

Marqueurs discursifs en traduction à l'oral

Les sous-titres de film sont-ils le reflet de
la conversation spontanée?

Liesbeth Degand
Université catholique de Louvain

Présentation sur: <http://hdl.handle.net/2078.1/158799>

4th International Symposium on Discourse Markers in Romance Languages; Heidelberg 6-9 May 2015

Discourse Markers in the Translation of Speech

Do film subtitles reflect spontaneous spoken conversation?

Liesbeth Degand

Université catholique de Louvain

Download this presentation from: <http://hdl.handle.net/2078.1/158799>

Discourse markers

- linguistic expressions whose primary function lies at the discourse level, i.e. relating their host utterance to the discourse situation.
- Can play a threefold role:
 - contributing to the discourse organization (textual coherence),
 - to the speaker/hearer interaction (interpersonal meanings),
 - and/or to speaker attitudes

Discourse Markers as a thing of speech

- DMs are frequent and varied in spoken language
- The more spontaneous the language, the more variation and the higher their frequency (Georgakopoulou & Goutsos, 1998)
- **Working hypothesis:** DM use is an index of a discourse's degree of orality/spoken-likeness (see also, Schlamberger Brezar 2012)

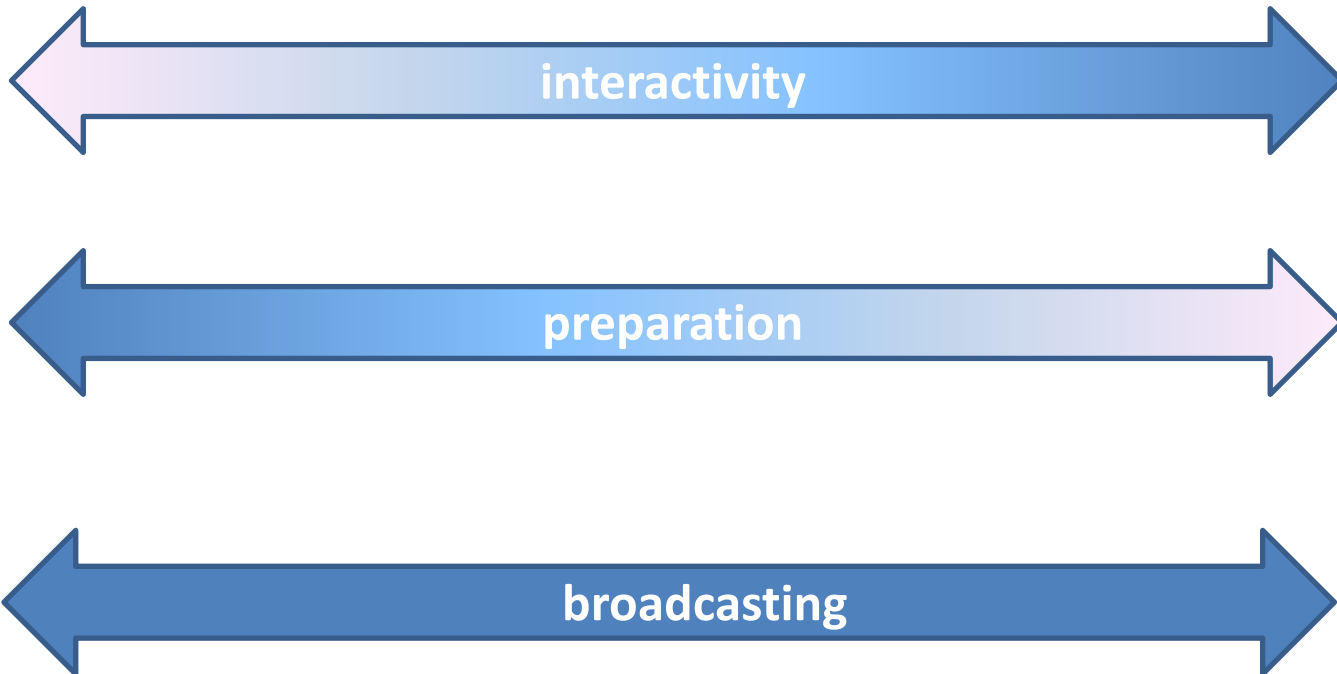
Discourse Markers as a thing of speech

- FRENCH

- *parce que*: objective in writing, objective/subjective/intersubjective in speech (Simon & Degand 2007)
- *au fond*: abstract discursive meaning in speech (61%) vs. in writing (35%) (Tanguy & Sarda 2013)
- *alors*: temporal in writing (35%), metadiscursive in speech (63%) (Degand & Fagard 2011, see also Schlamberger Brezar 2012)
- *mais*: “phatic” uses (metalinguistic, structuring) in speech vs. adversative in writing (Schlamberger Brezar 2012)

