

Discourse markers and coherence relations

Comparison across markers, languages and modalities

Maite Taboada

Simon Fraser University
Vancouver, Canada
mtaboada@sfu.ca

María de los Ángeles Gómez-González

University of Santiago de Compostela
Santiago de Compostela, Spain
mdelosangeles.gomez@usc.es

6th International Contrastive Linguistics Conference
Berlin
September 30-October 2, 2010

Research projects:

1. “A comparative perspective on the Grammar-Discourse Interface in English, with special reference to Coherence and Subjectivity” HUM2007-62220 (Spanish Ministry of Education and Science)
2. “Gramática, discurso e competencia comunicativa en lingua inglesa” INCITE09 204 155 PR (Xunta de Galicia, Spain)

<http://www.usc.es/scimitar>

<http://www.sfu.ca/~mtaboada>

Background

- Discourse markers as signals of coherence relations
 - Large body of research in English (Schiffrin, Fischer, Aijmer, Jucker)
 - Also from a diachronic perspective (Brinton 1996)
- **Fruitful use of contrastive studies** (Altenberg, 2002; Degand, 2009; Degand and Pander Maat, 2003; Fabricius-Hansen, 2005; Knott and Sanders, 1998; Taboada, 2004)
 - How markers are translated
 - Which ones are omitted or added in translation
 - Where they are placed in the sentence/clause

Concession

- Found to be one of the most important relations in vernacular argumentation (Trnavac and Taboada 2010)
 - More frequent in online informal evaluative text (online movie and book reviews)

	Number of words	Number of concessive relations	Concessive relations per 1,000 words
Penn Discourse Treebank	1,000,000	33	0.03
RST Discourse Treebank	176,000	230	1.31
SFU Review corpus	62,000	616	9.94

Coherence relations

- Coherence as fundamental property of discourse
- Coherence created through relations among propositions in discourse
- Relations are the building blocks of discourse
- Our view: coherence within Rhetorical Structure Theory (Mann and Thompson 1988; Taboada and Mann 2006a,b)

