Pero, Bueno, Pues

TESTING NEW METHODOLOGICAL APPROACHES FOR THE IDENTIFICATION AND DISAMBIGUATION OF DISCOURSE MARKERS IN SPOKEN SPANISH

Zoé Broisson
UCREL Research Seminar
25th October 2018
About me

Honours thesis: Cuantificación de la armonía vocálica en español andaluz oriental

Master’s thesis (work in progress): Discourse markers: For the speaker or for the hearer?
1. Introduction
   What are DMs? Why study them?

2. Previous taxonomies

3. This study

4. Methods

5. Results

6. Conclusions
1. Introduction
But actually, what are discourse markers?

“sequentially dependent elements which bracket units of talk” – Schiffrin 1987: 31

“a class of expressions, each of which signals how the speaker intends the basic message that follows to relate to prior discourse” – Fraser 1990: 387

“A [discourse marker] is a phonologically short item that is not syntactically connected to the rest of the clause (i.e., is parenthetical), and has little or no referential meaning but serves pragmatic or procedural purposes” – Brinton 2008: 1

actually, I mean, look, by the way, well, yeah, for example, however

pero, bueno, pues, vale, la verdad, porque, por ejemplo, además
What do DMs do? Why study them?

- **Structure discourse**
  (Crible & Zufferey, 2015: 14)

  - Relations
  - Interactions

- **Self-monitor**
  our communicative
  (pragmalinguistic) competence
  (Celce-Murcia & Olshtain, 2000: 493)

- **Interpret information**
  (speech: metadiscursive instructions)
  (Brinton, 2008; Hansen, 2008)

- **Implications for second language teaching and learning**
  (Svartvik, 1980: 171; Wei, 2011)
So... What is the issue?

Because of the **formal heterogeneity** of DMs, authors usually struggle to **categorize** them.

Crible and Zufferey (2015: 15)
« It has become standard in any overview article or chapter on DMs to state that reaching agreement on what makes a DM is as good as impossible, be it alone on terminological matters »

- Degand, Cornillie, Pietrandrea (2013: 5)
I mean, issues?

Pragmatic markers
- Brinton 1996; González 2005

Discourse markers
- Lenk 1998; Schiffrin 1987

Discourse particles
- Pragmatic expressions
- Modal particles

Discourse operators
- Discourse connectives

Function(s)?

Coherence Theory

Relevance Theory

Rouchota 1996
Blakemore 1987
DMs in the literature

Need for an open-class definition and categorisation!
2. Previous taxonomies
Penn Discourse Tree Bank 2.0 (Prasad et al. 2008)

- Wall Street Journal (WSJ) corpus
- 40,000+ discourse relations
- Discourse connectives (*because*, *after*, *so*, *when*, *if*, *but*, *however*)
González (2005)

- English and Catalan corpus of 40 oral narratives (20-20)
- Pragmatic markers and discourse coherence relations (*anyway, I mean, well, so...*)
  - 168 markers in English
  - 433 markers in Catalan

Speech-based
Martín Zorraquino & Portolés (1999)

Marcadores Conversacionales
("Conversational markers")
- Evidencia/Certeza (Confirmation/Manifestation of certainty – Epistemic modality)
- Aceptación (Agreement – Deontic modality)
- Alteridad (’Otherness’ – Monitoring the relationship with the interlocutor)
- Metadiscursivos (Metadiscursive function, structure the conversation)

Operadores Argumentativos
("Argumentative operators")
- De resfuerzo argumentativo (Reinforce a previously formulated argument, e.g. de hecho ‘in fact’)
- De concreción (Present an example)

Reformuladores
("Reformulation markers")
- Explicativos (Reformulation/specification)
- De rectificación (Correct a previous formulation)
- De distanciamiento (Convey the irrelevance of a previous formulation)
- Recapitulativos (Recapitulate previous information or present a conclusion)

Conectores
("Connectors")
- Aditivos (Addition)
- Consecutivos (Consequence)
- Contraargumentativos (Contrast)

Estructuradores de la Información
("Information organizers")
- Comentadores (Topic-shifting)
- Ordenadores (Ordering)
- Digresores (Digression)

Speech & Writing
Why worry about reliability & replicability?

