

The Coreference Annotation of the CSTNews Corpus

Thiago Alexandre Salgueiro Pardo¹, Jorge Baptista², Magali Sanches Duran¹,
Maria das Graças Volpe Nunes¹, Fernando Antônio Asevedo Nóbrega¹, Sandra
Maria Aluísio¹, Ariani Di Felippo³, Eloize Rossi Marques Seno⁴, Raphael
Rocha da Silva¹, Rafael Torres Anchiêta¹, Henrico Bertini Brum¹, Márcio de
Souza Dias⁵, Rafael de Sousa Oliveira Martins¹, Erick Galani Maziero¹,
Jackson Wilke da Cruz Souza³, and Francielle Alves Vargas¹

¹ Universidade de São Paulo

² Universidade do Algarve

³ Universidade Federal de São Carlos

⁴ Instituto Federal de São Paulo

⁵ Universidade Federal de Goiás

tasparado@icmc.usp.br, {jorge.manuel.baptista, magali.duran}@gmail.com,
gracan@icmc.usp.br, fernandoasevedo@gmail.com, sandra@icmc.usp.br,
{arianidf, eloizeseno, raphaelsilva2500}@gmail.com, rta@usp.br, {henrico.
brum, marciosouzadias, martins.rso, egmaziero, jackcruzsouza}@gmail.com,
francielleavargas@hotmail.com

Abstract. We report in this paper the coreference annotation process of the CSTNews corpus as part of a collective task of the IberEval 2017 conference. The annotated corpus is composed of 140 news texts written in Brazilian Portuguese language and counts with several annotation layers, including annotations in the morphosyntax/syntax, semantics, and discourse levels. The annotation, focused on nominal references, was conducted in a semi-automatic way by five teams, achieving satisfactory annotation agreement results.

1 Introduction

Coreference resolution is the task of finding linguistic expressions in a text that refer to the same entity [1]. As an illustration of coreference occurrence, we show below a short text with some coreferent elements in bold. In this short text, the referring expressions “a passenger plane”, “the airplane” and “it” refer to the same entity and form a “coreference chain”. It is interesting to notice that the coreference resolution task includes pronominal anaphora resolution, which is part of the problem.

*At least 17 people died after the crash of **a passenger plane** in the Democratic Republic of Congo. According to an ONU spokeswoman, **the airplane** was trying to land in the Bukavu airport in the midst of a storm. **It** failed to reach the runway and fell in a forest 15 kilometers away from the airport.*

Coreference resolution and its subtasks have been investigated for a long time in the Natural Language Processing (NLP) area. There are approaches based on heuristics [2, 3], machine learning [4, 5], and discourse theories, as Centering [6] and Veins Theory [7]. The task has also been the focus of a shared task in CoNLL [8]. Besides its history, the task is still a challenge, as it brings together the difficulties of automatically dealing with semantics and discourse. Such linguistic analysis levels usually require deep linguistic processing capabilities from the machines, which, in turn, demand automatic text interpretation techniques.

Coreference is a very important information for NLP systems. For instance, it is essential for summarization systems to produce coherent and cohesive summaries, allowing the systems to properly “glue” together text passages; it is useful to track, on the web, entities of interest; it may be necessary for information extraction about entities of interest and for performing the related inference; it may help simplifying texts, allowing changing pronominal anaphoras by nominal antecedents, as certain anaphoric mentions cause difficulties for people with cognitive disabilities; and so on. Therefore, research efforts in such task are very relevant. Although some coreference annotated corpora and automatic coreference annotation softwares are available for English (the interested reader may refer to the shared task in CoNLL), similar initiatives for other languages are still rare, including Portuguese, which is the focus of this paper.

We report here the coreference annotation process of a corpus of news texts written in Brazilian Portuguese language - the CSTNews corpus [9]. The annotation, focused on nominal references, was conducted in a semi-automatic way by 5 annotation teams as part of a collective task of the IberEval 2017 conference, achieving satisfactory annotation agreement results (considering the difficulty of the task).

