The RST Basque TreeBank: an online search interface to check rhetorical relations

Mikel Iruskieta[1], **María Jesús Aranzabe**[2], **Arantza Diaz de Ilarraza**[3], **Itziar Gonzalez-Dios**[3], **Mikel Lersundi**[2], **Oier Lopez de Lacalle**[3]

[1]Department of Didactics of Language and Literature
University of the Basque Country (UPV/EHU)
Postcode 48940 – 0034.94601.7569 – Leioa – Basque Country

mikel.iruskieta@ehu.es

[2]Department of Basque Language and Communication (UPV/EHU)

[3]Department of Computer Science (UPV/EHU)

Abstract. *This paper introduces the first Basque discourse TreeBank annotated with rhetorical relations following Rhetorical Structure Theory. We report the main features of the corpus, such as the annotation criteria, inter-annotator agreement and harmonization procedure. We describe an online search system to check the annotation of discourse relations.*

1. Introduction

In computational linguistics discourse analysis covers a wide range of structural phenomena, such as identification of referential and relational structures. The main task when studying referential structures is correference resolution [Mitkov 2002, Recasens et al. 2010] while relational structures are related to coherence relation assignment [Asher and Lascarides 2003, Mann and Thompson 1988].

Annotated corpus are necessary in order to build advanced applications such as automatic text generation systems [Bouayad-Agha 2000], automatic summarizers [Marcu 2000b] or machine translation systems [Marcu et al. 2000]. These systems rely on different linguistic information, including the discourse level. Consequently, it is important to have a corpus which is annotated at different linguistic levels. Aforementioned systems could take advantage of the available *automatic discourse analyzers* [Marcu 2000b, Pardo et al. 2004], in order to improve their output.

There are a few works that deal with the annotation of referential structures for corpus written in languages such as English [Carlson et al. 2002, Taboada and Renkema 2011], German [Stede 2004], Dutch [van der Vliet et al. 2011], Portuguese [Pardo and Seno 2005] and Spanish [da Cunha et al. 2011a].

In the case of corpus annotation for Basque, we can find studies on referential structure [Goenaga et al. 2012, Ceberio et al. 2009] and relational structure [Iruskieta et al. 2013, Iruskieta et al. 2011]. From the linguistic point of view it is interesting to study languages with a different typology as Basque and to offer annotated corpus to the scientific community.

This work is the first RST corpus for Basque created to serve as a reference for several NLP applications for this language. The annotations follow the RST theory introduced by [Mann and Thompson 1988]. From our point of view: *i*) RST facilitates the representation of coherence in real texts, establishing relations among all the units in a tree-like structure; *ii*) RST has been applied to different languages and used for advanced applications and, *iii*) there are tools which facilitate working with RST annotated corpora: RSTTool [O'Donnell 2000] and Rhetorical DataBase [Pardo 2005]. We present the annotated corpus and we describe an online search interface to check the annotated discourse structure.

The remainder of this paper is structured as follows. Section 2 lays out the theoretical framework and Section 3 the methodology utilized to annotate the corpus. Section 4 sets out the results of the annotation and presents the online search interface. Finally, Section 5 presents the discussion and establishes directions for future work.

2. Annotation in Rhetorical Structured Theory

Rhetorical Structured Theory is a language-independent theory describing coherence between text fragments. It combines the idea of nuclearity, i.e. the importance of an individual fragment from within the discourse, with the presence of rhetorical relations (R) (hypotactic and paratactic relations) between these fragments. Hypotactic and paratactic relations connect discourse units, either a single unit (EDU) or groups of units (span). According to the theory, these relations can be paratactic (N-N) —when they establish relations between fragments that are equally important to the author (LIST, CONTRAST, DISJUNCTION, etc.)— or hypotactic (N-S) —when they connect a less-important unit with a unit the author views to be more important (ELABORATION, MEANS, PREPARATION, CONCESSION, CAUSE, RESULT, etc.). Relations are defined in light of the restrictions established between the nucleus and satellite and by describing the effect they have on the reader. A more detailed explanation of RST can be found in [Mann and Thompson 1988] and in [Mann and Taboada 2010].

