Detecting central units in argumentative answer genre: signals that influence annotators' agreement

La detección de las unidades centrales en respuestas argumentativas: señales que influenciam el acuerdo entre anotadores

Juliano Desiderato Antonio
Universidade Estadual de Maringá
Av. Colombo, 5790 – Jd. Universitário – CEP
87020-900 – Maringá – PR – Brasil
jdantonio@uem.br

Resumen: Este trabajo tiene como objetivo la investigación del acuerdo entre anotadores en la detección de unidades centrales en un corpus de cien textos de género respuesta argumentativa. La teoría que subyace la investigación es Rhetorical Structure Theory (RST), propuesta por Mann y Thompson (1988), cuyo objetivo es investigar la coherencia del texto con respecto a las relaciones entre las partes del texto, tanto con la macro y la microestructura. La investigación dio a conocer el patrón más común de unidades centrales: la unidad central es la declaración inicial del texto; la respuesta reformula la pregunta y la unidad central presenta la estructura "El secreto de vestibular + verbo copulativo + fator(es)" o una estructura que tiene en cuenta más de un factor; la declaración inicial puede ser formulada con los verbos modales de actitud proposicional y adverbios también pueden utilizarse, especialmente epistémicos aseverativos. Las dicrepancias han sido motivados por errores de escritores o de anotadores. Los Errores de los anotadores fueron los siguientes: identificar la conclusión como la unidad central porque la conclusión parece más completa y más fuerte que la declaración inicial; considerar alguna palabra en el texto un indicador más fuerte de la unidad central que el patrón presentado anteriormente. Los errores de los escritores fueron los siguientes: la elección de un factor y más adelante del texto eligir otro factor(es) como el secreto de vestibular; transmitir la misma respuesta en dos ocasiones por paráfrasis. Los resultados de este trabajo pueden ser útiles para el desarrollo de aplicaciones de evaluación automáticos.
Palabras clave: RST, unidad central, género respuesta argumentativa.

Abstract: This paper aims at investigating annotators' agreement in central unit detection in a corpus of one hundred texts of argumentative answer genre. The theory underlying the investigation is Rhetorical Structure Theory (RST), proposed by Mann and Thompson (1988), which aims at investigating text coherence regarding relations held between parts of text, both in macro and microstructure. The research unveiled the most common pattern of central units: the central unit is the initial statement of the text; the answer resumes the question and the central unit presents the structure "The secret of Vestibular + copula verb + factor(s)" or a structure which takes into account more than one factor; the initial statement may be embedded in evidential verbs of propositional attitude and adverbs may also be used, specially epistemic asseverative adverbs. Discrepancies were motivated by annotators' mistakes or by writers' mistakes. Annotators' mistakes were the following: identifying the conclusion as the central unit because the conclusion seems more complete and stronger than the initial statement; considering some word in the text a stronger indicator of the central unit than the pattern presented previously. Writers' mistakes were the following: choosing one factor and further up the text choosing another factor(s) as the secret of Vestibular; conveying the same answer twice by paraphrasing. The results of this paper may be useful for the development of automatic evaluation applications.
Keywords: RST, central unit, argumentative answer genre.

1   Introduction

According to Iruskieta, Ilarraza and Lersundi (2015), detecting the central unit of the text is a key step in annotating rhetorical relations since "a greater degree of agreement with regard to the central unit leads to a greater degree of agreement in rhetorical relations linked to the central unit".

Mann and Thompson (1988) state that nuclearity is a central organizing principal of discourse coherence. Thus, the detection of the central unit focuses the most relevant and essential piece of information, the nucleus to which all the other texts spans[1] are related. Discourse coherence is conceived in this paper in van Dijk's (1980) sense: a property that applies not only to adjacent sentences, but also to the global level of the text.

Within Rhetorical Structure Theory (RST henceforth) framework, rhetorical relations are of meaning and not of form (Mann and Thompson, 1988). This means that relations can be held and interpreted without being formally marked by connectives. Thus, other means of signaling relations must be investigated in Phonology, Morphology, Syntax, discourse genre etc (Taboada, 2006).

