a very important aspect in multi-document summarization, since good summaries not only condense the most relevant information, but also present it in a well-organized structure. One of the most investigated models for local coherence is the Entity-based model, which has been successfully used, once it facilitates the computational approach for coherence measurement. Particularly, this model was used for the evaluation of local coherence in multi-document summaries, achieving promising results. In order to improve the potential of the Entity-based model, we propose the creation of a language model for multi-document summaries that integrates the Entity-based model with discourse knowledge, mainly from Cross-document Structure Theory. Our results show that this type of information enriches the Entity-based Model by capturing other phenomena that are inherent to multi-document summaries, such as redundancy and complementarity, which improves the performance of the original model.**

***Keywords*—multi-document summarization; entity-based model; discourse models**

I. Introduction

Textual Coherence is a characteristic of texts that denotes the capability of a text being understandable to the reader in a communication scenario [5, 27]. Coherence may be categorized as Local or Global [28]. Local coherence refers to parts of the text, such as sentences or sentence sequences, and it occurs when there is related content and good usage of linguistic elements among these parts in order to make sense for communication. Global coherence refers to the message of the text as a whole and it depends of local coherence.

Several recent works have focused on local coherence [4, 6, 10, 14, 15] in text generation and evaluation tasks, since it may be analyzed in shorter textual portions and is more computationally feasible. One of the most relevant models for local coherence is the Entity-based model [6], which supposedly captures aspects of local coherence by analyzing entity transitions among sentences.

In Multi-document Summarization (MDS), local coherence is as important as informativity. A summary must not only contain relevant information but also present it in a coherent way, readable and understandable to the users. A "coherence model" (following the language modeling tradition in the Machine Translation area) for summarization should, therefore, "score" summaries by their coherence, capturing the phenomena that may affect coherence. As stated in [3], some of these multi-document phenomena (that must be avoided or properly treated in summaries) are redundancy, complementarity and contradiction, which may happen because information from different source texts may be used to compose the summary. In several MDS approaches [1, 3, 20, 23, 25], the CST (Cross-document Structure Theory) [22] discourse model has been successfully applied to deal with these issues.

In this paper, we propose a language modeling approach that integrates discourse knowledge from CST and the Entity-based model in order to "judge" coherence in multi-document summaries. In our experiments, a combination of syntactical and discourse features are extracted from multi-document summaries of news texts and given as input for machine learning (with SVMlight [7]), following the same methodology of [6]. Results showed that CST improved the performance of the Entity-based model at classifying coherent and non-coherent summaries. We also conducted other experiments incorporating information from RST (Rhetorical Structure Theory) [16], which is a widely used discourse theory, mainly for single document summarization [17, 18, 19, 21, 24]. Results were similar to the experiments using CST, confirming that discourse information is useful for modeling local coherence properties.

This paper is organized as follows: in the next section, we briefly introduce the Entity-based model [6], the discourse models CST and RST, and the related work on local coherence; in Section 3, our proposals of integrated entity-discourse-based models are presented; in Section 4, we report our experimental methodology and show the obtained results; finally, in Section 5, we present some final remarks.

II. Related Work

[6] developed the Entity-based model, which explores entity transitions among sentences in order to obtain patterns for (locally) coherent texts. The idea behind this model is that coherent texts present regular entity transition patterns. The method consists in the creation of a bi-dimensional grid (matrix), also called entity grid, where lines represent sentences and columns represent entities, for each text. An

978-1-4799-5618-0/14 $31.00 © 2014 IEEE
DOI 10.1109/BRACIS.2014.19

example of this grid is illustrated in Fig. 1, where it is shown part of a text and its correspondent entity grid.

S_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]$_O$.
S_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]$_O$.
S_3 [The case]$_S$ revolves around [evidence]$_O$ of [Microsoft]$_S$ aggressively pressuring [Netscape]$_O$ into merging [browser software]$_O$.
…

	Department	Trial	Microsoft	Evidence	Competitors	Markers	Products	Brands	Case	Netscape	Software
1	S	O	S	X	O	–	–	–	–	–	–
2	–	–	O	–	–	X	S	O	–	–	–
3	–	–	S	O	–	–	–	–	S	O	O

Fig. 1: Example of entity grid reproduced from [6]