Refering to the annotation process, it is well known that agreement is higher when there is training among coders. Works in which annotators did not have a training phase present a similar agreement [van der Vliet et al. 2011]. This fact is reported in the work carried out on the English language [Carlson et al. 2003]; a total of six professional annotators tagged the corpus measuring inter-annotator agreement in different texts (53 to be precise) in a pairwise manner (and in a few cases three-wise manner). There are methods for improving inter-annotator agreement: in [Carlson et al. 2003], for example, it is reported that at the beginning of the project the highest level of agreement attained between the three annotators in a small sample was a Kappa score of 0.602, while at the end of the project, after training, it was 0.755. In this project, in addition to the professional annotators, the authors also measured the agreement between two non-profesional annotators, with very different results: Kappa scores of between 0.597 and 0.792 (1918 EDUs, 30 texts).

The size of the corpus is another aspect to take into acount. We can say that, while the size of our corpus is smaller than that of the corpora found in the bibliography, the fragment tagged in a pairs was comparable as regards both size and number of annotators.

Although the delivery phase is important in annotation [Hovy 2010], it is usually forgotten. This is not the case in the RST Spanish Treebank [da Cunha et al. 2011b]. Relation extraction from a corpus is very helpful for a better understanding of the relation itself or for the study of patterns (this information will be useful to be on the design of automatic rules or as features in machine learning algorithms). In the RST Basque TreeBank the delivery phase is of great importance as we will see in the Section 4.

3. Methodological principles

Our corpus is composed by abstracts, short but well structured texts, written in Basque.[1]

Regarding coherence relations, abstracts function as independent discourse and summarize the main idea of the paper. The percentages of each relation —which are available on the web— are similar to the ones of [Pardo and Nunes 2004].

As regards relational structure, agreement between annotators was measured manually, using the evaluation system based on rhetorical relations presented in [da Cunha and Iruskieta 2010]. We decided not to use the evaluation system that assesses the tree structure [Marcu 2000a], mainly in order to avoid the shortfalls described in [Iruskieta et al. 2013]. According to these authors, span and nuclearity factors are not independent phenomena in the tree structure evaluation proposed in [Marcu 2000a], since they influence the evaluated factor of rhetorical relations. In contrast, [da Cunha and Iruskieta 2010] propose an evaluation method based on rhetorical relations where three factors are assessed: satellite unit or composition span (C), nuclear unit or attachment span (A),[2] and rhetorical relations (R).

3.1. Annotated corpus

The corpus utilized in this study is composed of abstracts from three specialized domains: medicine, terminology and science. Medical texts include the abstracts of all medical articles written in Basque in the Medical Journal of Bilbao (GMB) between 2000 and 2008. Texts related to terminology were extracted from the proceedings of the International Conference on Terminology (TERM) organized in 1997 by UZEI, while scientific articles are papers from the University of the Basque Country's Faculty of Science and Technology (ZTF) Research Conference, which took place in 2008. We have collected 60 documents that contain 15566 words (803 sentences). The created gold standard contains 1355 EDUs and 1292 Rs.

3.2. Annotators

The corpus was annotated by two linguists. The two annotators had previously annotated other linguistic levels (morphosyntax, syntax and semantics), and were familiar with RST and its annotation interface, RSTTool, but no training was provided.

3.3. Annotation phases

The process of tagging the rhetorical structure was divided into four phases. Each phase was evaluated and harmonized by a judge, in order to ensure that all annotators started each new phase from the same basic criteria. The four phases were as follows:

i) **Segmentation:** annotators were asked to divide the text into EDUs; in general, each EDU is either a subordinate clause containing a verb or an independent clause (more details in [da Cunha and Iruskieta 2010]).

[1] In the same sense as [Swales 1990] mentions that abstracts follows an IMRaD (*Introduction*, *Method*, *Results* and *Discussion*) structure.

[2] In multinuclear relations any of the nucleus can be considered as composition or attachment span.

$ii)$ **Identifying the macrostructure:** before identifying the rhetorical relations, annotators were asked to identify most important part of the text or central unit (CU).

$iii)$ **Representing the relational structure:** bearing in mind the CU, rhetorical structure was annotated in a modular and incremental way as proposed in the work by [Pardo 2005] and with the extended classification of rhetorical relations [Mann and Taboada 2010].

$iv)$ **Annotating the signals of relations:** one annotator has tagged the signals of rhetorical relations, as proposed in [Taboada and Das Forthcoming]. The cause subset (CAUSE, RESULT and PURPOSE) was annotated by two annotators and evaluated.