Discourse Markers as a thing of speech

- Variation according to the speech situation



Discourse Markers in LOCAS-F

- **Louvain Corpus of Annotated Speech – French**
 - **12** “genres”
 - Communicative situations (degree of interactivity, degree of preparation, degree of broadcasting)
 - **42** audio samples
 - 3 hours 38 minutes
 - ca. 15 min/genre
 - **74** speakers (**48** different speakers)
 - A mean of 4 speakers/genre (max. 11)
 - 12 speakers in more than one genre

Genre	Duration	Interactivity	Preparation	Broadcast
Conference talk	16:44	0	2	0
Debate	19:19	2	1	2
Academic address	15:18	0	2	1
Political address	20:26	0	2	2
Homily	13:23	0	2	1
Survey	16:04	1	1	0
Free interview	15:33	2	0	2
Radio interview	20:29	1	1	2
Radio news	14:45	0	2	2
Reading	15:18	0	2	0
Free conversation	10:22	2	0	0
Spont. Form. Conv.	12:51	2	0	0
Spont. Inform. Conv.	14:24	2	0	0
Radio creation	13:24	0	1	2
Total	3:38:20			

Genre	Duration	Interactivity	Preparation	Broadcast
Conference talk	16:44	0	2	0
Debate	19:19	2	1	2
Academic address	15:18	0	2	1
Political address	20:26	0	2	2
Homily	13:23	0	2	1
Survey	16:04	1	1	0
Free interview	15:33	2	0	2
Radio interview	20:29	1	1	2
Radio news	14:45	0	2	2
Reading	15:18	0	2	0
Free conversation	10:22	2	0	0
Spont. Form. Conv.	12:51	2	0	0
Spont. Inform. Conv.	14:24	2	0	0
Radio creation	13:24	0	1	2
Total	3:38:20			

Discourse Markers in Locas-F

- Macro-syntactic definition:
 - It has to be syntactically detachable from a sentence (Schiffrin 1987)
 - They do not enter into the construction syntactically with other elements of the sentence (Sankoff et al. 2007)
 - [DMs are] “either outside the syntactic structure or loosely attached to it” (Brinton, 1996: 34).

Discourse Markers in Locas-F

- Degree of interactivity:
 - High
 - Types: 50
 - Tokens: 868
 - Top markers: ***mais, et, donc, ben, bon, alors, hein, puis, quoi, en fait, voilà, quand même, ou sinon*** (> 10)
 - Low
 - Types: 51
 - Tokens: 656
 - Top markers: ***et, donc, mais, puis, alors, quoi, en fait, voilà, ben, bon, d'ailleurs*** (> 10)

Discourse Markers in Locas-F

- Degree of preparation:
 - High
 - Types: 39
 - Tokens: 314
 - Top markers: ***et, mais, donc, alors, en fait*** (> 10)
 - Low
 - Types: 47
 - Tokens: 904
 - Top markers: ***et, mais, donc, puis, ben, bon, enfin, quoi, alors, hein, en fait, voilà, quand même, ou sinon*** (> 10)

Discourse Markers in Locas-F

- Degree of broadcasting:
 - High
 - Types: 31
 - Tokens: 234
 - Top markers: ***et**, mais, alors, puis* (> 10)
 - Low
 - Types: 50
 - Tokens: 856
 - Top markers: ***et**, mais, **donc**, puis, alors, bon, ben, enfin, en fait, quoi, hein, ou, quand même* (> 10)