QUALITY & EXCHANGE OF RESEARCH

In this particular context...

• Implicit or underspecified information
• Subjectivity = Interpretation = Low inter-rater agreement scores

(Spooren & Degand 2010)
1. Critical review of the literature and selection of the most recurrent and relevant criteria for DM identification

2. Intuitive selection of DM candidate tokens in a balanced bilingual corpus (FR-EN) and confrontation of identified criteria with description in context - Which criteria are stronger or weaker predictors of DM membership?

3. Elaboration of a definition and coding scheme

4. Annotation experiments and revision of the scheme for replicability
Crible’s (2017:106) definition

“DMs are a grammatically heterogeneous, multifunctional type of pragmatic markers, hence constraining the inferential mechanisms of interpretation. Their specificity is to function on a metadiscursive level as procedural cues to situate the host unit in a co-built representation of on-going discourse”

“I claim that any categorical definition is only useful insofar as it is endorsed by an empirical model of identification and annotation”
Crible (2017:106-107)

**Syntactic features**
- DMs are optional
- DMs are relatively mobile in the utterance
- DMs belong to diverse grammatical classes
- DMs have a fixed form as a result of grammaticalisation and high-frequency use
- DMs have a variable scope
- The host unit must be autonomous both syntactically and semantically

**Functional features**
- DMs have a procedural meaning
- DMs are multifunctional
  - A single member can perform different functions in different contexts (i.e. DMs are polyfunctional)
  - A single member can perform different functions simultaneously in the same context (i.e. DMs can be polysemous)

**Interjections, question tags**
Crible (2014)

<table>
<thead>
<tr>
<th>IDEATIONAL</th>
<th>RHETORICAL</th>
<th>SEQUENTIAL</th>
<th>INTERPERSONAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>cause</td>
<td>motivation</td>
<td>punctuation</td>
<td>monitoring</td>
</tr>
<tr>
<td>consequence</td>
<td>conclusion</td>
<td>opening</td>
<td>face-saving</td>
</tr>
<tr>
<td>concession</td>
<td>opposition</td>
<td>boundary</td>
<td>disagreeing</td>
</tr>
<tr>
<td>concession</td>
<td>specification</td>
<td>topic-resuming</td>
<td>agreeing</td>
</tr>
<tr>
<td>contrast</td>
<td>relevance</td>
<td>topic-shifting</td>
<td>elliptical</td>
</tr>
<tr>
<td>alternative</td>
<td>emphasis</td>
<td>quoting</td>
<td></td>
</tr>
<tr>
<td>condition</td>
<td>comment</td>
<td>addition</td>
<td></td>
</tr>
<tr>
<td>temporal</td>
<td>approximation</td>
<td>enumeration</td>
<td></td>
</tr>
<tr>
<td>exception</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**Objective**  
**Subjective**  
**Intersubjective**
How to improve reliability?

- Make categories independent
- Reduce number of categories
  - Bite-size procedural steps

(Spooren & Degand 2010)
Crible & Degand (2017b)

<table>
<thead>
<tr>
<th>IDEATIONAL</th>
<th>RHETORICAL</th>
<th>SEQUENTIAL</th>
<th>INTERPERSONAL</th>
</tr>
</thead>
<tbody>
<tr>
<td>[addition]</td>
<td>[alternative]</td>
<td>[cause]</td>
<td>[concession]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[condition]</td>
<td>[consequence]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[contrast]</td>
<td>[punctuation]</td>
</tr>
<tr>
<td></td>
<td></td>
<td>[specification]</td>
<td>[temporal]</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>[topic]</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Objective</th>
<th>Subjective</th>
<th>Intersubjective</th>
</tr>
</thead>
</table>

French and English
(Crible & Zufferey 2015)

French, English & Polish
(Crible & Degand 2017b)

Belgian French SL
(Gabbaró-López 2017)
3. This study
Why (yet) another study?