In what follows (Section 2), we present some basic concepts regarding coreference. In Section 3, we introduce the corpus we annotated. Section 4 reports the annotation process and the achieved results, describing some challenges of the task. Some final remarks are presented in Section 5.

2 Coreferences

Detecting coreferent elements and composing coreference chains is a task that demands sophisticated linguistic knowledge. Coreference mainly happens at the semantics/discourse interface, as it requires the identification of the meaning of the elements (i.e., to what they refer to) in the same sentence or across sentences in a text. There are also initiatives for detecting coreference relations across different texts, which is relevant for multi-document processing purposes.

As presented in [1], the entities in a text are evoked by “referring expressions”. The element that refers to a previous one in the text (e.g., “the airplane” in the sample text in the Introduction section) is the “anaphoric element”, while the element that is referred to (“a passenger plane”) is called the “antecedent” (or “the referent”, according to some authors). Such elements are said to corefer, that is, they refer to the same extra-linguistic entity. When we track and store

all the references to an entity, we perform “coreference resolution”, and the set of references is a “coreference chain”.

Referring expressions may happen in several forms in a text, usually syntactically realized as noun phrases. For instance, we may use pronouns (e.g., “it”), common nouns (“airplane”) and proper nouns (“William Shakespeare”), optionally presenting determiners, pre- and/or post-modifiers (e.g., “the beautiful girl”) and having high size variance (e.g., “the beautiful and charming girl that was looking at me”). In [10], the authors discuss such variations and the difficulties that they bring. The authors comment that the reference may happen *directly*, when the same noun is used to refer to another one, as in the text passage “The **letter** was signed yesterday. In the **letter**, the scientists argue that...”, or in an *indirect* way, when different terms are used, as in “The **letter** was signed yesterday. In the **text**, the scientists argue that...”. In such cases, the reference to the original expression may be recovered by accessing linguistic and world knowledge, as synonymy, hypernymy/hyponymy, and meronymy/holonymy relations, several types of pronouns, acronyms and abbreviations, and verb nominalizations, among several others. We suggest consulting the work of [10] for the interested reader.

To the best of our knowledge, the only manually annotated corpus in Brazilian Portuguese that is specifically focused on coreference chains is the Summ-it corpus [11], which is generally used for training and testing systems. There are other corpora annotated with named entities, which might be used for such end, but that were not specifically built to address the task of coreference resolution. As argued in [12], “the Portuguese coreference resolution area is at an early stage of development”, but some initiatives exist. The CORP system¹ [13, 14], for instance, has been used by the research community.

In such scenario, producing coreference annotated corpora and developing and/or improving coreference resolution systems for Portuguese are key issues to be pursued. The annotation effort reported in this paper is a step towards new advances in the NLP area for Portuguese. In what follows, we introduce the corpus that was annotated.

3 The CSTNews Corpus

The CSTNews corpus [9] was originally developed for multi-document summarization purposes during the SUCINTO project². The corpus includes 140 news texts written in Brazilian Portuguese, from some main online news agencies in Brazil, as *Folha de São Paulo*, *Estadão*, *Jornal do Brasil*, *Gazeta do Povo*, and *O Globo*. The texts are grouped in 50 clusters (each cluster has 2 or 3 texts), and the texts of a cluster are on the same topic. The texts are about Economy, Politics, Sports, Science, Daily News, World News, and Financial subjects. Several summaries are associated to each cluster, including single and multi-document

¹ <http://ontolp.inf.pucrs.br/corref/>

² <http://www.icmc.usp.br/~taspardo/sucinto/>

summaries, extractive and abstractive summaries, and automatically and manually produced summaries, as the corpus was originally intended for application in the summarization area.