The method mainly used in RST to increase annotator agreement on rhetorical relations is to establish a training phase. From our point of view this could carry a circular process between relations and their signals [Spenader and Lobanova 2009]. To provide a more reliable annotated corpus and do not fall in this circular problem, we analyzed the problems arising amongst annotators, and, in order to achieve our aim (a reference corpus annotated with relational structure), we established the criteria for annotation and we designed a manual for a judge to decide the cases of disagreement.

3.4. Results

We carried out an evaluation to assess each of the annotation steps by means of different agreement measures. This way, we calculated the agreements of segmentation (EDU), the agreement on CU identification, the agreement on rhetorical structure and the agreement on signals of the cause subset. At the rhetorical structure level we provide an analysis of the source of the disagreement, categorizing them in different types.

Segmentation (EDU). Inter-annotator agreement between annotators is 81.35%.

CUs identification. The overall mean agreement between annotators is 81.67%.[3]

Relational structure level. Based on the factors we defined —composition span (C), attachment span (A) and rhetorical relations (R)— the following types of agreements: *i*) **CAR**: agreement in composition span, attachment span and relation, *ii*) **CR**: agreement in composition span and relation, *iii*) **AR**: agreement in attachment span and relation and *iv*) **R**: agreement only in relation. Table 1 shows the agreement level obtained on the four types of measurements.

Agree	K. α	%	Gain
CAR	0.394	47.76%	-
CR	0.458	54.03%	6.27%
AR	0.431	51.17%	3.41%
R	0.561	61.47%	13.71%

Table 1. Types of agreement

Disagree	%	Disagree	%
No-Match	0.23%	Different R	13.62%
Nuclearity	6.73%	Similar R	5.88%
N/N-N/S	8.90%	MissMatch R	2.01%
Attachment	0.08%	Specificy	0.93%
Composition	0.15%	Segmentation	0.15%

Table 2. Types of disagreement

The results show how the agreement increases as the relaxation of the agreement increases too, being CAR the most demanding agreement, and R the more relaxed one.

[3] Agreement related to CU has been different in the three domains. The agreement is related to the number of candidates (text size) and to the enough explicit linguistic evidence which highlights the CU.

The inter-annotator agreement level [Krippendorff 2012] is moderate for relations. It must be noted that we are in the initial phase of the annotation project. Nevertheless, the results obtained are comparable to those achieved in the initial phases of the main work of rhetorical relation annotation carried out for English [Carlson et al. 2003].

On the other hand, we defined different types of disagreement, taking into account the following phenomena: *i*) **No-match**: The composition of the tree results in relations that cannot be compared. *ii*) **Nuclearity**: Different choices in nuclearity entailed discrepancy in hypotactic relations. *iii*) **N/N vs N/S**: Different choices in nuclearity entailed a paratactic/hypotactic mix-up. *iv*) **Attachment span**: Different choices in attachment span entailed a different relation. *v*) **Different R**: A relation has the same composition and attachment span, but not the same relation. *vi*) **Similar R**: Relations chosen are similar in nature. *vii*) **Mismatch R**: Relations with mismatched RST trees. *viii*) **Specificity**: The relation chosen is more specific in one annotation than in the other. *ix*) **Segmentation**: Segmentation does not match.

As shown in Table 2, although the Different R label is the main source of disagreement (13.62% of the times), one of the main disagreement comes from the choice of nuclearity: in total, 15.63% of the annotation disagree on Nuclearity or the N/N-N/S factors. The other types of disagreement (the 8.82% of the annotations) can easily be resolved explaining how the annotator understand the relations involved in Similar R, Mismatch R and Specificity labels.

Signals for rhetorical relations. Finally, a judge resolved the disagreements between annotators, establishing the relational structure model and specifying the signals for rhetorical relations. The average agreement between annotators of the cause subset —which is often signalled— was 78.11% (PURPOSE 90%, CAUSE 76.79% and RESULT 59.7%).

4. The RST Basque TreeBank

When entering in the website,[4] you can find information of the general characteristics of the RST Basque TreeBank and facilities to consult the contents of the tagged corpus, as for example: *i*) discourse units, the central unit and relations linked to the central unit (4.1 subsection); *ii*) all instances of a selected rhetorical relation in the corpus (4.2 subsection); *iii*) the rhetorical structure of a desired text (4.3 subsection); *iv*) all the signals of relations (4.4 subsection) and, *v*) searching facilities for further studies about typical patterns about combination of word-forms, lemma and POS present in the corpus (4.5 subsection).