In this grid, which makes use of syntactical information, we indicate for each entity if it happens as a subject (S), an object (O) or as any other syntactical function (X) in each sentence. The hyphen indicates that the entity did not happen in the sentence. When one entity happens more than once in a sentence, its most prominent function is adopted (it is considered that subjects are more prominent than objects, which are more prominent than other functions). If syntactical information is not available for building the grid, one may simply indicate if the entity happens in a sentence or not, as suggested in [6]. The authors of the model also suggest two more variations, using co-reference and salience information. The use of co-reference information allows to join entities in the grid (as it happens for "Microsoft" in the above grid, which represents the entities "Microsoft Corp.", "the company", and "Microsoft") in order to let the grid more consistent. When no co-reference information is available, each different noun may be an entity in the grid. By salience, the authors refers to the frequency of the entities in the text, suggesting that different grids may be built for such entities (e.g., one grid for the entities that happen only once and another grid for the more frequent/salient ones).

Given a grid, it is possible to see the entity transitions and to compute their probabilities. Considering the "Evidence" entity, one may see in the grid that there is a transition from X to - and another one from - to O in the 3 sentences in the example. In [6], transitions of size 2 are suggested, but others might be used as well. Transition probabilities are computed as the ratio between the frequency of a specific transition in the grid and the total number of transitions. For instance, the transition from - to O (usually represented as [- O]) has a probability of 0.18 (or 18%), since it happens 4 times out of the 22 possible transitions in the grid. The transition probabilities are then used to compose a feature vector for each text. Such vectors become training instances for a machine learning process using the SVM^{light} package [7]. If provided with instances of coherent and incoherent texts, the machine learning may learn to distinguish such texts, as showed in [6]. These authors produced several models considering the usage or not of salience, syntactical and co-referential information and showed that the best model for distinguishing coherence in multi-document summaries was the one that used only syntactical and salience information, obtaining 83.8% of accuracy.

Some other works investigated new approaches using the Entity-based model. For instance, [8] conducted experiments for textual ordering for news texts written in German. The authors also clustered entities according to the semantic relations among them, by using the WikiRelate API [9]. Their best model (that used co-referential and salience information) obtained 75% of accuracy. [15] investigated the applicability of the Entity-based model in the evaluation of coherence for scientific summaries written in Portuguese and obtained 74.4% of accuracy by using syntactical and salience information.

[11] was one of the inspiring works for this paper. The authors used in their model the stemmed forms of the open class words (instead of entities), combined with discourse information, assuming that local coherence favors certain types of transitions among discourse relations. The relations used in this work were the ones included in the Penn Discourse Treebank [13]. The authors proposed the Discourse Role Matrix, which is composed of sentences (rows) and selected terms (columns), with discourse relations used over their arguments. For example, Fig. 2 shows an example of this grid (b) for part of a sample text (a).

(S_1) Japan normally depends heavily on the Highland Valley and Cananea mines as well as the Bougainville mine in Papua New Guinea. (S_2) 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)

Fig. 2: Part of a text and its grid reproduced from [11]

In this grid, a cell $C_{Ti,Sj}$ contains the set of the discourse roles of the term T_i that appears in sentence S_j. For example, the term "depend" from 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 roles (as in $C_{copper,S2}$, as "copper" in S_2 participates in two relations). The authors obtained results over 90% for a corpus of news texts in English.

In our proposal, we also integrate the syntactical information provided by the entity grid with discourse information for multi-document summaries, in order to improve the performance of the Entity-based model. Our proposal differs from the previous works in the fact that we also use information that models the multi-document phenomena. In what follows, we briefly introduce the discourse models we use in this paper.

Rhetorical Structure Theory (RST) was proposed by [16] as a theory for textual organization. It identifies discourse segments in a text (usually expressed as clauses or sentences) and establishes a set of discourse relations among them. The theory is based on the presupposition that these segments are organized in a text by the author in order to express a message to the reader. Each segment is also classified as nucleus (N) or satellite (S): the nuclei or nuclear segments contain the most important pieces of information in the relations and are considered more relevant than the satellites; the satellites, on the other hand, present additional information that helps the reader in the interpretation of the nuclei. According to [16], coherent texts should have a complete RST structure (which usually adopts the form of a tree). As expected, if one may not find any relation between two sentences, the text is not fully coherent, therefore. In this paper, we suppose (as [12] also claims) that there are patterns of relation occurrence that may help the coherence judgment.