Discourse Markers in Locas-F

- DM-friendliness of situational features:
 - High: interactive/non-prepared/not broadcasted
 - Types: 39
 - Tokens: 418
 - Top markers: *et, mais, donc, bon enfin, ben, puis, alors, hein, quoi, en fait* (> 10)
 - Low: non-interactive/prepared/broadcasted
 - Types: 19
 - Tokens: 117
 - Top markers: *et, mais* (> 10)

Discourse Markers in Locas-F

- Situation-specific DMs
 - **Non-interactive:** *du coup, car, c'est que, ainsi, au contraire, ensuite, de ce fait, voilà que, de plus, néanmoins, par ailleurs, certes, de même, toutefois, pour autant*
 - **Interactive:** *ou sinon, beh, en même temps, après, tout de même, par contre, encore que, quoique, bien que, m'enfin*

Discourse Markers in Locas-F

- Situation-specific DMs
 - **Non-prepared:** *quoi, là, tiens, ou sinon, en plus, allez, beh, en même temps, déjà, et puis, après, par contre, bref, ensuite, encore que, bien que, quoique*
 - **Prepared:** *en effet, car, ainsi, au contraire, par conséquent, cependant, de ce fait, de plus, par ailleurs, certes, de même, toutefois, pour autant*

Specific DMs as a thing of speech

- Focus on non-relational DMs (Degand & Simon-Vandenberghe 2011, Crible subm.):
- DisFren Corpus (Crible 2014):
 - Multi-genre spoken FR/EN corpus (8 genres x 1h x lgg -176.000 words), 5444 annotated DMs → conversation 2467, news 210/political 351
- NRDMs in French and English conversation
 - *quoi, mais, hein, ben, tu vois, enfin, bon, en fait, voilà, alors*
 - *well, you know, but, I mean, so, actually, like, anyway, in fact*

Non-relational DMs in FR and EN conversation

FR Non-Rel DM	Freq.	Rel. Freq.
mais	242	138,82
quoi	216	123,91
enfin	93	52,20
ben	70	40,16
hein	64	36,71
tu vois	57	32,70
alors	50	28,68
bon	17	9,75
en fait	14	8,03
voilà	7	4,01

FR Non-Rel DM	Freq.	Rel. Freq.
well	132	75,52
but	123	70,37
so	83	47,48
you know	68	38,90
I mean	60	34,33
actually	25	14,30
like	9	5,15
anyway	9	5,15
in fact	6	3,43

Non-relational DMs in French conversation

- *le loup s'en s'empresse/ [M chemin] pour arriver chez la mère-grand avant le petit chaperon rouge **mais** là faut mettre pas pris le bon chemin*
- *[about computers] ces derniers temps ça prend beaucoup de d' expansion **quoi** il y a beaucoup de de d' imprimeries et d' entreprises qui emploient ça /*
- **bon** moi heureusement je n'ai pas eu de choses très graves **mais enfin** quand même / on sent / on commence à se sentir qu'on / qu'on / qu'on diminue /
- - ah **mais** je ne regrette pas **hein**
 - non / et ton papa il fait quoi exactement dans l'imprimerie
 - **ben** il fait de la photogravure

Subtitles as a type of speech

- Movie subtitles of “movie conversations”
 - “Written to be spoken as if not written” (Gregory & Carroll 1978, 47 cited in Biagini 2010)
 - Realistic representation of “real-world” conversations because basis is co-construction and interaction (Biagini 2010)
 - Corpus of American Soap Operas (100 M words) more informal (lexis and grammatical constructions) than BNC and COCA (spoken)

Subtitles as a type of speech

- Subtitles: "diasemiotic translation in polysemiotic media (including films, TV, video and DVD), in the form of one or more lines of written text presented on the screen in sync with the original dialogue" (Gottlieb 2004, 220)
- Technical constraints → translation?
 - Time-and-space constraints → condensation
 - Diasemiotic: from speech to writing (not from writing to writing)

Subtitles as a type of speech

- Reliable source of lexical (frequency) norms (New *et al.* 2007, Keuleers *et al.* 2010)
- “Film subtitles represent highly interactive involved informal communication. At the same time, they contain relatively few pause fillers, repetitions, reformulations and other discourse markers, which are typical of spontaneous discourse produced under real-time constraints.” (Levshina, *subm.*)