4.1. Consulting EDUs and CU of a tree

The application offers the possibility to check the linear segmentation (EDUs) of a document as well as its CU. Table 3 shows the segmentation for the GMB0301 document. The text has seven EDUs[5] and the last one, EDU$_7$, has a button called *See* in the CU column. If you click on this button, you will see all the relations linked to the CU of this text.

4.2. Dealing with rhetorical relations

The web application allows you to look up all the occurrences of a specific relation, or restrict your search to a particular sub-corpus (GMB, TERM or ZTF). If the segments are

[4] http://ixa2.si.ehu.es/diskurtsoa/en/

[5] Translations thereof are found underneath these.

	GMB0301-GS.rs3 (7)		
EDU	Segment	Annotator	CU
1	Estomatitis Aftosa Recurrente (I): Epidemiologia, etiopatogenia eta aspektu klinikopatologikoak.	GS	
	Recurrent aphthous stomatitis (I): epidemiologic, etiologic and clinical features.		
2	“Estomatitis aftosa recurrente” deritzon patologia, ahoan agertzen den ugarienetako bat da.	GS	
	“Recurrent aphthous stomatitis” is one of the most frequent oral pathologies.		
3	tamainu, kokapena eta iraunkortasuna aldakorra izanik.	GS	
	having a variable size, location and duration.		
4	Honen etiologia eztabaidagarria da.	GS	
	It has a controversial etiology.		
5	Ultzera mingarri batzu bezela agertzen da,	GS	
	It is characterized by the apparition of painful ulcers,		
6	Hauek periodiki beragertzen dira.	GS	
	These ulcers appear recurrently.		
7	Lan honetan patologia arrunt honetan ezaugarri epidemiologiko, etiopatogeniko eta klinikopatologiko garrantsitsuenak analizatzen ditugu.	GS	See
	In this paper we analyze the most important epidemiological, etiological, pathological and clinical features of this common oral pathology.		

Table 3. Example of the EDUs section, GMB0301

very long and you are only interested in the beginning of each, you can also limit the size.

Table 4 shows a fragment of a search conducted in the relation database. Since the search was limited to the Term corpus, there are only 27 Cause relations, rather than the 56 shown in corpus. The first 3 columns of Table 4 describe the order and direction of the discourse units. Since the segments —left span and right span— follow the order in where they appear in the text, the second column specifies the nuclearity of the relations: if the relation is NS (nucleus on the left and satellite on the right), then the arrow points left (<–), towards the nucleus. If it is SN, then the arrow points right (–>). The fourth column specifies the relation and relation type: in this case, a single nucleus relation (N/S) Cause; when there are multiple nuclei, this is indicated by the letters (N/N). Finally, the source of the example (Ref.) and annotator (Annot.) is specified.[6]

		Relation: Cause (27)			
Left span	NS	Right span	Relation	Ref.	Annot.
Aurreko hamarkadetan, serbierako zientzia-arloko ikertzaile askok joera bat nabaritu dute eta horren berri eman dute: ingeleseko unita[. . .]	<–	Izan ere, iritzi ezberdinetako zientzialari serbiarrek adostasuna lortu dute eta aurreko hamarkadetan ingelesari eman diote [. . .]	Cause	TERM18	GS
In recent decades, many Serbian researchers working in different scientific fields have noticed a tendency and this is outlined here: the English unit [. . .]		Indeed, Serbian scientists from different schools of thought have reached a consensus and have given English [. . .]			
Terminologiak berak ere, uztartu egin behar ditu joera orokor horiek, eransten zaizkien beste batzuekin batera, hala nola: teknologien [. . .]	<–	gizartearekin lotuta dagoen jarduera denez,	Cause	TERM19	GS
Terminology itself must seek to unite these general trends, along with others related to them, for example: technology [. . .]		since it is an activity linked to society,			

Table 4. Example of a Cause relation search

[6] Note: due to space limitations we only mention here the most important information contained in the database. The signals for rhetorical relations are underlined in Table 4.