Differently from RST (which is intended for single document analysis), the Cross-document Structure Theory (CST) was created by [22] as a model for structuring multiple texts on the same topic, establishing discourse relations among textual segments of different texts. These relations model the multi-document phenomena of redundancy, complementarity, and contradiction, mainly. According to the theory, relations may be established at any level of segmentation (words, phrases, sentences, paragraphs, or even entire documents). [26] formalized better the theory by refining and organizing CST relations in a hierarchy of categories according to their meaning and the multi-document phenomena that they indicate. In this paper, we expect that such relations may properly handle the multi-document phenomena in multi-document summaries and help improving the coherence modeling.

III. Discourse Knowledge and Entity-based Model

For our proposal, besides the syntactical transitions, we also incorporate the discourse information that occurs among sequences of sentences in a multi-document summary. We want to make use of the distribution of discourse knowledge and syntactical information of entities in coherent summaries. In order to illustrate this idea, we show in Fig. 3 an example of a multi-document summary (in Portuguese) extracted from the CSTNews corpus [26] with syntactical information given by the Palavras parser [2] and CST relations that were manually identified.

1. Ao menos 17 [pessoas]$_S$ morreram após a [queda]$_X$ de um [avião]$_X$ de passageiros na [República Democrática do Congo]$_X$.
2. As [vítimas]$_S$ do acidente foram 14 [passageiros]$_X$ e três [membros]$_X$ da [tripulação]$_X$.
3. Segundo [fontes]$_X$ aeroportuárias, os [membros]$_S$ da [tripulação]$_S$ eram de [nacionalidade]$_X$ russa.
4. O [avião]$_S$ explodiu e se incendiou, acrescentou o porta-voz da ONU em Kinshasa, Jean-Tobias Okala.
5. [Ele]$_S$ havia saído da [cidade]$_X$ mineira de Lugushwa em direção a Bukavu, numa [distância]$_X$ de 130 [quilômetros]$_X$.

Elaboration

Elaboration

Follow-up

Fig. 3: Example of discourse relations in a multi-document summary

In this example, there are three discourse relations occurring among the sentences. For instance, there is an elaboration relation between sentences 1 and 2, which means that sentence 2 gives complementary information about the main facts in sentence 1. The same occurs between sentences 2 and 3. In the case of sentences 4 and 5, there is a Follow-up relation, which indicates that facts in sentence 4 happened after the facts in sentence 5. From this example, two grids may be generated: one representing the traditional entity grid and the other representing the relations among sentences in the summary. These grids are illustrated in Fig. 4. Notice that we do not use co-referential information (since, to the best of our knowledge, there is no widely available co-reference resolution system for Portuguese).

	pessoas	queda	avião	R.D.Congo	vítimas	passageiros	membros	tripulação	fontes	nacionalidade	ele	cidade	distância	kilometros
S_1	S	X	X	X	–	–	–	–	–	–	–	–	–	–
S_2	–	–	–	–	S	X	X	X	–	–	–	–	–	–
S_3	–	–	–	–	–	–	S	S	X	X	–	–	–	–
S_4	–	–	S	–	–	–	–	–	–	–	–	–	–	–
S_5	–	–	–	–	–	–	–	–	–	–	S	X	X	X

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

Fig. 4: Grids for entities and for relations

In the discourse grid, each cell represents the CST relations that occur for a pair of sentences in the summary. When there is no relation, the cell is filled with '–'.

From these two grids, it is possible to produce new features that incorporate both information. For instance, it may be observed that the Elaboration relation between sentences S1 and S2 co-occurs with the syntactical transitions of the entities in the same pair of sentences, which are of the types [S–], [X–], [–S], [–X] and [– –]. In this case, for each cell of the entity grid, we count the number of times each syntactical transition occurs together with the relation marked for the corresponding pair of sentences. Then these values are divided by the total number of transitions of length two in the entity grid. The idea behind this coding is that CST relations occur with certain patterns in coherent texts, for example, contiguous sentences in a coherent summary tend to be complementary because they should expose and detail a certain fact or topic. For the text we are analyzing, the probability for each transition would be as illustrated in the feature vector in Fig. 5.

