

EXPLORATIONS AT THE DISCOURSE LEVEL: BASIC DISCOURSE UNITS, DISCOURSE MARKERS, AND MORE

Liesbeth Degand

University of Louvain, Belgium

Institute for Language & Communication

liesbeth.degand@uclouvain.be

Outline

- Explorations at the discourse level
 - TextLink COST Action: Structuring Discourse in Multilingual Europe
 - Fluency and Disfluency Markers: A multimodal contrastive perspective
 - LOCAS-F: Louvain Corpus of Annotated Speech (French)
- Basic Discourse Units
 - Segmentation and annotation principles
 - Discourse strategies
- Discourse Markers
 - Identification and (functional) Annotation
 - Position
 - DMs as Fluencemes
- Research perspectives

Explorations at the discourse level

- Discourse is a crucial notion for understanding human communication (Sanders & Spooren 2007)
- “Discourse is what makes us human, what allows us to communicate ideas, facts, and feelings across time and space.” (Graesser, Millis, & Zwaan 1997: 164)
- Discourse phenomena to investigate:
 - Spoken Discourse Segmentation
 - The paradox of Discourse Markers

Ongoing projects

- TextLink COST Action: *Structuring Discourse in Multilingual Europe*
 - 29 partner countries, 20+ languages (written, spoken, signed)
 - **Aim:** Portal of interconnected corpora/resources annotated at the discourse level (Identification and semantic/functional annotation of « discourse relational devices »)
 - **Main challenge:** Compatibility of identification and annotation schemes
 - More info: <http://textlink.ii.metu.edu.tr/>

Ongoing projects

- *Fluency and Disfluency Markers: A multimodal contrastive perspective*
 - **Aim:** Multimodal (corpus-based) description of fluency/disfluency in English (L1/L2), French prosody (L1), English/French (DMs as fluencemes), LSFB → fluency and disfluency are the two sides of the same coin and the forms and functions of (dis)fluency markers systematically vary according to language, speaker and genre.
 - **Main challenge:** Common annotation scheme of « fluencemes »; from fluencemes to (Dis)Fluency → complexity of the language production system (Bock 1996)
 - More info: <http://uclouvain.be/415256.html>

Ongoing projects

- *Louvain Corpus of Annotated Speech (French)*
 - **Aim:** Identification and Annotation of « Basic Discourse Units » in a balanced multi-genre corpus of Spoken French + investigation of the left periphery
 - **Main challenges:** Syntactic and prosodic annotation, Cognitive status of BDUs
 - More info: Publication list

BASIC DISCOURS UNITS IN LOCAS-F

Spoken Discourse Segmentation

- “The stepwise structure of speech can give us many clues about cognitive processes. It suggests units for planning, production, perception and comprehension.” (Holsánová 2008: 1)

Spoken Discourse Segmentation

- Basic Discourse Units segmentation model accounts for
 - Discourse as **product**: structured and organized object in terms of coherence and discourse relations between states and events → coherence relations approach (cf. Grosz & Sidner, 1986; Mann & Thompson, 1988; Polanyi, 1988; Roulet et al., 2001; Sanders, Spooren & Noordman, 1992)
 - Discourse as **process**: temporal event that is dynamic, transitional, subject to linearization and interactional achievement; i.e. transitory, irreversible, synchronized cf. discourse **flow** (of thoughts/ideas and of sounds/forms), (cf. Chafe, 1994; Couper-Kuhlen, 2012 ; Auer, 2009)

Linguistic correlates of discourse segmentation

- Both syntax and prosody provide linguistic cues that may be used to segment the discourse flow into delineated units
 - Discourse as product → priority to syntax
 - Discourse as process → priority to prosody
- Not a strict divide!

Priority to prosody

- the *Intonation Unit* is "a unit of mental and linguistic processing [...] that seems to be of exactly the right size to be processed in its entirety with the help of echoic memory" (Chafe, 1994: 55)
- "one new idea constraint"
 - "Conversational language appears subject to a constraint that limits an intonation unit to the expression of no more than one new idea" (Chafe, 1994: 119).
 - "It may be that neither the speaker nor the listener is able to handle more than one new idea at the time" (Chafe, 1994: 109).

Priority to syntax

- “In actual speech, the processing of syntactic structures in itself is largely independent of prosodic information. [...] [I]t appears to be a truism that prosody, unlike syntax, is not inherently essential to the conveyance of information in human communication.” (Monschau, Kreyer and Mukherjee, 2003: 584-585)