4.3. Checking all relations of a RST tree

You can also consult the database file by file: viewing the rhetorical relations of the chosen file or its image in JPG format. The rhetorical structure can be consulted in different formats (XML and RS3). Other information can be consulted here: text file in TXT format, morphosyntactic information annotated automatically in KAF format [Bosma et al. 2009], and the signals for relations annotated in RHETDB format.

4.4. Signals of rhetorical relations

You can check if a signal is in more that one relation. We show as an example a query based on the adversative conjunction *baina* 'but' in Table 5, which signals two similar relations (CONTRAST and CONCESSION).[7]

Signal: *baina* 'but'			
Gainerakoan, prokasu adierazle egokiak daude,	Kontzesioa	baina altan dagoen gaixoaren ahalmen funtzionalaren erregistro urria antzematen da,	GMB0504
With respect to the other aspects, the indicators of process are good	Concession	but there is poor recording of the patient's functional capacity on discharge,	
Bestalde, Euskaltzaindiak hitz elkartuen bidea (1995eko urtarrilaren 27an onartutako araua) proposatzen du adjektibo erreferentzialak itzultzeko,	Kontrastea	baina arauan bertan esaten denez, "... ahal den guztian...",	TERM22
Euskaltzaindia proposed a mechanism of compound words (in a standard approved on January 27th 1995) for the translation of referential adjectives.	Contrast	However the academy also confirmed, ... "whenever possible",	

Table 5. Example of the SIGNALS section, the discourse marker *baina* 'but'

4.5. Word form, lemma and POS search interface

Searches combining word-form, lemma and POS features can be done in the application due to the fact that all the words in the texts have associated morphological and syntactical information in KAF format.

	Doc.	**Sent Id**	**Word**	**CU**	**Sentence**
1	TERM50	sent2	taldeek / helburua	BAI	[...] Hitzaldi honek azken hiru urteotan lau unibertsitate hauen *talde*ek egindako ikerkuntzaren ondorioetako batzuk azaltzeko *helburua* izango luke.
			groups / aim	YES	"[...] The aim of this talk is to present some of the results of the research carried out by groups from these four universities over the last three years."
2	ZTF13	sent1	taldearen / helburu	BAI	[...] Gure *ikerkuntza talde*aren *helburu* nagusia, [...]
			group's / aim	YES	[...] Our research group's principal aim, [...]
3	ZTF13	sent17	taldearen / helburu	EZ	Alor honetan, gure *ikerkuntza talde*aren *helburu* nagusiak bi dira.
			group's / aim	NO	In this field, our research group has two main aims.
1	ZTF15	sent7	helburu / talde	EZ	[...] bestelako galdera zailagoei ere erantzutea dute *helburu*, hala nola, espezieen biogeografia, *talde*aren filogenia, eta abar.
			aim / group	NO	[...] the aim is to answer other such difficult questions, such as species biogeography, group phylogeny, etc.

Table 6. Example of the SEARCH section

These searches provide the option of searching patterns. For example, in a two-word search, you can specify to show the sentences which contain words starting with the forms *talde* 'group' or 'team' and *helburu* 'goal' or 'aim'. You can also define whether or not other words can be located between the target terms. Table 6 shows a search for the

[7] More information about ambiguity in this corpus can be read in [Iruskieta and da Cunha 2010] and in [Iruskieta et al. 2009].

terms *talde* 'group' and *helburu* 'aim' results in two YES responses for CU, but another search with the terms the other way round (aim and group) would only give one NO response for CU.

5. Discussion and Future Work

This paper presents the first RST Basque TreeBank, where the gold standard files that have been used to compile the database are at the disposal of anyone who wishes to use them. Moreover, the study also served to design the harmonization processes for the different annotation phases (segmentation, identification of central units, rhetorical relations and its signals), as well as giving the judge the opportunity of consulting both their annotations and those of the annotators, seeing at a single glance the frequency of each relation and its signals. This in turn enabled the detection of errors and incoherence during the establishment of the gold standards.

The work carried out is useful for certain language processing tasks. Indeed, during the course of the project we established a segmented gold standard for 60 texts, on the road towards automatic segmentation. As regards rhetorical relations, after establishing a gold standard for 60 texts, we marked the signals of those relations, being the size of the work similar to that of others in the literature [Taboada and Das Forthcoming]. In the future, this work will help us define rhetorical relation patterns, and this in turn will help us achieve automatic detection of those most commonly signaled relations.