- Make categories independent
- Reduce number of categories
- Bite-size procedural steps

French and English
(Crible & Zufferey 2015)

French, English & Polish
(Crible & Degand 2017b)

Belgian French SL
(Gabbaró-López 2017)

Spanish?
Research question

Will the use of Crible and Degand’s (2017b) revised version of Crible’s (2017) taxonomy in combination with a step-wise annotation protocol allow for the consistent disambiguation of discourse markers in a selected sample of spoken peninsular Spanish?
4. Methods
Corpus data

Sample from the spoken Spanish component of the Backbone corpora

- 4 face-to-face interviews, each between 2 adult speakers of peninsular Spanish
- 2 males (interviewees), 3 females (1 interviewer + 2 interviewees)
- Audio available for annotation

<table>
<thead>
<tr>
<th>CORPUS SAMPLE</th>
<th>NUMBER OF WORD TOKENS</th>
<th>LENGTH (IN MINUTES)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interview 1* (bb_es008_rosa)</td>
<td>1159</td>
<td>5:12</td>
</tr>
<tr>
<td>Interview 2* (bb_es0012_alejandropena)</td>
<td>1221</td>
<td>6:26</td>
</tr>
<tr>
<td>Interview 3 (bb_es0021_irene)</td>
<td>2325</td>
<td>14:05</td>
</tr>
<tr>
<td>Interview 4 (bb_es005_santiago)</td>
<td>3618</td>
<td>16:41</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>8323</strong></td>
<td><strong>42:24</strong></td>
</tr>
</tbody>
</table>
Annotation: 3 steps

Software: EXMARaLDA (Schmidt & Wörner, 2012)

• **Step 1:** chronological manual annotation of DMs according to the functional definition (no closed list)

• **Step 2:** chronological manual annotation of domains and then functions, or vice-versa

• **Step 3:** chronological manual annotation of domains and then functions, or vice-versa (same identified DMs) at a 2-3 weeks’ interval

No double-tagging
<table>
<thead>
<tr>
<th>TXT</th>
<th>sí</th>
<th>de</th>
<th>universidad</th>
<th>no</th>
<th>es</th>
<th>una</th>
<th>suerte</th>
<th>María</th>
<th>Y</th>
<th>que</th>
<th>suerte</th>
<th>te</th>
<th>lleva</th>
<th>más,</th>
<th>la</th>
<th>investigación</th>
<th>o</th>
<th>la</th>
</tr>
</thead>
<tbody>
<tr>
<td>MARÍA [words]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MARÍA [DM]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MARÍA [DM Domain]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>MARÍA [DM Function]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SANTIAGO [words]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SANTIAGO [DM]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SANTIAGO [DM Domain]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>SANTIAGO [DM Function]</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

(Comments)
Annotation of domains
Annotation of functions

**FUNCTIONS**

*Instructions*

Try to mentally take the annotated DM out of the highlighted sentence to answer the following questions. Do keep the original DM in mind during your interpretation, as to preserve the meaning of the sentence. Please note: ‘Segment 1’ and ‘Segment 2’ refer to the units of talk situated on either side of the isolated Discourse Marker (henceforth ‘DM’).

1. Can you replace the original DM by *además* [Eng.: *in addition*], AND/OR does the original DM introduce new information?
   - Yes, then the function of this DM is ADDITION. You have finished analysing the function of the DM.
   - No. Proceed to question 2.

2. Can you paraphrase the original sentence as follows: ‘(either) [segment 1] OR [segment 2]’?
   - Yes, then the function of this DM is ALTERNATIVE. You have finished analysing the function of the DM.
   - No. Proceed to question 3.

3. Can you use *because, due to or if* to connect the segments?
   - *Because* or *due to*, then the function of this DM is CAUSAL. You have finished analysing the function of the DM.
   - *If*, then the function of this DM is CONDITIONAL. You have finished analysing the function of the DM.
   - If neither can be used, proceed to question 4.

