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

Research projects:
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

6th International Contrastive Linguistics Conference
Berlin
September 30-October 2, 2010
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)
  - 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 (Trnavač and Taboada 2010)
  - More frequent in online informal evaluative text (online movie and book reviews)

<table>
<thead>
<tr>
<th></th>
<th>Number of words</th>
<th>Number of concessive relations</th>
<th>Concessive relations per 1,000 words</th>
</tr>
</thead>
<tbody>
<tr>
<td>Penn Discourse Treebank</td>
<td>1,000,000</td>
<td>33</td>
<td>0.03</td>
</tr>
<tr>
<td>RST Discourse Treebank</td>
<td>176,000</td>
<td>230</td>
<td>1.31</td>
</tr>
<tr>
<td>SFU Review corpus</td>
<td>62,000</td>
<td>616</td>
<td>9.94</td>
</tr>
</tbody>
</table>
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

I enjoyed reading *Along Came a Spider*, another James Patterson novel about Detective Alex Cross.

1. Contrast

2. Concession

3. "Kiss the Girls was OK, but there were too many unbelievable points about it that made it a bad story all together."
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](http://www.sfu.ca/~mtaboada/research/SFU_Review_Corpus.html)
# Corpus statistics

<table>
<thead>
<tr>
<th></th>
<th>Written</th>
<th></th>
<th>Spoken</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>English</td>
<td>Spanish</td>
<td>English</td>
<td>Spanish</td>
</tr>
<tr>
<td>Texts/conversations</td>
<td>100</td>
<td>100</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>Sentences</td>
<td>3,869</td>
<td>5,768</td>
<td>1,708</td>
<td>1,322</td>
</tr>
<tr>
<td>Words</td>
<td>62,090</td>
<td>90,338</td>
<td>11,457</td>
<td>8,694</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th></th>
<th>Written</th>
<th></th>
<th>Spoken</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>English</td>
<td>Spanish</td>
<td>English</td>
<td>Spanish</td>
</tr>
<tr>
<td>Number of relations</td>
<td>326</td>
<td>628</td>
<td>101</td>
<td>24</td>
</tr>
<tr>
<td>Relations per 1,000 words</td>
<td>5.25</td>
<td>6.95</td>
<td>8.82</td>
<td>2.76</td>
</tr>
</tbody>
</table>
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

<table>
<thead>
<tr>
<th>Order Type</th>
<th>Written</th>
<th>Spoken</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>English</td>
<td>Spanish</td>
</tr>
<tr>
<td>Nucleus first</td>
<td>31 (9.51%)</td>
<td>108 (17.20%)</td>
</tr>
<tr>
<td>Satellite first</td>
<td>294 (90.18%)</td>
<td>504 (80.25%)</td>
</tr>
<tr>
<td>Middle</td>
<td>1 (0.31%)</td>
<td>16 (2.55%)</td>
</tr>
</tbody>
</table>
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 aún 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)


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