Subtitles as a type of speech

- « [dans les sous-titres] les MD surgissent dans des positions stratégiques et efficaces d'un point de vue communicatif, à savoir notamment les ouvertures et clôtures des tours de parole » (Biagini 2010, 14)
- Simplification of DMs, especially interjections and conjunctions (Biagini 2010)
- Subtitles are a « discours écrit oralisé » (Biagini 2010, 15)

DMs in subtitles

- Redundancy with image?
 - priority to propositional content → condensation/simplification of DMs?
- Importance of interpersonal dynamics
 - les MD demeurent parmi les éléments les plus efficaces pour exprimer le rapport à autrui dans son propre discours (Biagini 2010, 6)

DMs in subtitles

- Corpora:
 - OpenSubtitles2013 (Tiedemann 2012)
- Corpus for this study (< Levshina forthc.):

	ENGLISH	FRENCH
ORIGINAL	48650	35548
TRANSLATION	38354	49402

DMs in Subtitles

- French subtitles vs. Conversation

	Subtitles		Conversation
DM	Freq.	Rel. Freq.	Rel. Freq.
Mais	133	37,41	138,82
Alors	67	18,85	28,68
Bon	45	12,66	9,75
Voilà	32	9,00	4,01
Hein	20	5,63	36,71
Ben	20	5,63	40,16
Quoi	12	3,37	123,91
Enfin	11	3,09	52,20
En fait	4	1,12	8,03
Tu vois	2	0,56	32,70
Total	345	97,05	476,13

DMs in Subtitles

- DMs in French subtitles and their translation

DM	Original subtitles	Translated	
Mais	133	34	26%
Alors	67	27	40%
Bon	45	33	73%
Voilà	32	12	37%
Hein	20	8	40%
Ben	20	6	30%
Quoi	12	4	33%
Enfin	11	1	9%
En fait	4	2	50%
Tu vois	2	2	100%
Total	346	129	37%

DMs in Subtitles

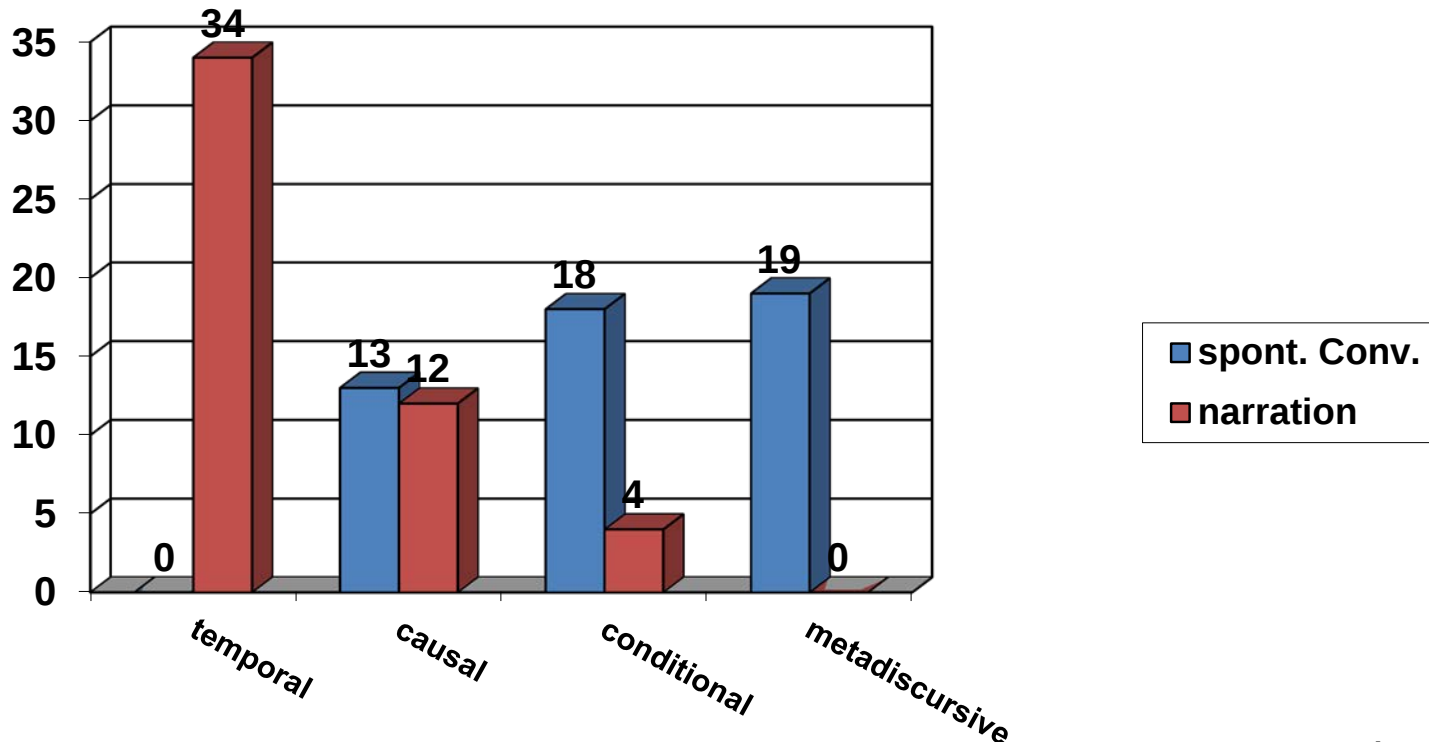
- DMs in French subtitles and their translations
 - Mais: **but** (*yet, and, not at all, you see, what, of course*)
 - Alors: **so** (*yes why, did it, all right, then*)
 - Bon: **well** (*oh yeah, good, all right, did I, OK, really, that's it, what about, come on*)
 - Voilà: *here, there, so, all right, that's it*
 - Hein: *hmm, did-n't you, can you, aren't you, right*