In the years following its creation, the corpus received several linguistic annotation layers (mainly of discourse nature), according to the uses it had. The corpus currently includes:

- manual multi-document discourse annotation, according to the Cross-document Structure Theory (CST) [15], which was the first annotation in the corpus and gave origin to its name;
- manual single document discourse annotation, according to the Rhetorical Structure Theory (RST) [16];
- manual subtopic segmentation, following the proposal of [17];
- manual informative aspect identification, according to the guidelines proposed at the Guided Summarization task³ of the TAC conference [18];
- manual identification and normalization of temporal expressions, following the proposal of [19];
- manual word sense disambiguation of verbs and (the most frequent) common nouns, using Princeton WordNet as sense repository [20];
- manual text-summary alignment, indicating which sentences from the source texts gave origin to the sentences of the summaries and through which rewriting operations;
- automatic morphosyntax and syntax annotation by the PALAVRAS parser [21].

Except for the last one, which was automatic, all of these annotations were manually carried out in systematic and controlled way. Satisfactory annotation agreement results were obtained (when applicable), which allows to infer that the data is reliable.

With such annotations, the corpus has subsidized several research efforts, including traditional multi-document summarization [22–24], the more recent update summarization [25], summary coherence evaluation [26], the study of human behavior for summary production [27], the development of discourse parsers [28, 29] and application on information extraction [30], phrase generalization [31], and theoretical studies on some discourse aspects [32–34], among several others that may be seen in the SUCINTO project website.

The coreference annotation reported here constitutes a new annotation layer in the CSTNews corpus. The annotation process is described in the following section.

4 The Annotation

The coreference annotation of the CSTNews corpus was carried out as a collective task, named “Collective Elaboration of a Coreference Annotated Corpus

³ <https://tac.nist.gov//2010/Summarization/Guided-Summ.2010.guidelines.html>

for Portuguese Texts”⁴, of the IberEval 2017 (Evaluation of Human Language Technologies for Iberian Languages) conference⁵.

In the collective task, annotation teams with at least 3 members should submit their corpus for annotation. If selected to participate, the members of the team would receive pre-processed texts to annotate. Part of the texts was from the corpus submitted by the team (4 of them were used to compute the annotation agreement of the team), and another part was from the corpora submitted by other teams. Each member should then individually annotate the coreference chains in his/her texts. Only nominal referring expressions were intended to be annotated.

As said above, the texts were presented in a pre-processed form. They were automatically annotated by the CORP coreference resolution system [13,14], which performs (in some cases, with the aid of some other tools) part of speech tagging, shallow syntactic parsing, and construction of coreference chains by using syntactic heuristics (as string matching, copular constructions, and juxtaposition of linguistic expressions in the text) and semantic knowledge (as synonymy and hyponymy relations) obtained from the Onto.PT resource [35]. The collective annotation task consisted, therefore, in reviewing and eventually correcting the automatic coreference annotation that was presented, which was done with the aid of another tool, the CorrefVisual, which is a graphical interface that allows to visualize the text and the available preprocessed coreference chains, and to edit the chains.

Figure 1 shows the CorrefVisual interface with an annotated text loaded. The text is in Portuguese, as it is in the corpus. It is about an accident in an airport, where an airplane crashed into a building. One may see the text at the left of the panel, the coreference chains in the middle, and the remaining referring expressions that did not form any chain, called “unique mentions”, on the right side. Above these unique mentions, there is an auxiliary panel to help dealing with the chains. In the interface, each chain is associated to a different color, and, every time that an element is selected, its occurrence in the text is highlighted in the corresponding color.

Correcting a chain consisted, therefore, in moving the referring expressions among the windows in the tool, e.g., removing an expression from a chain and adding it to another one, incorporating a unique mention to some existing chain, creating new chains, etc. If necessary, the auxiliary panel would help in grouping and moving the elements across the windows. The tool also allowed to edit the referring expressions, by adding or removing words next to it (to the left or to the right of the expression). This is an important step, as such expressions were automatically detected and some errors probably occurred (in fact, during the annotation, we have noticed that such errors were very frequent). Unfortunately, the tool does not allow to use new expressions that were not previously detected by the tool itself. The edition of the target expressions had also been explicitly and highly discouraged by the organizers of the collective annotation task. Such