S – Elaboration	X – Elaboration	– S Elaboration	– X Elaboration	– – Elaboration	X S Elaboration	S – Follow-up	X – Follow-up	– S Follow-up
.03	.07	.01	.08	.25	.03	.01	0	.01

– X Follow-up	– – Follow-up	X S Follow-up	S – –	X – –	– S –	– X –	– – –	X S –
.05	.16	0	.03	.03	.01	0	.16	0

Fig. 5: Feature vector combining information from the grids

For illustration purposes, in the example we have only included as features the transitions that appear in the grids in Fig. 4. In a real scenario, the number of features would be 224, which is the result of multiplying 16 (number of possible combinations of syntactical patterns of the entity-based model) * 14 (total number of CST relations that are considered). In this example, the probability values are the result of dividing the corresponding count for each pattern by 56, which is the total number of transitions for the entity grid in Fig. 4. For instance, for the pattern [S – Elaboration], the probability value 0.03 was obtained by dividing the number of times this pattern appeared in the text (2) by 56.

In the real scenario, the number of features is too big, which may generate sparse data and decrease the performance in the classification task. In order to ameliorate this situation, we modified the model so that the discourse information considered are not the CST relationships, but the CST categories that represent these relations (in the CST hierarchy proposed in [26]). The advantage of using categories instead of using the complete set of CST relations is that there is a fewer number of categories that condense the discourse information of the CST relations. In this work, five categories from the hierarchy were considered, which are: redundancy, complement, contradiction, authorship, and style categories. We chose these five categories since they enclose the main types of relations, according to their meaning and purpose. Considering these categories instead of the CST relations, the total number of features reduces to 80, which is the result of multiplying 16 by 5. The new CST grid, considering CST categories for the example above, and the correspondent feature vector are shown in Figs. 6 and 7, respectively.

	S_1	S_2	S_3	S_4	S_5
S_1	–	Complement	–	–	–
S_2	–	–	Complement	–	–
S_3	–	–	–	–	–
S_4	–	–	–	–	Complement
S_5	–	–	–	–	–

Fig. 6: CST categories grid

S – Complement	X – Complement	– S Complement	– X Complement	– – Complement	X S Complement	S – –	X – –	– S –	– X –	– – –	X S –
.03	.07	.01	.09	.25	.03	.03	.03	.01	0	.16	0

Fig.7: Feature vector with CST categories

Again, for illustration purposes, the features shown in Fig. 7 are just the ones that appear in the example. As it may be observed in the example, the dimensionality of the feature vector is reduced to half the size of the feature vector with the whole set of CST relations. However, 80 features may still be a high number of features, so we propose another model in which only Boolean features are considered. In other words, it will only be considered when a CST relation is present (indicated by '1') or absent (indicated by '0') between two sentences, which will reduce the number of features to 32 (16 * 2). In this scenario, the CST grid and feature vector for our example will be as shown in Figs. 8 and 9, respectively.

	S_1	S_2	S_3	S_4	S_5
S_1	0	1	0	0	0
S_2	0	0	1	0	0
S_3	0	0	0	0	0
S_4	0	0	0	0	1
S_5	0	0	0	0	0

Fig. 8: Boolean CST grid

S – 1	X – 1	– S 1	– X 1	– – 1	X S 1	S – 0	X – 0	– S 0	– X 0	– – 0	X S 0
.03	.07	.01	.08	.25	.03	.03	.03	.01	0	.16	0

Fig. 9: Feature vector with boolean CST

An important characteristic of multi-document summaries is that sentences that compose them may come from different documents, but may also come from the same document. Sentences that come from the same document may also have discourse relations that are not represented by CST relations. To tackle this issue, we complement the information in the boolean CST grid with boolean information of single document discourse relations, provided by RST. In other words, we consider if an RST relation is present or not between two contiguous sentences in the summary. The result of this combination is a grid filled with values 1 or 0 when a discourse relation (single or multi-document) is present or not between two contiguous sentences.

From the three previous proposals, only the boolean grid is complemented with RST information. In the case of the CST categories grid, complementing it with RST information would require to classify the RST relations into the same categories (and maybe other new categories) of the CST taxonomy, which is a task that we still have not accomplished. In the case of the CST relations grid, including RST relations would generate a feature vector of higher dimensionality, which may impair the performance of the method.

IV. Experiments and Results

Our experiments were conducted over the CSTNews corpus [26], which is composed of 50 clusters of news texts written in Brazilian Portuguese, with the texts of each cluster on a specific topic. Each cluster contains 2 or 3 texts, their CST annotation, RST annotation and the corresponding manually built multi-document summary, which is an extract. Each text and each summary (extract) in the corpus are also automatically parsed with the Palavras parser, which is the state of the art syntactical parser for Portuguese.