The authors are currently striving to achieve the following aims: in the short medium term, their goal is to annotate texts from another genre: newspaper articles, texts from the EPEC corpus and to study deeply the signals of relations in the RST Basque TreeBank. With the data provided by the RST Basque TreeBank, they are implementing an automatic discourse segmentation program. Besides, and considering how time consuming the tagging and evaluation processes are, the authors are working on the implementation of a new interface to facilitate the editing of rhetorical relations and programs for automatic evaluation program based on rhetorical relations.

Acknowledgments

This study was carried out within the framework of the following projects: Ber2Tek (IE12-333); Hibrido Sint (TIN2010-20218); NewsReader project (FP7- ICT-2011-8-316404); IXA group, Research Group of type A (IT344-10). We would like to thank Esther Miranda and Kike Fernandez for their help in designing the web page.

References

[Asher and Lascarides 2003] Asher, N. and Lascarides, A. (2003). *Logics of conversation*. Cambridge Univ Pr, Cambridge.

[Bosma et al. 2009] Bosma, W., Vossen, P., Soroa, A., Rigau, G., Tesconi, M., Marchetti, A., Monachini, M., and Aliprandi, C. (2009). KAF: a generic semantic annotation format. In *GL2009 Workshop on Semantic Annotation*, Italy.

[Bouayad-Agha 2000] Bouayad-Agha, N. (2000). Using an abstract rhetorical representation to generate a variety of pragmatically congruent texts. In *Annual Meeting-ACL*, volume 38, pages 16–22.

[Carlson et al. 2003] Carlson, L., Marcu, D., and Okurowski, M. E. (2003). *Building a discourse-tagged corpus in the framework of Rhetorical Structure Theory*, pages 85–112. Current and new directions in discourse and dialogue. Springer, Berlin.

[Carlson et al. 2002] Carlson, L., Okurowski, M. E., and Marcu, D. (2002). *RST Discourse Treebank, LDC2002T07 [Corpus].* PA: Linguistic Data Consortium, Philadelphia.

[Ceberio et al. 2009] Ceberio, K., Aduriz, I., Díaz de Ilarraza, A., and Garcıa, I. (2009). Empirical study of the relevance of semantic information for anaphora resolution: the case of adverbial anaphora. In *7th Discourse Anaphora and Anaphor Resolution Colloquium (DAARC09)*, pages 56–63, Goa, India.

[da Cunha and Iruskieta 2010] da Cunha, I. and Iruskieta, M. (2010). Comparing rhetorical structures in different languages: The influence of translation strategies. *Discourse Studies*, 12(5):563–598.

[da Cunha et al. 2011a] da Cunha, I., Torres-Moreno, J. M., and Sierra, G. (2011a). On the Development of the RST Spanish Treebank. In *5th Linguistic Annotation Workshop (LAW V '11)*, pages 1–10, Portland, USA.

[da Cunha et al. 2011b] da Cunha, I., Torres-Moreno, J. M., Sierra, G., Cabrera-Diego, L.-A., and Castro-Rolón, B.-G. (2011b). The RST Spanish Treebank On-line Interface. In *International Conference Recent Advances in NLP*, Bulgaria.

[Goenaga et al. 2012] Goenaga, I., Arregi, O., Ceberio, K., de Ilarraza, A. D., and Jimeno, A. (2012). Automatic coreference annotation in basque. In *Eleventh International Workshop on Treebanks and Linguistic Theories*, Portugal.

[Hovy 2010] Hovy, E. (2010). Annotation: A Tutorial. In *48th Annual Meeting of the ACL*, Uppsala, Sweden.

[Iruskieta and da Cunha 2010] Iruskieta, M. and da Cunha, I. (2010). Marcadores y relaciones discursivas en el ámbito médico: un estudio en español y euskera. In *XXVIII Congreso Internacional AESLA: Analizar datos > Describir variación*, pages 13–159, Vigo.

[Iruskieta et al. 2009] Iruskieta, M., de Ilarraza, A. D., and Lersundi, M. (2009). Correlaciones en euskera entre las relaciones retóricas y los marcadores del discurso. In *Proceedings of 27th AESLA International Conference*, pages 963–971, Ciudad Real, Spain.