This paper aims at investigating annotators' agreement in a corpus of one hundred texts of argumentative answer genre produced for the Vestibular (entrance exam) of Universidade Estadual de Maringá (UEM henceforth). Four different annotators with basic RST training detected the central unit of each text. Based on the frequency of agreement among the annotators, the paper also intends to describe the features of the units in which there is a higher frequency of agreement. Features investigated are: position of the central unit within the text; resumption of the question; frequency of nouns, adverbs, verbs. Another aim of the paper is to investigate features responsible for annotators' disagreement such as annotators' mistakes and/or writers' mistakes that compromise text structure and text coherence.

The identification of such features may be helpful to the implementation of automatic evaluation applications and even for the training of the Portuguese teachers committee which corrects the texts manually.

In order to achieve the goals presented previously, this paper is organized in three more sections. In section 2, the main tenets of the theory that underlies the paper, the research corpus and the methodology used in the research are unveiled. Section 3 focuses on the presentation and on the discussion of the results. In the final section, conclusions of this study are presented.

2   Theory, corpus and methodology

2.1 Theory

One of the main tenets of RST is that besides the propositional content conveyed by the clauses of a text, implicit propositions arise from the combining of the clauses or of text spans which hang together. RST is a theory proposed by Mann and Thompson (1988) which aims at investigating text coherence regarding relations held between parts of text, both in macro and microstructure. RST has been used in Descriptive Linguistics and in Computational Linguistics, and implicit propositions have been labeled by various authors as coherence relations, discourse relations, rhetorical relations or relational propositions (Taboada, 2009).

Matthiessen and Thompson (1988) state that rhetorical relations permeate the whole text and can be held by text spans of any size. RST assumes that these relations are essential to the coherence of the text (Mann and Thompson, 1988), as they confer unity to the text and enable the producer of the text to reach the aims that the text was intended to.

A list of 25 relations was initially suggested by Mann and Thompson (1988) after the analysis of hundreds of texts with RST. Afterwards an extended version of the list has been provided at RST website (http://www.sfu.ca/rst/01intro/definitions.html). This list is not a closed one, and other relations can be defined according to the investigated type or genre of text.

---

[1]   Spans are discourse units which can vary from one clause or sentence to the combination of other spans (Taboada and Habel, 2013).

Regarding organization, relations can be of two types: nucleus-satellite (hypotactical), in which satellite is ancillary to the nucleus; multinuclear (paratactical), in which a text span is not ancillary to the other (each span is a distinct nucleus).

In figure 1, the RST diagram of a text of the corpus is presented. As our paper does not aim at analyzing the structure of the whole text, spans have not been segmented into elementary units, i.e., units which cannot be divided into smaller units.

The text is divided into 5 spans. Span 1 is the central unit, i.e., the unit which presents the main idea of the text. As the other spans are satellites regarding the central unit, the arrows point towards span 1. Spans 2-4 hold elaboration relation with the nucleus. In elaboration relation, the satellite provides additional detail about the elements of the nucleus (Mann and Thompson, 1988). In other words, the writer provides more information about the three secrets to achieve success in the entrance exams presented in span 1. In span 2 he/she elaborates intelligence, in span 3 he/she elaborates effort and in span 4 he/she elaborates luck. The relation held among spans 2, 3 and 4 is list, a multinuclear one. Finally, span 5 is a conclusion satellite.
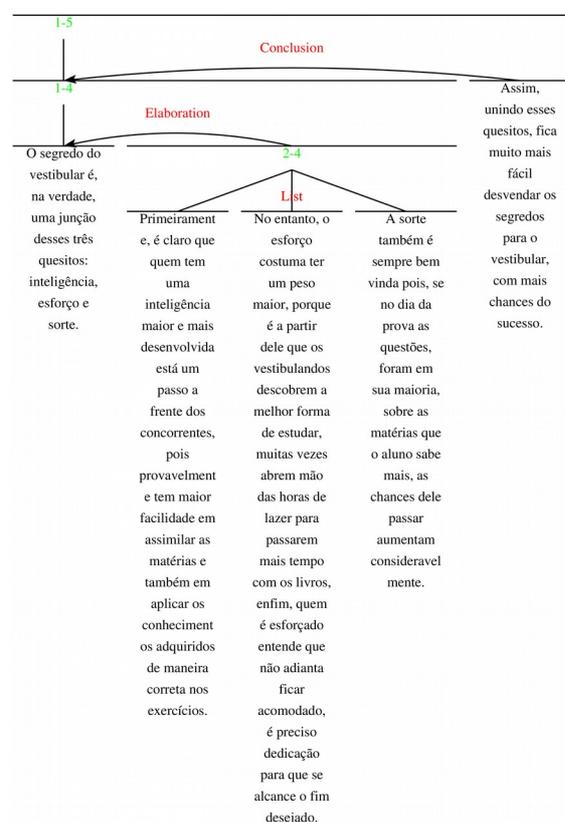