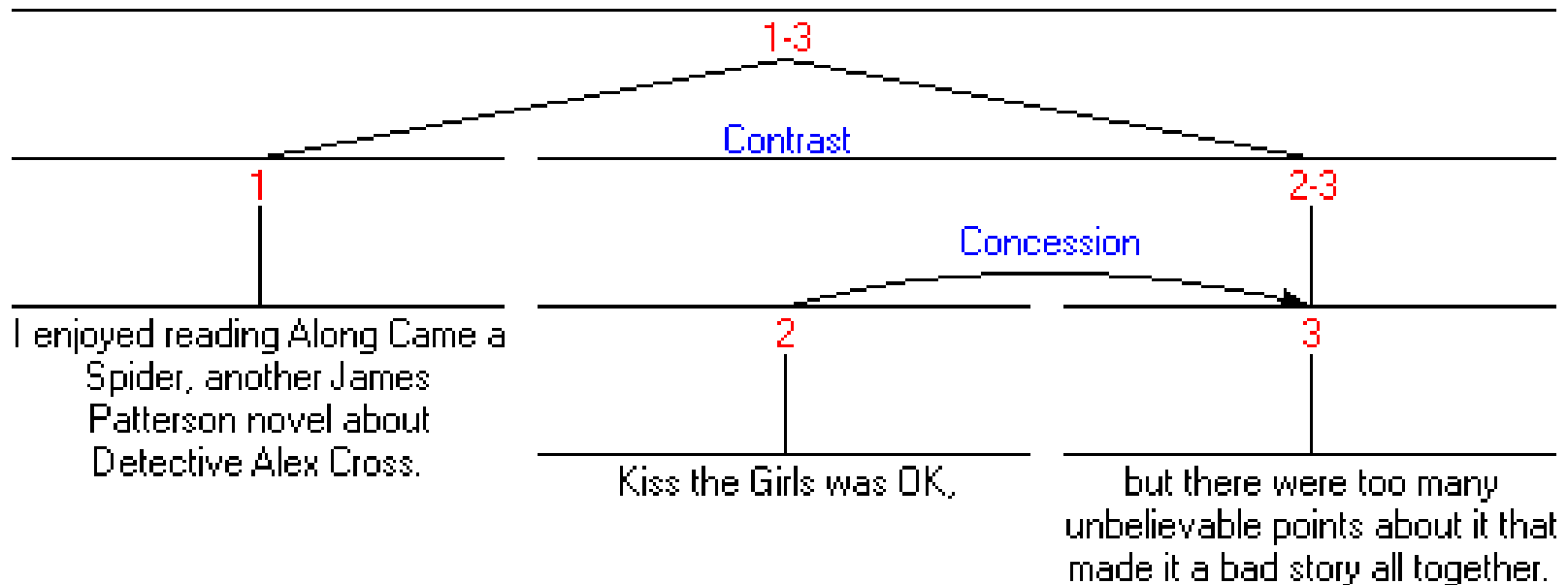
Rhetorical Structure Theory

- Explains text coherence through the relations holding among parts of a text
- Components
 - Text spans
 - Relation schemas
- Hierarchical: spans and schemas might become part of other spans and enter in new relations
- Based on hypotactic and paratactic relations in language

Relations in RST

- Relations hold between two non-overlapping text spans
- Text spans are called *nuclei* and *satellite*
- Most of the relations described hold between a nucleus and a satellite, although there are also multi-nuclear relations
- A relation consists of:
 1. Constraints on the Nucleus,
 2. Constraints on the Satellite,
 3. Constraints on the combination of Nucleus and Satellite,
 4. The Effect.

RST analysis example



Concessive, contrast, adversative relations

- Characteristics

- Introduced by conjunctions considered as concessive
- Can be pre- or post-posed to the main clause or verb (marker-dependent restrictions; pragmatic changes)
- Cannot be replaced by a semantically equivalent adverb

- (1) a. *Although the ending was a happy one, it was also a little sad.* [M, no3]
- b. *La banda sonora es excelente, aunque se repite.* [P, no_2_20]
'The soundtrack is excellent, although repetitive.'

Concession in RST

- **Constraints on the nucleus**
 - The writer has positive regard for the nucleus
- **Constraints on the satellite**
 - The writer is not claiming that S does not hold
 - The writer acknowledges a potential or apparent incompatibility between nucleus and satellite
 - Recognizing the compatibility between nucleus and satellite increases the reader's positive regard for the nucleus
- **Effect**
 - The reader's positive regard for the nucleus is increased

Markers of concession in English

- **Conjunctions and conjuncts**
 - albeit, although, but, but even so, come what may, despite (everything), despite the fact that, even if, even though, even when, even while, howbeit, much as, though, when, whereas, whether, while
- **Sentence adverbials**
 - above all, after all, and even then, anyway, at any cost, even, even yet, for all that, for one thing, however, in any case, in spite of all things, in spite of everything, nevertheless, no matter what, nonetheless, of course, only, over all, rather, regardless, still, too, withal, yet
- **Gerunds introducing subordinate clauses or noun phrases**
 - admitting, allowing that, even supposing, granting (all this), supposing, without considering
- **Prepositional phrases with certain prepositions**
 - against, aside from, distinct from, even after, even before, even as, even with, in contempt of, in defiance of, in spite of, in the face of, notwithstanding, regardless of, without regard to

Markers of concession in Spanish

- **Concessive conjunctions/conjuncts**
 - a pesar de (que), a pesar de todo, a pesar de + Inf., así, aunque, cuando, no obstante, (Conditional / Future +) pero, pese a (que), si bien, sin embargo, (tan) siquiera
- **por + AdjP / AdvP + que-relative clause**
 - por más que, por mucho que
- **para + NP / InfP / que- relative clause**
- **con + NP / InfP / que-relative clause or con lo + AdjP / AdvP + que-relative clause**
- **Gerund**
- **Gerund / Participle / AdjP + y todo**
- **Repetition of two (identical or different) verbal expressions**
 - sea cual sea, (lo) quieras o no
- **Impersonal clausal**
 - (si) bien es cierto, lo cierto es que, la verdad es que, está claro que
- **Adverbs and adverbial expressions**
 - ciertamente, efectivamente
- **Combination of markers**
 - aún así, aún con eso/esto, aún cuando, aún + Gerund, así y todo, pero no obstante, y sin embargo