[Iruskieta et al. 2011] Iruskieta, M., Díaz de Ilarraza, A., and Lersundi, M. (2011). Unidad discursiva y relaciones retóricas: un estudio acerca de las unidades de discurso en el etiquetado de un corpus en euskera. *Procesamiento del Lenguaje Natural*, 47:144.

[Iruskieta et al. 2013] Iruskieta, M., Díaz de Ilarraza, A., and Lersundi, M. (2013). A critical analysis of rhetorical annotation: fundamental principles of discourse segmentation in basque. *Corpus Linguistics and Linguistic Theory*, 0(0):1–32.

[Krippendorff 2012] Krippendorff, K. (2012). *Content analysis: An introduction to its methodology*. SAGE, London.

[Mann and Taboada 2010] Mann, W. C. and Taboada, M. (2010). RST web-site. *http://www.sfu.ca/rst/*.

[Mann and Thompson 1988] Mann, W. C. and Thompson, S. A. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. *Text*, 8(3):243–281.

[Marcu 2000a] Marcu, D. (2000a). The rhetorical parsing of unrestricted texts: A surface-based approach. *Computational Linguistics*, 26(3):395–448.

[Marcu 2000b] Marcu, D. (2000b). *The theory and practice of discourse parsing and summarization*. The MIT press, Cambridge.

[Marcu et al. 2000] Marcu, D., Carlson, L., and Watanabe, M. (2000). The automatic translation of discourse structures. In *1st North American chapter of the Association for Computational Linguistics conference*, pages 9–17, Seattle (USA).

[Mitkov 2002] Mitkov, R. (2002). *Anaphora resolution*, volume 134. Longman London.

[O'Donnell 2000] O'Donnell, M. (2000). Rsttool 2.4: a markup tool for rhetorical structure theory. In *6th European Workshop on Natural Language Generation*, Germany.

[Pardo 2005] Pardo, T. A. S. (2005). Métodos para análise discursiva automática. PhD Thesis. Instituto de Ciências Matemáticas e de Computação, Universidade de São Paulo.

[Pardo and Nunes 2004] Pardo, T. A. S. and Nunes, M. G. V. (2004). Relações Retóricas e seus Marcadores Superficiais: Análise de um Corpus de Textos Científicos em Português do Brasil. Technical Report NILC-TR-04-03.

[Pardo et al. 2004] Pardo, T. A. S., Nunes, M. G. V., and Rino, L. H. M. (2004). DiZer: An automatic discourse analyzer for Brazilian Portuguese. *Advances in Artificial Intelligence–SBIA 2004*, pages 224–234.

[Pardo and Seno 2005] Pardo, T. A. S. and Seno, E. R. M. (2005). Rhetalho: um corpus de referência anotado retoricamente. *Anais do V Encontro de Corpora*, pages 24–25.

[Recasens et al. 2010] Recasens, M., Màrquez, L., Sapena, E., Martí, M. A., Taulé, M., Hoste, V., Poesio, M., and Versley, Y. (2010). Semeval-2010 task 1: Coreference resolution in multiple languages. In *5th International Workshop on Semantic Evaluation*, pages 1–8, Sweden. Association for Computational Linguistics.

[Spenader and Lobanova 2009] Spenader, J. and Lobanova, A. (2009). Reliable discourse markers for contrast relations. In *Proceedings of the 8th International Conference on Computational Semantics*, Tilburg, The Netherlands.

[Stede 2004] Stede, M. (2004). The Potsdam Commentary Corpus. In *2004 ACL Workshop on Discourse Annotation*, pages 96–102, Barcelona, Spain.

[Swales 1990] Swales, J. M. (1990). *Genre analysis: English in academic and research settings*. Cambridge Univ Pr, Cambridge, UK.

[Taboada and Das Forthcoming] Taboada, M. and Das, D. (Forthcoming). Annotation upon annotation: Adding signalling information to a corpus of discourse relations. *Dialogue and Discourse*.

[Taboada and Renkema 2011] Taboada, M. and Renkema, J. (2011). Discourse Relations Reference Corpus. http://www.sfu.ca/rst/06tools/discourse_relations_corpus.html.

[van der Vliet et al. 2011] van der Vliet, N., Berzlánovich, I., Bouma, G., Egg, M., and Redeker, G. (2011). Building a discourse-annotated dutch text corpus. *Bochumer Linguistische Arbeitsberichte*, 3:157–171.