DMs in Subtitles

- DMs in French subtitles and their translations
 - Ben: *well, same here, then, gosh, come on*
 - Quoi: *or what*
 - Enfin: *so be it*
 - En fait: *in fact, so, that's why*
 - Tu vois: *see, what*
- 50% of the translated DMs belong to the most speech-like DMs in English (*so, but, well, in fact, you see*)

A case study: *alors*

- *Alors* in spontaneous conversation
 - From connective to NRDM in speech → orality marker (cf. also Schlamberger Brezar 2012)



A case study: *alors*

- *Alors* in spontaneous conversation
 - mm et ce qui te plaît le plus **alors** c' est la mise en page
 - **alors** c'est/ euh: et ben et où habite-t-elle cette mère-grand ?
 - pourquoi ils ils ils arrivent difficilement à téléphoner [mm] à s' exprimer au téléphone écrire c' est encore pis / **alors** euh ces immigrants viennent et ils ont souvent des problèmes un peu complexes

A case study: *alors* in subtitles

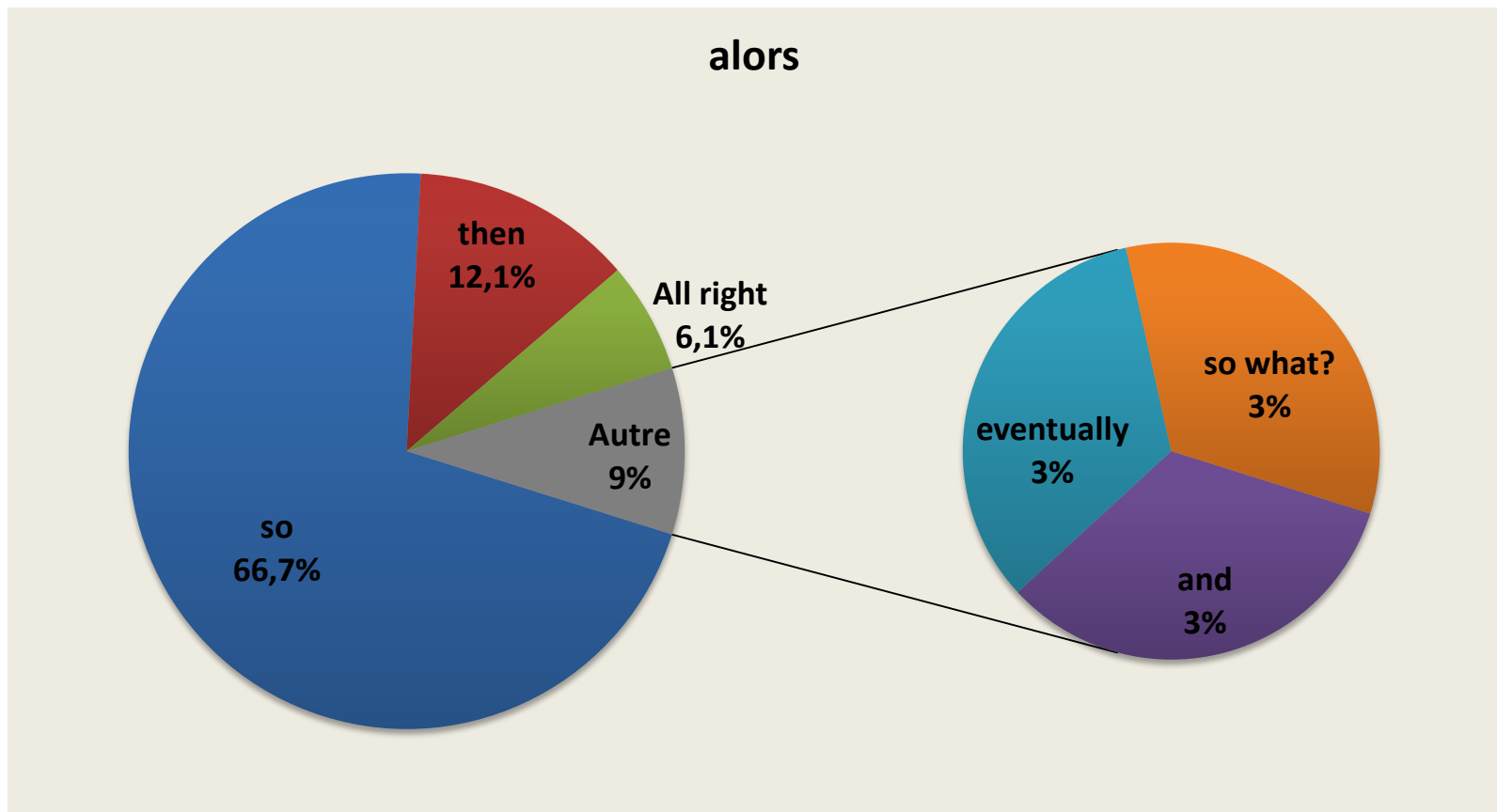
- *Alors* in original subtitles (FR)
 - 67 / 27 translated
 - Translations: **so** (19/70%) *yes why* (1), *did it* (1), *all right* (1), *then* (5)
 - ***Alors*** *c'est l'autre là? Le type qui lève la main?*
 - ***So*** *it's this man here, with his hand up?*
 - *Il dormira devant ma porte. | He'll sleep on a mat in the doorway.*
 - *Pas ici! | Not here, he won't.*