Substitution and paraphrasing tests inspired by Scholman et al. (2016)
5. Results
Identified DMs

<table>
<thead>
<tr>
<th>CORPUS SAMPLE</th>
<th>TOTAL NUMBER OF WORD TOKENS</th>
<th>TOTAL NUMBER OF DM TOKENS</th>
<th>PROPORTION OF DMs</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interview 1</td>
<td>1159</td>
<td>79</td>
<td>6.81%</td>
</tr>
<tr>
<td>(bb_es008_rosa)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interview 2</td>
<td>1221</td>
<td>127</td>
<td>10.40%</td>
</tr>
<tr>
<td>(bb_es0012_alejandropena)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interview 3</td>
<td>2325</td>
<td>184</td>
<td>7.91%</td>
</tr>
<tr>
<td>(bb_es0021_irene)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interview 4</td>
<td>3618</td>
<td>347</td>
<td>9.59%</td>
</tr>
<tr>
<td>(bb_es005_santiago)</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>8323</strong></td>
<td><strong>737</strong></td>
<td><strong>8.85%</strong></td>
</tr>
</tbody>
</table>
Functional distribution

SEQ 37%

RHE 25%

INT 26%

IDE 12%

ADD 17%

ALT 3%

CAU 3%

CONC 5%

COND 1%

CONS 7%

CONT 4%

PUNCT 35%

TOPIC 6%

SPE 14%

TEMP 5%
Results in context of Crible & Degand (2017b)
Formal distribution

<table>
<thead>
<tr>
<th>Rank</th>
<th>DM Type</th>
<th>Number of Occurrences</th>
<th>Proportion of the Overall Number of DM Tokens (737)</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>y</td>
<td>177</td>
<td>24.1%</td>
</tr>
<tr>
<td>2</td>
<td>pues</td>
<td>10</td>
<td>177</td>
</tr>
<tr>
<td>3</td>
<td>no</td>
<td>48</td>
<td>6.5%</td>
</tr>
<tr>
<td>4</td>
<td>pero</td>
<td>44</td>
<td>6.0%</td>
</tr>
<tr>
<td>5</td>
<td>bueno</td>
<td>42</td>
<td>5.7%</td>
</tr>
<tr>
<td>6</td>
<td>entonces</td>
<td>30</td>
<td>4.1%</td>
</tr>
<tr>
<td>7</td>
<td>es decir</td>
<td>21</td>
<td>2.8%</td>
</tr>
<tr>
<td>8</td>
<td>o</td>
<td>21</td>
<td>2.8%</td>
</tr>
<tr>
<td>9</td>
<td>por ejemplo</td>
<td>21</td>
<td>2.8%</td>
</tr>
<tr>
<td>10</td>
<td>Porque</td>
<td>20</td>
<td>2.7%</td>
</tr>
<tr>
<td>TOTAL</td>
<td></td>
<td>501</td>
<td>68.0%</td>
</tr>
</tbody>
</table>

![Bar chart showing distribution of DM Types and Functions Types](chart.png)
Round 1 vs. Round 2: Domains

Number of DM tokens

IDE  | INT  | RHE  | SEQ  
---   | ---   | ---  | ---   
92    | 189   | 188  | 268   
93    | 188   | 185  | 271   

Annotation Round 1  
Annotation Round 2
Round 1 vs. Round 2: Functions

The chart compares the number of DM tokens for various functions between Round 1 and Round 2. The bars represent the number of tokens for each function, with separate colors for Round 1 and Round 2.
# Intra-rater agreement