⁴ <http://ontolp.inf.pucrs.br/corref/ibereval2017/>

⁵ <http://nlp.uned.es/IberEval-2017/>

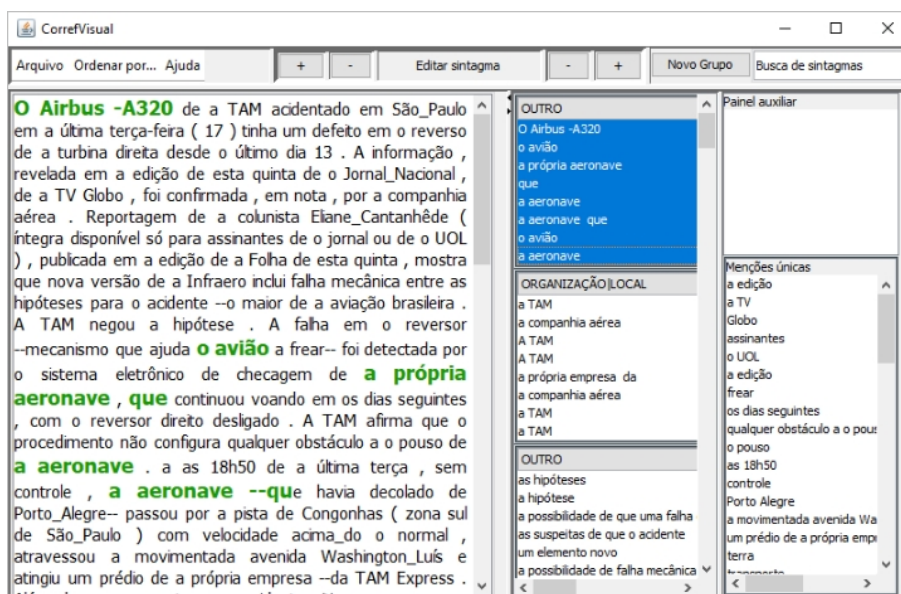


Fig. 1. CorreVisual interface

decision benefits the annotation agreement results, but is also based on the idea that we should make use of the available state of the art preprocessing tools, with their current limitations and potentialities.

Finally, one may also notice in the interface that a semantic category should be assigned to each chain, indicating the type of the entities of the chain. In the interface, the category appears above each chain, in capital letters. The available semantic categories in the tool are based on the named entity typology of the REPENTINO gazetteer [36], namely: person, organization/place, event, communication, product, document, abstraction, nature, another living being, substance, and “other”.

In order to fully annotate the CSTNews corpus, 5 annotation teams were assembled: one with four members and four with three members. Each team was led by its more senior member, usually a researcher with some experience in NLP and corpus annotation. All the members were native speakers of Portuguese.

Initially, all the teams read the material made available by the organizers of the collective annotation task (regarding the task instructions and how to use the CorreVisual tool) and also studied a didactic reference paper on coreference for Portuguese [10]. After that, a training step was performed over a single text that was provided by the task organizers. The teams had the chance to discuss some annotation issues and, after that, they provided feedback to the task organizers so that they could improve some functionalities.

After the training with a single text, the task organizers considered that the task was well understood and provided the teams with the actual texts to annotate. In average, each annotator had to annotate from 11 to 13 texts, in a period of a month and a half. In general, most of the teams managed to finish the task in 2 to 3 weeks, with each member trying to annotate one text per day.

Some issues were very challenging to deal with. Several members had severe problems with the annotation tool, which apparently has inconsistent functionalities in different operating systems. Given the difficulties with the tool, some annotators preferred to manually annotate the texts (in paper or in a different electronic edition application) before replicating the process in the tool. More serious than the inconsistencies of the tool were its limitations regarding the inclusion of referring expressions and the guideline for avoiding editing the expressions. Such problems are the main cause of some very poor annotations, with wrong referring expressions in the chains (e.g., see in Figure 1 the expression *a aeronave que* in the first chain; in English, “the airplane that”), non-nominal elements (as *que*; “that”), and expressions in the text that were not detected by the tool and, therefore, could not be included in any chain.