Corpus

- Two genres or modalities
 - Spoken – CallHome (American English and Spanish)
 - Telephone conversations between relatives and friends (Kingsbury et al. 1997; Wheatley 1996)
 - Written – SFU Review Corpus
 - Reviews of movies, books and consumer products posted online (Taboada et al. 2008)
 - http://www.sfu.ca/~mtaboada/research/SFU_Review_Corpus.html

Corpus statistics

	Written		Spoken	
	English	Spanish	English	Spanish
Texts/conversations	100	100	5	5
Sentences	3,869	5,768	1,708	1,322
Words	62,090	90,338	11,457	8,694

Methodology

- Extract all concessive/contrast/adversative relations based on the marker
- Compare (across languages)
 1. Distribution of concessives across written and spoken texts
 2. Position of concessives with regard to the macro-structure of texts in the Initial, Middle or Final stages of the generic structure
 3. Realization of the concessive relation
 - Concessive marker
 - Order of the nucleus and satellite

Results

- General distribution
- Comparison across genres
- Order of spans
- Multiple markers

General distribution

	Written		Spoken	
	English	Spanish	English	Spanish
Number of relations	326	628	101	24
Relations per 1,000 words	5.25	6.95	8.82	2.76

Comparison across genres

- Written genre

- Concession qualifies an opinion – often by acknowledging the opposite viewpoint
- Dismiss other people's (potential) objections

(2) Reconozco que tiene "pasajes" muy guapos, pero también hay otros (la mayoría) muy pesados y otros que ni siquiera resultan creíbles. [W, P no1_15]

'I acknowledge that it has very good "passages", but also that there are other (most) [that are] very tedious and other that are not even credible.'

(3) Despite what some people think, a kids movie can be good and appeal to adults, such as Toy Story or Space Jam. [W, M no20]

Comparison across genres

- Spoken genre

- Contrast between two situations
- Correction of potential misinterpretations

(4) B: it's hot I mean Tiberius is very hot too but it's dry
and this is humid
I don't know what's worse [S, en_4315]

(5) B: felt a little funny
he felt a little funny in the chest
but that could be a reaction because of the heat [S, en_4315]

Comparison across genres

- Spoken genre
 - Topic management
 - As in the written part of the corpus, acknowledgement of different viewpoints
- (6) B: y sigue igual, así bien despierta, igual a como era mamá
A: ahá
B: sólo que más despierta
y Mónica sin embargo ha crecido un montón. [S, sp_0753]
'and she's the same, like very lively, just like Mom was / only more lively / and Mónica however has grown a lot.'
- (7) B: because it is a regular fulltime job
even though it might not be the great the great school [S, en_4808]

Order of spans

- Coherence relations are presumed to have a canonical order
 - Mann and Thompson (1988)
 - Concession: satellite before nucleus
 - Note that this does not mean “subordinate before main”
- (8) [S] The idea of the film is not new either. It was like a different version of the Sixth Sense, but in a more perverse way. [N]
However, I forgive that because it seemed to work out at the end.
[M, yes17]

Order of spans – Annotation

- Annotation of each relation
 - Satellite-first, nucleus-first, or middle (inserted)
 - Spans may include multiple clauses and sentences
 - Identification of scope not always clear
 - Informal style, run-on sentences
- Markers with canonical order
 - Satellite first
 - *but, despite, however, while*
 - *a pesar de (todo), aún (así), cuando, pero*
 - Nucleus first
 - *even though, even when*
 - *aunque*
- Markers that appear in all (or both) orders
 - *although, even if*
 - *a pesar de (que), por mucho/muy/más (que)*

Order of spans – Results

- Confirmation: canonical order is satellite-first, in both languages and genres

	Written		Spoken	
	English	Spanish	English	Spanish
Nucleus first	31 (9.51%)	108 (17.20%)	1 (0.99%)	1 (4.17%)
Satellite first	294 (90.18%)	504 (80.25%)	100 (99.01%)	23 (95.83%)
Middle	1 (0.31%)	16 (2.55%)	-	-

Multiple markers

- Multiple markers common in both languages (9)
 - Sometimes in an ungrammatical construction (10)
- (9) *A pesar* estar destinada a un público claramente infantil, lo cierto es que con *Ratatouille* pasa lo que pasa con muchas otras películas de animación... [P, yes_5_7]
- ‘Despite (the fact that) it is clearly geared towards a children’s audience, the truth is that with *Ratatouille* you get what you get with many animation movies...’
- (10) *Aunque* todas sus amigas y familia la dijeran que era lo mejor que la había podido pasar, *pero* ella seguía dando vueltas a la cabeza si *Iain* aun amaba a su ex. [B, yes_5_15]
- ‘*Although* all her friends and family told her that it was the best thing that could have ever happened to her, *but* she was still considering whether *Iain* still loved his ex.’