 - ***Alors*** *dans l'écurie! | In the stable, then.*

A case study: *alors* in original subtitles

- Semantic distribution (n= 50)
 - Temporal: 0
 - Causal/Conclusive: 13
 - C'était le genre à écarter les cuisses, mais il fallait du satin. **Alors** mon mari s'est mis à piquer dans la caisse.
 - Metadiscursive (interactional, discourse structuring, turn management): 37
 - Alors, une figue et trois noisettes, comme d'habitude ?
 - Il vous dégoûte alors ?
- Matches spoken conversational data

A case study: *alors* in translated subtitles

- *Alors* in translation
 - 75! / 33 are translations



Conclusions

- DMs are a relevant index of orality
 - Variation with the discourse situation
 - High degree of interactivity: + tokens
 - Low degree of preparation: + types, + tokens, + distinctive markers
 - Low degree of broadcasting: // low degree of prep.
 - Situation-specific (low-frequency) DMs
 - High-frequency → non-relational DMs

Conclusions

- DMs *do* occur in (original and translated) subtitles
 - Nearly five times less frequent in original subtitles than in spont. Conversation (97,05 vs. 476,13)
 - Still less in translated subtitles (46,93)
 - And only 37% of the DMs present in original subtitles are translated
- BUT, the semantic distribution of DMs in subtitles and conversation is highly similar

Pour la route!