<table>
<thead>
<tr>
<th>Corpus Sample</th>
<th>Number of Selected DM Tokens</th>
<th>Number of Disagreements</th>
<th>Agreement Score</th>
<th>Number of Disagreements</th>
<th>Agreement Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>Interview 1 (bb_es008_rosa)</td>
<td>50</td>
<td>7</td>
<td>86%</td>
<td>7</td>
<td>86%</td>
</tr>
<tr>
<td>Interview 2</td>
<td></td>
<td>50</td>
<td>17</td>
<td>6</td>
<td>88%</td>
</tr>
<tr>
<td>(bb_es0012_alejandropena)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Interview 3 (bb_es0021_irene)</td>
<td></td>
<td>50</td>
<td>10</td>
<td>8</td>
<td>84%</td>
</tr>
<tr>
<td>Interview 4</td>
<td></td>
<td>50</td>
<td>9</td>
<td>4</td>
<td>92%</td>
</tr>
<tr>
<td>(bb_es005_santiago)</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>200</strong></td>
<td><strong>43</strong></td>
<td><strong>78.5%</strong></td>
<td><strong>25</strong></td>
<td><strong>87.5%</strong></td>
</tr>
</tbody>
</table>
## Disagreement analysis: Domains

<table>
<thead>
<tr>
<th>Disagreement Pair</th>
<th>Number of Occurrences</th>
<th>Proportion of Overall Number of Domain Disagreements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sequential-Interpersonal</td>
<td>12</td>
<td>27.9%</td>
</tr>
<tr>
<td>Sequential-Rhetorical</td>
<td>12</td>
<td>27.9%</td>
</tr>
<tr>
<td>Rhetorical-Ideational</td>
<td>8</td>
<td>18.6%</td>
</tr>
<tr>
<td>Sequential-Ideational</td>
<td>7</td>
<td>16.3%</td>
</tr>
<tr>
<td>Rhetorical-Interpersonal</td>
<td>4</td>
<td>9.3%</td>
</tr>
<tr>
<td><strong>Total</strong></td>
<td><strong>43</strong></td>
<td><strong>100.0%</strong></td>
</tr>
</tbody>
</table>

Disagreements at domain-level only

![Bar chart showing number of disagreements by domain pair, with categories for Sequential-Interpersonal, Sequential-Rhetorical, Rhetorical-Ideational, Sequential-Ideational, and Rhetorical-Interpersonal. Each bar is divided to show the number of disagreements and the number of disagreements at domain-level only.]

- **SEQ INT**: 11 disagreements, 1 at domain-level only
- **SEQ RHE**: 7 disagreements, 5 at domain-level only
- **RHE IDE**: 6 disagreements, 2 at domain-level only
- **SEQ IDE**: 5 disagreements, 2 at domain-level only
- **RHE INT**: 3 disagreements, 1 at domain-level only
[...] Y un día normal de mi vida (short pause) la verdad es que acabo de empezar y, más o menos, no hay una rutina diaria así muy normal, la verdad [...]  
(bb_es0021_irene – 00:42.25)  

[...] And regarding how a normal day of my life goes about (short pause) well, you know, I just started in this new job and I don’t really have a normal routine like you described [...]
Disagreement analysis: Functions

<table>
<thead>
<tr>
<th>Disagreement Pair</th>
<th>Number of Occurrences</th>
<th>Proportion of Overall Number of Function Disagreements</th>
</tr>
</thead>
<tbody>
<tr>
<td>Specification-Addition</td>
<td>5</td>
<td>20.0%</td>
</tr>
<tr>
<td>Punctuation-Addition</td>
<td>5</td>
<td>20.0%</td>
</tr>
<tr>
<td>Consequence-Addition</td>
<td>3</td>
<td>12.0%</td>
</tr>
<tr>
<td>Specification-Temporal</td>
<td>2</td>
<td>8.0%</td>
</tr>
<tr>
<td>Temporal-Causal</td>
<td>1</td>
<td>4.0%</td>
</tr>
<tr>
<td>Consequence-Punctuation</td>
<td>1</td>
<td>4.0%</td>
</tr>
<tr>
<td>Consequence-Contrast</td>
<td>1</td>
<td>4.0%</td>
</tr>
<tr>
<td>Contrast-Concession</td>
<td>1</td>
<td>4.0%</td>
</tr>
<tr>
<td>Punctuation-Topic-shifting</td>
<td>1</td>
<td>4.0%</td>
</tr>
<tr>
<td>Addition-Topic-shifting</td>
<td>1</td>
<td>4.0%</td>
</tr>
<tr>
<td>Causal-Consequence</td>
<td>1</td>
<td>4.0%</td>
</tr>
<tr>
<td>Punctuation-Contrast</td>
<td>1</td>
<td>4.0%</td>
</tr>
<tr>
<td>Addition-Temporal</td>
<td>1</td>
<td>4.0%</td>
</tr>
<tr>
<td>Specification-Punctuation</td>
<td>1</td>
<td>4.0%</td>
</tr>
<tr>
<td><strong>TOTAL</strong></td>
<td><strong>25</strong></td>
<td><strong>100.0%</strong></td>
</tr>
</tbody>
</table>