Conclusions

- Methodology for contrastive study of coherence relations and discourse markers
- Concession a prevalent relation in informal argumentative text
- Differences in usage are more pronounced across genres than across languages
 - In speech, concession is used to correct misunderstandings and contrast situations
 - In writing, used to qualify opinions
 - Distribution very similar across languages
- Confirmation that most frequent order is satellite first

Future work

- Larger corpus (especially speech)
- Distribution of concession across stages in discourse
- Differences in ordering
 - When the 'canonical' ordering is not used, what are the pragmatic or semantic implications?

References (I)

- Abraham, Werner. (1979). But. *Studia Linguistica*, 33 (2), 89-119.
- Alarcos Llorach, Emilio. (1994). *Gramática de la lengua española*. Madrid: Espasa-Calpe.
- Altenberg, Bengt. (2002). Concessive connectors in English and Swedish. In H. Hasselgård, S. Johansson, B. Behrens and C. Fabricius-Hansen (Eds.), *Information Structure in a Cross-Linguistic Perspective* (pp. 21-43). Amsterdam: Rodopi.
- Biber, Douglas, Stig Johansson, Geoffrey Leech, Susan Conrad and Edward Finegan. (1999). *Longman Grammar of Spoken and Written English*. Harlow, Essex: Pearson Education.
- Brinton, Laurel J. (1996). *Pragmatic Markers in English: Grammaticalization and Discourse Functions*. Berlin: Mouton de Gruyter.
- Couper-Kuhlen, Elizabeth and Sandra A Thompson. (2000). Concessive patterns in conversation. In E. Couper-Kuhlen and B. Kortmann (Eds.), *Cause, Condition, Concession, Contrast: Cognitive and Discourse Perspectives* (pp. 381-410). Berlin: Mouton de Gruyter.
- Crevels, Mily. (2000a). *Concession: A Typological Study*. Unpublished Ph.D. dissertation, University of Amsterdam, Amsterdam.
- Crevels, Mily. (2000b). Concessives on different semantic levels: A typological perspective. In E. Couper-Kuhlen and B. Kortmann (Eds.), *Cause, Condition, Concession, Contrast: Cognitive and Discourse Perspectives* (pp. 313-339). Berlin: Mouton de Gruyter.
- Degand, Liesbeth. (2009). Describing polysemous discourse markers: What does translation add to the picture? In S. Slembrouch, M. Taverniers and M. Van Herreweghe (Eds.), *From will to well. Studies in Linguistics offered to Anne-Marie Simon-Vandenbergen*. Gent: Academia Press.
- Degand, Liesbeth and Henk Pander Maat. (2003). A contrastive study of Dutch and French causal connectives on the Speaker Involvement Scale. In A. Verhagen and J. van de Weijer (Eds.), *Usage based approaches to Dutch* (pp. 175-199). Utrecht: LOT.
- Di Tullio, Ángela. (1997). *Manual de gramática del español*. Buenos Aires: Edicial.

References (II)