The task organizers provided the annotation agreement results to the teams. Agreement was computed over all possible pairs of referring expressions that formed coreference chains, and, for each pair of expressions, it was indicated how many annotators said that the pair was in fact coreferring. Kappa measure [37] was used for computing agreement. This measure is highly adopted in the area because it corrects the results for expected chance agreement. The author proposes that a minimum agreement value of 0.67 is important for tentative conclusions to be drawn from the data. However, the NLP area has learned that different values may be expected in different tasks, as such values highly depend on the difficulty and subjectivity inherent in the task. In our case, we expected lower values.

Table 1 shows the kappa results for each team, including the general average. As said before, for each team, 4 texts were used for computing kappa. One may see that, excepting team 2, all the other teams reached an agreement of at least 0.50. In average, the teams presented an agreement of 0.54, which we consider satisfactory given the annotation conditions (limited training during a short time period, and difficulties and limitations with the annotation tool).

Table 1. Agreement results for the annotation teams

Teams	Kappa
1	0.50
2	0.48
3	0.55
4	0.64
5	0.57
<i>Average</i>	<i>0.54</i>

Overall, including the other participants of the collective annotation task, the minimum agreement value was 0.41, and the maximum was 0.64 (achieved by our team 4).

Some very interesting cases of disagreements may be found in the data, which may be partially explained by the limited training, and partially by different principles regarding the task and the high level of subjectivity in some cases. For instance, to cite a few (related to the text shown in Figure 1):

- while one annotator has built a coreference chain with all the elements that refer to the airport (that is, *Congonhas* - which is the name of the airport - and *o aeroporto* - “the airport”, in English), another one has divided this chain in two, one for the sense of airport and another one for the sense of the location of the airport, being this difference very difficult to realize (and this may explain why the semantic categories “organization” and “place” have been joined by the task organizers);
- one annotator has considered that the expressions *hipótese* (“hypothesis”) and *falha mecânica* (“mechanical failure”) were coreferent (as the “hypothesis” for the accident with the airplane was a “mechanical failure”), but another one considered that these expressions are in different generalization levels and built different chains for them;
- some annotators have built chains for time expressions that refer to the same event (in such cases, some inference was necessary to determine the correct time that was referred to), while others did not, considering that time is not a proper element to a coreference chain (the attentive reader probably realized that the semantic categories in the annotation tool did not include “time”, which is a traditional class in named entity recognition tasks);
- some annotators have included in the chains the occurrences of relative pronouns (e.g., *que*; “that/which”, in English), as they refer to some previous entities and were detected by the annotation tool, but other annotators did not consider such items because they are not strictly nominal expressions (which were the focus of the annotation).

Differences as the previous ones may result in broad variations in the final annotation. For instance, for the text of Figure 1, the three annotators produced from 32 to 46 coreference chains (with unique mentions varying from 49 to 86 referring expressions).

The annotation task required a lot of attention and dedication from the annotators, which had to read several times each text and its referring expressions in order to find all the chains. In many situations, background and domain knowledge was necessary for correctly identifying the chains. Usually, annotators consulted the web (mainly Wikipedia) to solve these cases. In average, the annotators took above 1 hour to annotate each text.

We present some final remarks in the next section.

5 Final Remarks

All the annotated data is in an XML format, which is a traditional way of marking and making data available. It shall be available in the SUCINTO project website, as it constitutes an additional linguistic annotation layer of the CST-News corpus.

We expect that the produced coreference annotation fosters other research initiatives on discourse processing tasks. For the short term, the new data may help to improve summarization models, specifically those involving coherence and cohesion evaluation, for which the occurrence and distribution of referring expressions are very important features (see, e.g., the entity-based model proposed in [38]).

For future work, concerning the CSTNews corpus, the task of pronominal anaphora resolution remains to be done, as it was not directly tackled in the reported annotation effort.