Figure 1: Rhetorical Structure of an argumentative answer

This analysis is supported by the organization of argumentative answer genre, which starts with the resumption of the question followed by the answer to the question (Menegassi, 2011). However, some analyst without knowledge of the organization of the genre might propose another analysis and consider spans 2-4 the central unit. In such a case, the rhetorical three would be completely different. This shows the relevance of using criteria for the detection of the central unit of the text.

2.2 The corpus

The 100 texts of the corpus were written by candidates for Summer 2013 entrance exams at UEM[2]. Candidates were asked to write two texts, a different genre each, but, for the aims of this paper, only argumentative answer genre is analyzed. There are excerpts the candidates can base upon to write the texts demanded by the instructions. On Summer 2013 the instructions for argumentative answer were:

> *Como vestibulando, redija, em até 15 linhas, uma resposta argumentativa à pergunta: "Qual o segredo do vestibular: inteligência, esforço ou sorte?". Você pode basear-se nas informações dos textos de apoio, mas não deve apresentar cópia deles[3].*

According to Menegassi (2011), argumentative answer genre belongs to scholar/academic sphere. It is initiated by the resumption of the question followed by the answer to the question, which is the thesis defended by the author. The remainder of the text presents arguments that support the thesis in order to try to convince or persuade the reader.

Antonio and Santos (2014) have described the rhetorical structure of argumentative answer genre. The initial statement is the central unit of the text and its development is the satellite. In the corpus investigated by the authors, evidence relation is very common to be held between the satellite and the nucleus as the writer intends to increase the reader's belief on the content of the nucleus. Elaboration relation also occurs in the corpus when writers intend to present additional information to the content of the nucleus. It must be noticed that not only these relations may be held between the central unit and the remainder of the text in argumentative answers. These relations were frequent in Antonio and Santos' (2014) investigation because of the characteristics of the investigated corpus.

2.3 Methodology

After the evaluation period, researchers requested UEM's entrance committee access to a lot of one hundred texts, as the lots that are evaluated by UEM's Portuguese teachers committee. The request did not mention any criteria for the selection of the texts by the committee because the aim of the research was not only to describe a certain pattern for the central unit, but also to investigate the motives for discrepancies among annotators. Furthermore, if evaluation is to be made automatically one day, all types of texts will have to be evaluated by the machine, not only the well structured ones. In other words, machines will have to face the same challenges human evaluators face when they have to assign grades to more than a hundred texts a day.

The texts were digitalized and segmented into elementary discourse units manually. Four annotators with basic RST training who take part in the research were asked to indicate the central unit.

After the annotator's agreement was calculated, the central units identified by the four annotators without discrepancy were investigated in order to find a pattern of frequency of nouns, adjectives, adverbs, verbs and also the position of the central unit in the text and whether there was resumption of the question in the beginning of the answer.

For the texts in which there was discrepancy the methodology was different. A super annotator (Hovy, 2010) also identified the central unit of each text. The super annotator

confirmed one identification and discarded the other(s), which was/were investigated in order to find out the features responsible for the discrepancy among annotators.

3 Results

Table 1 presents the frequency of agreement and of discrepancy among the annotators. The four annotators identified the same central unit in 75% of the texts (full agreement). In 23% of the texts, two different central units were identified by the annotators. In 2% of the texts, annotators identified three different central units.

	N	%
Full agreement	75	74%
Discrepancy: 2 central units annotated	23	24%
Discrepancy: 3 central units annotated	2	2%
Total	100	100%

Table 1: Frequency of agreement and discrepancy of central unit identification

3.1 Full agreement

In the case of full agreement, for the argumentative answer genre, the relevant features investigated were:

a) position of the central unit: it is expected that the texts begin with the answer, which is the central unit;

b) resumption of the question: it is expected that the answer follows the resumption of the question;

c) frequency of nouns, adverbs, verbs: it is expected that high frequency is associated with some pattern of answer.

Table 2 presents the frequency of the position of the central unit in the text (initial vs. non-initial) and the resumption or not of the question.