For each multi-document summary (in a total of 50, being 1 per cluster), the following grids were built: the traditional entity grid (the nuclei of noun phrases were considered entities, according to the Palavras automatic analyses), without considering salience or co-referential information; the CST relation grid; the CST category grid; the boolean CST grid; and the boolean CST grid complemented with RST information. The five grids give rise to the feature vectors for the corresponding summary. The probabilities in the feature vectors were computed considering entity transitions of size 2, as suggested by [6].

After this, we followed a methodology similar to [6] for building a model of coherence for multi-document summaries. For each multi-document summary (coherent summary), 20 sentence-permuted summaries (incoherent summaries) were generated. For each permuted summary, five feature vectors result from the use of the corresponding entity grid with the discourse grids.

In total, 1000 pairs of feature vectors (50*20) were produced. They compose the instances for the learning process with SVM^{light}. 10-fold cross-validation was used to train and test the models for each type of feature vector. Accuracy was computed for each fold, and was calculated by dividing the number of times the model ranked correctly the pair of the original text and its permuted version by the total of pairs. In TABLE I, we show the average accuracy for the traditional entity grid and the combination of the entity grid with the discourse grids.

TABLE I. Results for the traditional and discourse grids

Entity-based Model only	Entity-based Model with CST relations	Entity-based Model with CST categories	Entity-based Model with boolean CST	Entity-based Model with boolean CST and RST
73.65%	68.34%	73.22%	81.39%	79.03%

We also performed an experiment in which the boolean grid is filled considering only RST information. In other words, the grid is filled with 1 or 0 if there is an RST relation or not between the two corresponding sentences. The results of this experiment showed 80% of accuracy, very close to the accuracy value of the Entity-based model with boolean CST information.

In order to confirm the statistical significance of these results, we applied the t-test for each pair of methods and it showed a statistical significant difference (with 95% of confidence) for every pair of methods except for Entity-based model with boolean CST information and Entity-based model with boolean CST and RST information. There was also no statistical difference between these two methods and the Entity-based model with boolean RST information only.

As expected, including detailed CST information (all CST relations) degrades the performance of the Entity-based model. In this case, the total number of instances (considering all pairs) used for training is small for the high dimensional feature vector generated from the combination of the Entity-based model and CST information such as CST categories and relations, which harms the performance. On the other hand, boolean CST information combined with the Entity-based model is enough to outperform the original Entity-based model, showing that discourse information enriches and complements syntactical information patterns.

It is also interesting that using only boolean CST information with the Entity-based model had a higher level of accuracy than the model that combines RST and CST. This result may be explained by the fact that sentences that compose the summary tend to come from different sources and not from only one source, which means that multi-document discourse relations tend to appear more frequently among sentences than single document relations.

The model of [11], which integrated discourse relations in the Entity-based model, obtained a maximum gain of 6.13% in accuracy, compared to the Entity-based model of [6]. Our best model obtained a gain of 7.74% in accuracy compared to the Entity-based model from [6]. This shows that the use of discourse relations, even though in a different way of the Penn Discourse Treebank, and applied in texts of different nature (source texts and multi-document summaries), produced good results in their contexts and may help in the task of evaluating local coherence.

V. Final Remarks

In this paper, we have investigated how well-known discourse models as RST and CST may help in local coherence modeling. We show that such knowledge do improve the results for multi-document summaries.

As current summary evaluations focus on informativity, such coherence models may be useful for distinguishing coherence between summaries produced by different systems or for guiding summarizers in the content selection process in order to build better summaries.

Future work includes the investigation of other discourse phenomena for coherence modeling, as subtopics and informative aspects in the summaries.

Acknowledgements

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

REFERENCES

[1] S.D. Afantenos, “Reflections on the Task of Content Determination in the Context of Multi-Document Summarization of Evolving Events”, Advances on Natural Language Processing. Borovets/Bulgaria, 2007.

[2] E. Bick, “The Parsing System “Palavras”: Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework.” Phil. Thesis. Aarhus University. Aarhus, Denmark: Aarhus University Press, November 2000.

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

[4] B. Grosz, K.J. Aravind, W. Scott, “Centering: A framework for modeling the local coherence of discourse.” Computational Linguistics, vol. 21, pp. 203-225, 1995.