Acknowledgments

The authors are grateful to FAPESP, CAPES and CNPq for supporting this work.

References

1. Jurafsky, D., Martin, J.H.: *Speech and Language Processing: An Introduction to Natural Language Processing, Computational Linguistics and Speech Recognition*. Prentice Hall (2009)
2. Hobbs, J.R.: Resolving pronoun references. *Lingua* **44**(4) (1978) 311–338
3. Mitkov, R.: *Anaphora Resolution*. Pearson Education (2002)
4. Haponchyk, I., Moschitti, A.: A practical perspective on latent structured prediction for coreference resolution. In: *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics*. Volume 2. (2015) 143–149
5. Wiseman, S., Rush, A.M., Shieber, S.M.: Learning global features for coreference resolution. In: *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. (2016) 994–1004
6. Grosz, B., Joshi, A., Weisten, S.: Centering: A framework for modeling the local coherence of discourse. *Computational Linguistics* **21**(2) (1995) 203–225
7. Cristea, D., Ide, N., Romary, L.: Veins theory. an approach to global cohesion and coherence. In: *Proceedings of the 36th Annual Meeting of the Association for Computational Linguistics and 17th International Conference on Computational Linguistics*. (1998) 281–285
8. Pradhan, S., Moschitti, A., Xue, N., Uryupina, O., Zhang, Y., eds.: *Joint Conference on EMNLP and CoNLL – Shared Task*, Association for Computational Linguistics (2012)

9. Cardoso, P.C.F., Maziero, E.G., Castro Jorge, M.L.R., Seno, E.M.R., Di Felippo, A., Rino, L.H.M., Nunes, M.G.V., Pardo, T.A.S.: 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. (2011) 88–105
10. Vieira, R., Gonçalves, P.N., Souza, J.G.C.: Processamento computacional de anáfora e correferência. *Revista de Estudos da Linguagem* **16**(1) (2008) 263–284
11. Collovini, S., Carbonel, T.I., Fuchs, J.T., Coelho, J.C., Rino, L.H.M., Vieira, R.: Summ-it: Um corpus anotado com informações discursivas visando a sumarização automática. In: Proceedings of V Workshop em Tecnologia da Informação e da Linguagem Humana. (2007) 1605–1614
12. Fonseca, E., Vieira, R., Vanin, A.: Improving coreference resolution with semantic knowledge. In: Proceedings of the International Conference on Computational Processing of the Portuguese Language. (2016a) 213–224
13. Fonseca, E., Vieira, R., Vanin, A.: Corp: Coreference resolution for portuguese. In: Proceedings of the International Conference on Computational Processing of the Portuguese Language - Demonstration Session. (2016b) 9–11
14. Fonseca, E., Sesti, V., Atonitsch, A., Vanin, A., Vieira, R.: CORP: Uma abordagem baseada em regras e conhecimento semântico para a resolução de correferências. *LinguaMÁTICA* **9**(1) (2017) 3–18
15. Radev, D.R.: A common theory of information fusion from multiple text sources, step one: Cross-document structure. In: Proceedings of the 1st ACL SIGDIAL Workshop on Discourse and Dialogue. (2000) 74–83
16. Mann, W.C., Thompson, S.A.: Rhetorical structure theory: A theory of text organization. Technical Report Technical Report ISI/RS-87-190 (1987)
17. Hearst, M.: Texttiling: Segmenting text into multi-paragraph subtopic passages. *Computational Linguistics* **23**(1) (1997) 33–64
18. Owczarzak, K., Dang, H.T.: Who wrote what where: Analyzing the content of human and automatic summaries. In: Proceedings of the Workshop on Automatic Summarization for Different Genres, Media, and Languages. (2011) 25–32
19. Baptista, J., Hagège, C., Mamede, N.: Identificação, classificação e normalização de expressões temporais do português: A experiência do segundo HAREM e o futuro. In Mota, C., Santos, D., eds.: *Desafios na avaliação conjunta do reconhecimento de entidades mencionadas: O Segundo HAREM*. (2008) 33–54
20. Fellbaum, C.: *WordNet: An Electronic Lexical Database*. MIT Press (1998)
21. Bick, E.: *The Parsing System PALAVRAS: Automatic Grammatical Analysis of Portuguese in a Constraint Grammar Framework*. Aarhus University Press (2000)
22. Silveira, S.B., Branco, A.: Enhancing multi-document summaries with sentence simplification. In: Proceedings of the 14th International Conference on Artificial Intelligence. (2012) 742–748
23. Ribaldo, R., Cardoso, P.C.F., Pardo, T.A.S.: Exploring the subtopic-based relationship map strategy for multi-document summarization. *Journal of Theoretical and Applied Computing* **23**(1) (2016) 183–211
24. Cardoso, P.C.F., Pardo, T.A.S.: Multi-document summarization using semantic discourse models. *Procesamiento del Lenguaje Natural* **56** (2016) 57–64
25. Nóbrega, F.A.A., Pardo, T.A.S.: Update summarization for portuguese. In: Proceedings of the 6th Brazilian Conference on Intelligent Systems (To appear). (2017)
26. Dias, M.S., Pardo, T.A.S.: A discursive grid approach to model local coherence in multi-document summaries. In: Proceedings of the 16th Annual SIGdial Meeting on Discourse and Dialogue. (2015) 60–67