	Initial		Non-initial	
	N	%	N	%
With question resumption	62	82,7%	4	5,3%
Without question resumption	6	8%	3	4%
	68	90,7%	7	9,3%

Table 2: Position of the central unit and resumption of the question to start the answer

The data presented in the table shows that the central unit is in the beginning of 90,7% of the texts and the question is resumed in 88% of the central units. The expected initial statement of the argumentative answer is fulfilled by 82,7% of the central units. Example 1 presents a prototypical central unit found in the corpus.

(1) O segredo do vestibular é o esforço.

Due to the lack of space, it is not possible to present in this paper all the 55 nouns found in the central units of the 75 texts in which there was full agreement among the annotators. Thus, only the nouns with higher frequency will be presented. As it is possible to notice on table 3, five nouns were responsible for 75% of all the 317 noun occurrences.

	N
Esforço (*effort*)	66
Vestibular (*university entrance exam*)	66
Segredo (*secret*)	62
Inteligência (*intelligence*)	27
Sorte (*luck*)	19
Total	240 (75,1% out of 317)

Table 3: Nouns with higher frequency

The higher frequency of the nouns in table 3 is motivated by the fact that most central units are in compliance with the expected form for the initial statement of the argumentative answer. Words "segredo" and "vestibular" are resumed from the question and words "effort", "intelligence" and "luck" are the answer.

Another group of nouns must be focused. It is the group of nouns used by the writers who did not pick only one of the factors suggested in the question (intelligence, effort or luck), but elected more than one factor for the answer, as in example 2, extracted from the corpus.

(2) O segredo do vestibular é certamente uma mistura de inteligência, esforço e sorte.

Such nouns are presented in table 4. All of them are related to the meaning of junction,

combination etc. It is important to notice that the word "misciegenation" was used inadequately meaning union of the factors intelligence, effort and luck.

	N
Junção (*junction*)	4
Combinação (*combination*)	3
Conjuntura (*conjuncture*)	1
Miscigenação (*miscegenation*)	1
Mistura (*mixture*)	1
Mix (*mixture*)	1
Soma (*sum*)	1
União (*union*)	1
Total	13

Table 4: Junction nouns

Most verb tokens had copula function. Copula verbs such as "ser" (*be*) and "estar" (*be*) presented frequency of 94,7%, which can be explained by the structure of the question which was resumed by the answer: "The secret of Vestibular + *copula verb* + **factor**".

Evidential verbs which express propositional attitude (Hattnher, 2007) were used by the writers to embed the answer as in example 3, extracted from the corpus.

(3) Acredito que o segredo do vestibular é uma mistura desses três fatores.

The frequency of evidential verbs is presented in table 5.

	N
Acredito (*I believe that...*)	3
Creio (*I believe that...*)	1
Penso (I think that...)	1
Total	5

Table 5: Evidential verbs

Adverbs and adverbial phrases were also used by the writers in some central units. Frequencies are presented in table 6.

		N	%
Epistemic (Neves, 2000)	Sem dúvida (*no doubt*)	5	26,2%

	Na verdade (*as a matter of fact*)	1	5,3%
	Indubitavelmente (*undoubtedly*)	1	5,3%
	Certamente (*certainly*)	1	5,3%
Intensifier (Neves, 2000)	Muito (*very, much, a lot*)	3	15,7%
	Pouco (*little, a little*)	2	10,5%
Inclusion with exclusivity (Neves, 2000)	Apenas (*only*)	2	10,5%
	Só (*only*)	1	5,3%
Focalizer (Castilho, 2010)	Principal-mente (*mainly*)	1	5,3%
	Também (*too, also*)	1	5,3%
Frequency (Neves, 2000)	Geralmente (*generally*)	1	5,3%
Total		19	100%

Table 6: Adverbs

It is important to remark that all epistemic adverbs used by writers were asseverative in an attempt to make their propositions more credible, as in example 4.

> (4) O segredo do vestibular é sem dúvida o esforço.

Thus, the most frequent pattern found for the central units in which there was full agreement by all the annotators may be summarized as follows:

(a) The central unit is the initial statement of the text;

(b) The answer resumes the question and the central unit presents the structure "The secret of Vestibular + *copula verb* + *factor(s)*" or a structure which takes into account more than one factor:"The secret of Vestibular + *copula verb* + [junction of] *factor(s)*";

(c) The initial statement may be embedded in evidential verbs of propositional attitude such as "think" and "believe": "[I + *evidential verb* + that] the secret of Vestibular + *copula verb* + [junction of] *factor(s)*";

(d) Adverbs may also be used, specially epistemic asseverative adverbs: "[I + *evidential verb* + that] the secret of Vestibular + *copula verb* + [*adverb*] [junction of] *factor(s)*".