- **Added DMs in translation subtitles**
 - French → English:
 - 37% of DMs are translated (n = 64)
 - Most frequent translations: *so/but* > *well*
 - Number of English NRDM in French translations: 180!
 - 80/180 (44%) correspond to French DM → 100 (56%) were ADDED
 - Most added DMs: *well/so* > *but*

Et maintenant la pub!

- COST Action IS1312: TextLink: Structuring Discourse in Multilingual Europe
 - http://www.cost.eu/COST_Actions/isch/Actions/IS1312
 - <http://textlink.ii.metu.edu.tr>
- Thanks to
 - MA students of Constrastive linguistics class
 - Ludivine Crible
 - Natalia Levshina

- Biagini, M. 2010. Les sous-titres en interaction : Le cas des marqueurs discursifs dans des dialogues filmiques sous-titrés. *GLOTTOPOL Revue de sociolinguistique en ligne* 15 [http://glottopol.univ-rouen.fr/numero_15.html] (last accessed 5 May 2015)
- CRIBLE L. 2014. "Reaching cross-linguistic comparability across eight speech situations : a challenge for corpus design". Paper presented at the 7th Biennial IVACS Conference, June 19-21, Newcastle, UK
- Crible, L. 2015. Etude contrastive des marqueurs de discours français et anglais: approche onomasiologique sur corpus comparable. 4th International Symposium « Discourse markers in romance languages : a contrastive approach », May 6-9 2015, Heidelberg.
- Davies, Mark Corpus of American Soap Operas, 2001-2012 [<http://corpus.byu.edu/soap/>]
- DEGAND, L. et SIMON, A-C. (2009a). On identifying basic discourse units in speech: theoretical and empirical issues. In *Discours* 4 [En ligne]. URL [hjp://discours.revues.org/index.html](http://discours.revues.org/index.html).
- DEGAND, L. et SIMON, A-C. (2009b). Mapping prosody and syntax as discourse strategies: How Basic Discourse Units vary across genres. In WICHMANN, A., BARTH- WEINGARTEN, D. et DEHÉ, N. (eds) : *Where prosody meets pragmatics : research at the interface*, Studies in Pragmatics. Bingley: Emerald, pages 79--105.
- DEGAND, L. et SIMON, A-C. (2011). L'analyse en unités discursives de base : pourquoi et comment ?. In *Langue française* (2011), pages 45-59.
- Degand, L. and Simon-Vandenberg, A-M. 2011. Grammaticalization and (inter-subjectification of discourse markers. *Linguistics* 49, 287–294.

- Georgakopoulou, A. and Goutsos, D. 1998. Conjunctions versus discourse markers in Greek : The interaction of frequency, position, and functions in context. *Linguistics* 36(5) : 887-917.
- Gottlieb, H. 2004. Subtitles and International Anglification. *Nordic Journal of English Studies* 3 (1): 219-30.
- Keuleers, Emmanuel, Marc Brysbaert, et Boris New. 2010. SUBTLEX-NL: A New Measure for Dutch Word Frequency Based on Film Subtitles . *Behavior Research Methods* 42 (3): 643-50.
- Levshina, N. à par. Verbs of letting in Germanic and Romance languages: A quantitative investigation based on a parallel corpus of film subtitles. *Languages in Contrast*.
- Levshina, N. subm. Film subtitles as a corpus: An n-gram approach
- New, Boris, Marc Brysbaert, Jean Veronis, et Christophe Pallier. 2007. The use of film subtitles to estimate word frequencies . *Applied Psycholinguistics* 28 (04): 661-77.
- Tiedemann, J. 2012, Parallel Data, Tools and Interfaces in OPUS. In Proceedings of the 8th International Conference on Language Resources and Evaluation (LREC 2012)

Ongoing

- Fraser (1999: 933) DMs are “prosodically independent, being both accented and prosodically separated from their surrounding context by pauses, intonation breaks, or both”.
 - DMs do not systematically fulfill the assumption of prosodic independence. Rather prosodic independence results from syntactic and discursive constraints that need to be uncovered.
 - Correlate the prosodic characteristics of with the other levels of annotation (BDU, syntactic clauses)
- Specialization of LPE-S as coherence markers and LPE-I as topic management marker → which linguistic expressions, which discourse relations?