[5] P. Foltz, W. Kintsch, T.K. Landauer, “The Measurement of textual coherence using latent semantic analysis.” Discourse Processes, vol. 25, n. 2-3, pp. 285-307, 1998.

[6] R. Barzilay, M. Lapata, “Modeling local coherence: An entity-based approach.” Computational Linguistics, vol. 34, pp. 1-34, 2008.

[7] T. Joachims, “Optimizing search engines using clickthrough data.” Proceedings of the eighth ACM SIGKDD international conference on Knowledge discovery and data mining, pp. 133–142. New York, NY, USA, 2002.

[8] K. Filippova, M. Strube, “Extending the entity-grid coherence model to semantically related entities.” Proceedings of the Eleventh European Workshop on Natural Language Generations, pp. 139-142, 2007.

[9] M. Strube, S.P. Ponzetto, “Wikirelate! computing semantic relatedness using wikipedia.” Proceedings of the 21st National Conference on Artificial Intelligence, 2006.

[10] J. Burstein, J. Tetreault, S. Andreyev, “Using entity-based features to model coherence in student essays.” Proceedings of the Human Language Technologies: The 2010 Annual Conference of the North American Chapter of the Association for Computational Linguistics, pp. 681-684, 2010.

[11] Z. Lin, H. Ng, M.Y Kan, “Automatically evaluating text coherence using discourse relations.” Proceedings of the 49th Annual Meeting of the Association for Computational Linguistics: Human Language Technologies. vol 1. pp. 997-1006. Stroudsburg, PA, USA, 2011.

[12] B. Webber, “D-ltag: extending lexicalized tag to discourse.” Cognitive Science, pp. 751-779, 2004.

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

[14] R. Iida, T. Tokunaga, “A metric for evaluating discourse coherence based on coreference resolution.” Proceedings of COLING 2012: Posters, Mumbai, India, 2012.

[15] P. Freitas, V.D. Feltrim, “Análise Automática de Coerência Usando o Modelo Grade de Entidades para o Português.” Proceedings of the IX Brazilian Symposium in Information and Human Language Technology, pp. 69-78. Fortaleza, CE, Brazil, 2013.

[16] W.C. Mann, S.A. Thompson, “Rhetorical Structure Theory: A Theory of Text Organization.” Technical Report ISI/RS-87-190, 1987.

[17] D. Marcu, “The Theory and Practice of Discourse Parsing and Summarization.” The MIT Press. Cambridge, Massachusetts, 2000.

[18] M. O’Donnell, “Variable-Length On-Line Document Generation.” Proceedings of the 6th European Workshop on Natural Language Generation, 1997.

[19] K. Ono, K. Sumita, S. Miike, “Abstract generation based on rhetorical structure extraction.” Proceedings of the International Conference on Computational Linguistics, Japão, 1994.

[20] J.C Otterbacher, D. R. Radev, A. Luo, A. “Revisions that improve cohesion in multi-document summaries: a preliminary study.” Proceedings of the Workshop on Automatic Summarization, pp. 27-36, USA, 2002.

[21] T.A.S Pardo, L.H.M. Rino, “DMSumm: Review and Assessment.” In E. Ranchhod and N. J. Mamede (eds.), Advances in Natural Language Processing, pp. 263-273 (Lecture Notes in Artificial Intelligence 2389). Springer-Verlag/Germany, 2002.

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

[23] D.R. Radev, H. Jing, M. Budzikowska, “Centroid-based summarization of multiple documents: sentence extraction, utility-based evaluation and user studies.” Proceedings of the ANLP/NAACL Workshop, pp. 21-29, 2000.

[24] V.R. Uzêda, T.A.S. Pardo, M.G.V. Nunes, “A Comprehensive Comparative Evaluation of RST-Based Summarization Methods.” ACM Transactions on Speech and Language Processing, vol. 6, n. 4, pp. 1-20. 2010.

[25] Z. Zhang, S.B. Goldenshon, D.R. Radev, “Towards CST-Enhanced Sumarization.” Proceedings of the 18th National Conference on Artificial Intelligence (AAAI-2002), Canadá, 2002.

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

[27] I.V. Koch, L.C. Travaglia, “A Coerência Textual.” Editora Contexto, São Paulo, 2002.

[28] T.V. Dijk, W. Kintsch, “Strategics in discourse comprehension.” Academic Press. New York, 1983.