3.2 Discrepancies

Out of the twenty-six discrepancies among the annotators, twelve can be considered mistakes of the annotators, as the central unit could be easily identified using the pattern described in the previous section.

Five out of the twelve mistakes happened because the annotators identified the conclusion of the text as the central unit. It is plausible that the fact the conclusion may be stronger then the initial statement in terms of expressiveness. The conclusion is usually started by a discourse marker or a finisher expression such as "portanto" (*therefore*), "enfim" (*ultimately*), "a partir disso" (*taking this into account*), "sendo assim" (*thus*). Besides that, the fact that the answer and the stronger arguments are restated makes the conclusion seem more assertive than the initial statement. In example 5 conclusion is started with discourse marker "portanto", the question is resumed ("o segredo para um bom vestibular"), the answer is restated ("esforço") adding more words of the same semantic frame are added such as "força de vontade" (*willpower*), "determinação" (*determination*), and a bonus factor is also given: "manter a calma" (*keep calm*). Thus the conclusion seems more complete and stronger than the initial statement, misguiding the annotator.

> (5) Portanto, o segredo para um bom vestibular não requer apenas a inteligência, o que conta muito é o esforço, a força de vontade, determinação e principalmente manter a calma.

The remaining seven mistakes were motivated by some word in the text which seems to indicate the central unit more strongly than the pattern presented in the previous section. It is the case of qualifier adjectives (Neves, 2000) such as "essencial" (*essential*), "fundamental" (*fundamental*), "melhor" (*the best*), "único" (*unique*); nouns such as "chave" (*key*), "fruto" (meaning *result*); adverb "primeiramente" (*firstly*); verb form "acredito"

(*I believe that…*). In example (6) the writer states that "conhecimento amplo e diversificado" (*broad and diverse knowledge*) are extremely necessary using adjective "fundamental" (*fundamental*), misguiding the annotator.

> (6) O conhecimento amplo e diversificado é fundamental para a realização do vestibular.

Problems in the structure of the texts were responsible for the other fourteen discrepancies. Two kinds of problems were found: (i) two different answers; (ii) restatement of the same answer in two different positions in the text.

Nine texts presented two different answers. Writers choose one factor and further up the text choose another factor(s) as the secret of Vestibular, as in example 7.

> (7) Para ter um bom desempenho no vestibular é necessário acima de tudo estudar. (…) … então podemos concluir que é preciso inteligência, esforço e sorte para conseguir entrar no curso desejado.

In example (7) the writer states initially that "estudar" (*studying*) is necessary beyond everything to succeed at Vestibular. However the writer concludes the text picking three other factors: inteligência (*intelligence*), esforço (*effort*) and sorte (*luck*). As the writer provides two different answers it is impossible to identify the central unit of the text.

In five texts the writers restated the same answer twice by paraphrase. In example 9 the writer states that "esforço" (*effort*) is extremely important. Further up the text the same information is restated by paraphrasing: *The most important thing is to give the maximum of yourself*. As the same answer was given twice, it is impossible to identify one and only central unit for the text.

> (8) O esforço é de extrema importância para qualquer um… (…) O importante é dar o máximo de si.

It is possible now to summarize the motives for discrepancies in the annotation of the central unit taking into account the two types of problems pointed in this section:

a) annotators' mistakes: identifying the conclusion as the central unit because the conclusion seems more complete and stronger than the initial statement; considering some word in the text a stronger indicator (Paice 1980) of the central unit than the pattern "[I believe that] the secret of Vestibular + copula verb + [adverb] [junction of] factor(s)" presented as initial statement;

b) writers' mistakes: writers choose one factor and further up the text choose another factor(s) as the secret of Vestibular; writers convey the same answer twice by paraphrase. In both cases, two spans may be considered central unit, misguiding the annotator.

4   Conclusion

This paper aimed at investigating annotators' agreement in a corpus of one hundred argumentative texts produced for UEM's entrance exam. It also aimed at describing the pattern used by the writers to indicate the central unit by means of investigating position of the central unit, resumption of the question in the beginning of the answer, frequency of nouns, adverbs and verbs. Another aim of the paper was to investigate the features responsible for annotators' disagreement.