27. Camargo, R.T., Di Felippo, A., Pardo, T.A.S.: On strategies of human multi-document summarization. In: Proceedings of the 10th Brazilian Symposium in Information and Human Language Technology. (2015) 141–150
28. Maziero, E.G., Hirst, G., Pardo, T.A.S.: Semi-supervised never-ending learning in rhetorical relation identification. In: Proceedings of the Recent Advances in Natural Language Processing. (2015) 436–442
29. Braud, C., Coavoux, M., Søgaaard, A.: Cross-lingual RST discourse parsing. In: Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics. Volume 1. (2017) 292–304
30. Ponti, E.M., Korhonen, A.: Event-related features in feedforward neural networks contribute to identifying causal relations in discourse. In: Proceedings of the 2nd Workshop on Linking Models of Lexical, Sentential and Discourse-level Semantics. (2017) 25–30
31. Di Felippo, A., Nenkova, A.: Phrase generalization: a corpus study in multi-document abstracts and original news alignments. In: Proceedings of the 10th Linguistic Annotation Workshop. (2016) 151–159
32. Cardoso, P.C.F., Pardo, T.A.S., Taboada, M.: On the contribution of discourse to topic segmentation. In: Proceedings of the 14th Annual Meeting of the Special Interest Group on Discourse and Dialogue. (2013) 92–96
33. Maziero, E.G., Castro Jorge, M.L.R., Pardo, T.A.S.: Revisiting cross-document structure theory for multi-document discourse parsing. *Information Processing & Management* **50**(2) (2014) 297–314
34. Souza, J.W.C., Di Felippo, A.: O corpus cstnews e sua complementaridade temporal. In: PROPOR Workshop on Tools and Resources for Automatically Processing Portuguese and Spanish. (2014) 105–109
35. Oliveira, H.G., Gomes, P.: Eco and onto.pt: a flexible approach for creating a portuguese wordnet automatically. *Language Resources and Evaluation* **48**(2) (2014) 373–393
36. Sarmiento, L., Pinto, A.S., Cabral, L.: REPENTINO - a wide-scope gazetteer for entity recognition in portuguese. In: Proceedings of the 7th International Workshop on Computational Processing of the Portuguese Language. (2006) 31–40
37. Carletta, J.: Assessing agreement on classification tasks: the kappa statistic. *Computational Linguistics* **22**(2) (1996) 249–254
38. Barzilay, R., Lapata, M.: Modeling local coherence: An entity-based approach. *Computational Linguistics* **34**(1) (2008) 1–34