Disagreements at function-level only

- SPE ADD: 3
- PUNCT ADD: 2
- CONS ADD: 2
- SPE TEMP: 3
- TEMP CAU: 2
- CONS...: 1
- CONS...: 1
- CONT...: 1
- PUNCT...: 1
- ADD TOPIC: 1
- CAU CONS: 1
- PUNCT...: 1
- ADD TEMP: 1
- SPE PUNCT: 1
Example

[...] mi padre es empresario, tiene una empresa de transporte (short pause) y, bueno, pues siempre me ha gustado mucho el mundo de la empresa [...]

(bb_es0012_alejandropena – 00:29:66)

[...] my father is a businessman, he has a transport/shipping company (short pause) and, well, actually I’ve always been attracted to the business world [...]

Discussion

• Function disambiguation is quite consistent
  • SEQ domain = less reliable due to new combinations (Crible & Degand 2017b)
  • Agreement concentrated over ‘Objective’ end of continuum
  • Less cognitively ‘costly’ to annotate?

• Domain annotation is (a little bit) less consistent
  • More variation in high-frequency, polyfunctional DM ‘y’
  • ADD vs. SPE vs. ALT?
  • Difficult to identify functions in multi-DM sequences
  • ‘Pero, bueno, pues, la verdad es que’

• Strong vs. weak DMs?
Suggestions?

Train,
Hierarchise &
Systematise

Reformulation?
Suggestions?

8. Can you replace the original DM by ‘such as’, ‘like’, ‘for example’ AND/OR does the original DM introduce new or more information on the topic discussed by the speaker and the hearer?

☐ Yes, then the function of this DM is SPECIFICATION. You have finished analysing the function of this DM.
☐ No. Proceed to question 8.

*UPDATED: 8. Can you replace the original DM by ‘such as’, ‘like’, ‘for example’ AND/OR does the original DM express reformulation (can you substitute it with ‘in other words’, ‘in short’ for example)?

☐ Yes, then the function of this DM is SPECIFICATION. You have finished analysing the function of this DM.
☐ No. Proceed to question 8.
6. Conclusion
Conclusion

“Further operationalization to enhance the replicability of the functional taxonomy is particularly needed, along with intra-annotator reliability to check for consistency during the annotation process.”

– Crible & Degand (2017a)

Step-wise protocol = Higher agreement (to be tested in larger inter-annotator studies)

Crible’s (2017) Taxonomy = applicable to spoken peninsular Spanish

Raise awareness about methodological practices?

- Replicate study with more annotators
- More modalities (gestures?)
- Expert vs. Naïve coders?
- Transcriptions of speech only?
- Native vs. non-native speakers?
- NLP?
References


Gabbard-López, S. (2017, November). Defining the category of discourse markers in sign languages. Does the functional paradigm apply to discourse markers in a language of a different modality? Presented at the 5th International Conference "Discourse Markers in Romance Languages:
References


THANK YOU FOR YOUR ATTENTION!

ANY QUESTIONS?

Zoe Broisson
Zoe_Brsn
zoe.broisson@student.uclouvain.be