The most frequent pattern may be summarized as follows:

(a) The central unit is the initial statement of the text;

(b) The answer resumes the question and the central unit presents the structure "The secret of Vestibular + *copula verb* + factor(s)" or a structure which takes into account more than one factor:"The secret of Vestibular + *copula verb* + [junction of] *factor(s)*";

(c) The initial statement may be embedded in evidential verbs of propositional attitude such as "think" and "believe": "[I + *evidential verb* + that] the secret of Vestibular + *copula verb* + [junction of] *factor(s)*";

(d) Adverbs may also be used, specially epistemic asseverative adverbs: "[I + *evidential verb* + that] the secret of Vestibular + *copula verb* + [*adverb*] [junction of] *factor(s)*".

The features responsible for discrepancies in the annotation of the central are the following:

a) annotators' mistakes: identifying the conclusion as the central unit because the conclusion seems more complete and stronger than the initial statement; considering some word in the text a stronger indicator of the central unit than the pattern "[I believe that] the secret of Vestibular + copula verb + [adverb]

[junction of] factor(s)" presented as initial statement;

b) writers' mistakes: writers choose one factor and further up the text choose another factor(s) as the secret of Vestibular; writers convey the same answer twice by paraphrase. In both cases, two spans may be considered central unit, misguiding the annotator.

The results of this paper may be useful for the development of automatic evaluation applications and also for the training of the Portuguese teachers committee who evaluate University entrance exams writing tests. In the case of automatic evaluation, it is necessary to implement applications for automatic segmentation and central unit detection. In the case of the manual evaluation by the Portuguese teachers, trainings must be provided for two reasons. One of them is to distinguish well written argumentative answers, in which the central unit presents the expected pattern and the remainder of the text develops the central unit, from texts with structure and coherence problems, which present more than one central unit (if that is possible!). Another reason is to avoid mistakes as the ones made by the annotators who took part in the research.

Acknowledgements

References

Antonio, J. D. y J. A. Santos. 2014. A estrutura retórica do gênero resposta argumentativa. *Signum: Estudos da Linguagem*, 17(2):193-223.

Antonio J. D. y Iruskieta M. 2015. A RST e suas aplicações na linguística e no processamento de línguas naturais. En Souza, E. R. F. (Ed.) *Estudos de descrição sociofuncionalista: objetos e abordagens,* paginas 41-73, Ed Lincom.

Castilho, A. T. 2010. *Nova Gramática do Português Brasileiro*. Contexto, S. Paulo.

Hovy, E. 2010. Annotation: a tutorial. En *48th Annual Meeting of the Association for Computational Linguistics.* Uppsala.

Iruskieta, M., A. D. Ilarraza y M. Lersundi. 2015. The annotation of the Central Unit in Rhetorical Structure Trees: A Key Step in Annotating Rhetorical Relations. En *Proceedings of COLING 2014*: Technical Papers, Dublin.

Mann, W. C. y S. A. Thompson. 1988. Rhetorical Structure Theory: toward a functional theory of text organization. *Text*, 8(3):243-281.

Menegassi, R. J. 2011. A Escrita na Formação Docente Inicial: Influências da Iniciação à Pesquisa. *Signum: Estudos da Linguagem*, 14(1):387-419.

Neves, M. H. M. 2000. *Gramática de Usos do Português*. Ed. Unesp, S. Paulo.

Paice, Chris D. 1980. The automatic generation of literature abstracts: an approach based on the identification of self-indicating phrases. *Proceedings of the 3rd annual ACM conference on Research and development in information retrieval*.

Taboada, M. 2006. Discourse Markers as Signals (or Not) of Rhetorical Relations. *Journal of Pragmatics*, 48(4):567-592.

Taboada, M. Implicit and explicit coherence relations. 2009. En J. Renkema (Ed.). *Discourse, of course*. Amsterdam, J. Benjamins, páginas 125-138.

Taboada, M. y C. Habel. 2013. Rhetorical Relations in Multimodal Documents. *Discourse Studies*, 15(1):59-85.

Van Dijk, T. 1980. *Macrostructures*: an Interdisciplinary Study of Global Structures in Discourse, Interaction and Cognition. Lawrence Erlbaum, Hillsdale.