- Fabricius-Hansen, Cathrine. (2005). Elusive connectives: A case study on the explicitness dimension of discourse coherence. *Linguistics*, 43 (1), 17-48.
- Fischer, Kerstin (Ed.). (2006). *Approaches to Discourse Particles*. Amsterdam: Elsevier.
- Flamenco García, Luis. (1999). Las construcciones concesivas y adversativas. In V. Demonte (Ed.), *Gramática descriptiva de la lengua española* (Vol. 3: Entre la oración y el discurso. Morfología, pp. 3805-3877). Madrid: Espasa.
- Fuentes Rodríguez, Catalina. (1998). *Las construcciones adversativas*. Madrid: Arco.
- Gili Gaya, Samuel. (1955). *Curso Superior de Sintaxis Española*. Barcelona: Spes.
- Gutiérrez Ordóñez, Salvador. (1977-1978). A propósito de 'Cláusulas y oraciones'. *Archivum*, XXVII-XXVIII, 529-547.
- Hernández Alonso, César. (1995). *Nueva sintaxis de la lengua española*. Salamanca: Ediciones Colegio de España.
- Huddleston, Rodney and Geoffrey K. Pullum. (2002). *The Cambridge Grammar of the English Language*. Cambridge: Cambridge University Press.
- Izutsu, Mitsuko Narita. (2008). Contrast, concessive, and corrective: Toward a comprehensive study of opposition relations. *Journal of Pragmatics*, 40 (4), 646-675.
- Kingsbury, Paul, Stephanie Strassel, Cynthia McLemore and Robert McIntyre. (1997). CallHome American English Transcripts, LDC97T14 [Corpus]. Philadelphia, PA: Linguistic Data Consortium.
- Knott, Alistair and Ted Sanders. (1998). The classification of coherence relations and their linguistic markers: An exploration of two languages. *Journal of Pragmatics*, 30, 135-175.
- Kovacci, Ofelia. (1992). *El comentario gramatical II*. Madrid: Arco.
- Lavacchi, Leonardo and Carlota Nicolás. (1994). Oraciones de 'aunque' y 'pero'. *Verba*, 21, 257-278.
- López García, Ángel. (1999). Relaciones paratácticas e hipotácticas. In V. Demonte (Ed.), *Gramática descriptiva de la lengua española* (Vol. 3: Entre la oración y el discurso. Morfología, pp. 3507-3547). Madrid: Espasa.

References (III)

- Mann, William C. and Sandra A. Thompson. (1988). Rhetorical Structure Theory: Toward a functional theory of text organization. *Text*, 8 (3), 243-281.
- Molina Redondo, José Andrés de. (1985). En torno a la oración 'compuesta' en español. In *Philologica Hispaniensi in honorem Manuel Alvar II* (pp. 513-527). Madrid: Gredos.
- Moya Corral, Juan Antonio. (1996). *Los mecanismos de la interordinación: a propósito de 'pero' y 'aunque'*. Granada: Universidad de Granada.
- Narbona Jiménez, Antonio. (1990). *Las subordinadas adverbiales impropias en español*. Málaga: Agora.
- Quirk, Randolph, Sidney Greenbaum, Geoffrey Leech and Jan Svartvik. (1985). *A Comprehensive Grammar of the English Language*. London: Longman.
- Real Academia Española. (1931). *Gramática de la lengua española*. Madrid: Espasa Calpe.
- Rivarola, José Luis. (1976). *Las conjunciones concesivas en español medieval y clásico*. Tübingen: Max Niemeyer.
- Rudolph, Elisabeth. (1996). *Contrast: Adversative and Concessive Relations and their Expressions in English, German, Spanish, Portuguese on Sentence and Text Level*. Berlin: Mouton de Gruyter.
- Schiffrin, Deborah. (1987). *Discourse Markers*. Cambridge: Cambridge University Press.
- Taboada, Maite. (2004). Rhetorical relations in dialogue: A contrastive study. In C. L. Moder and A. Martinovic-Zic (Eds.), *Discourse across Languages and Cultures* (pp. 75-97). Amsterdam and Philadelphia: John Benjamins.
- Taboada, Maite. (2008). SFU Review Corpus [Corpus]. Vancouver: Simon Fraser University, http://www.sfu.ca/~mtaboada/research/SFU_Review_Corpus.html.
- Traugott, Elizabeth Closs. (1986). On the origins of 'and' and 'but' connectives in English. *Studies in Language*, 10 (1), 137-150.
- Trnavac, Radoslava and Maite Taboada. (2010). *The contribution of nonveridical rhetorical relations to evaluation in discourse*. Paper presented at the 37th International Systemic Functional Congress, Vancouver, Canada.
- Wheatley, Barbara. (1996). CallHome Spanish Transcripts, LDC96T17 [Corpus]. Philadelphia, PA: Linguistic Data Consortium

Discourse markers and coherence relations

Comparison across markers, languages and modalities

Maite Taboada
mtaboada@sfu.ca

María de los Ángeles Gómez-González
mdelosangeles.gomez@usc.es

Research group:

<http://www.usc.es/scimitar>

<http://www.sfu.ca/~mtaboada>