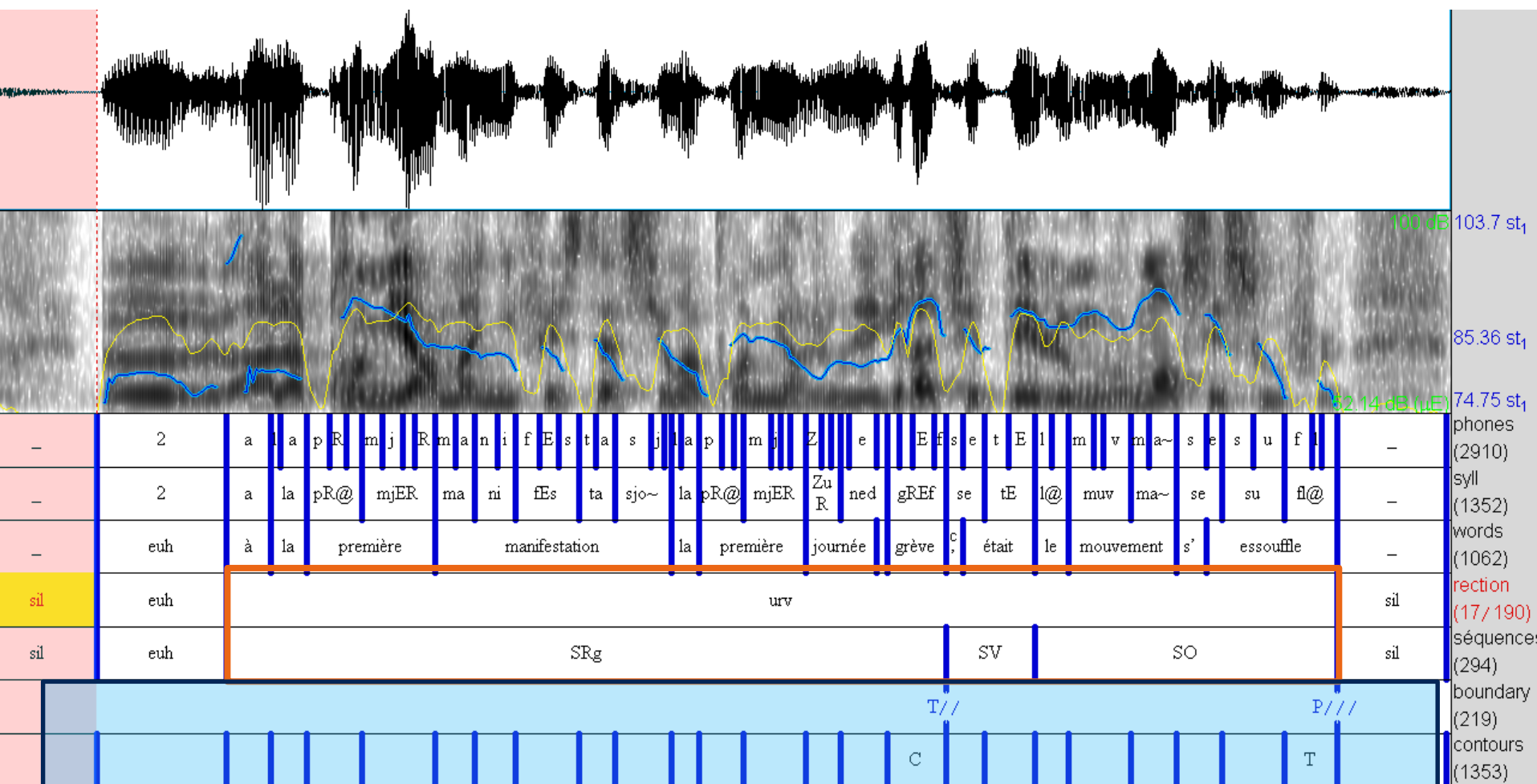
Basic Discourse Units model

- Syntax and prosody are two distinct, yet interdependent and complementary levels of surface segmentation, that work together to give rise to a specific discourse organizational level (Degand & Simon, 2009; in prep.)
- A basic discourse unit (BDU) emerges whenever a syntactic boundary coincides with a prosodic boundary. As long as the two levels do not coincide, the hearer awaits completion. More precisely, prosodic and syntactic structures do not always align in a one-to-one relation (see also Chafe, 1994; Croft, 1995, Ford & Thompson, 1996)

Segmentation into Basic Discourse Units

- Mapping of two **independently** performed levels of analysis generates specific discourse organising level
 - **Syntactic analysis**: segmentation into dependency clauses (and functional sequences)
 - **Prosodic analysis**: segmentation into major (and intermediary) prosodic units

Syntactic and Prosodic Segmentation



Syntactic segmentation and annotation

- Step 1: word-based orthographic transcription of the discourse to be analyzed in a Praat tier;
- Step 2: segmentation into syntactic dependency clauses;
- Step 3: annotation of functional sequences.
- Entirely manual, double-blind annotation on the basis of a (incremental) syntactic annotation manual (Tanguy et al., 2012). All cases of disagreement were solved by discussion between the two coders.

Syntactic segmentation

- Theoretical basis: dependency syntax (see e.g. Blanche-Benveniste 2002; Blanche-Benveniste et al., 1990; Berrendonner 2002; Deulofeu 2003)
 - 4 types of dependency clauses
 - a) verbal dependency clause (*urv*)
 - b) averbal dependency clause (*ura*)
 - c) elliptical dependency clause (*ure*)
 - d) interrupted dependency clauses (*urx-l*).
 - + macro-syntactic segments: adjuncts and discourse markers

Syntactic segmentation

- Functional sequences: clausal constituents that occupy a main syntactic function like Verb, Subject, Object, etc. (Bilger & Campione 2002: 119).
 - Verbal clauses: verb (SV, including the pronominalized subject and object, if any), (lexicalized) subject (SS), object (SO), (weak) governed constituent (SR) and insert.
 - Averbial clauses: morphosyntactic constituents, like nominal sequence (SN), adverbial sequence (SAdv), prepositional sequence (SPrep), etc.

Examples

- 1) <ben>_{md} [(on va aller voir)_{SV} (un autre film)_{SO}]^{urv} [(on n'allait pas)_{SV}]^{urv-l} <enfin>_{md} [(on va aller voir)_{SV} (un autre film)_{SO}]^{urv}

<well> [we are going to watch another movie] [we were not going to]

<well> [we are going to watch another movie]

- 2) euh [(Dreyfus)_{SS} <donc>_{md} euh (a valu)_{SV} (on va dire)_{insert} (à Zola)_{SO} euh (de s d'émigrer en en Angleterre)_{SO}]^{urv+} [(pourquoi)_{SAdv}]^{ura} [(parce qu'on on lui a reproché son intervention peut-être un peu trop radicale)_{SRg}]^{ure}

uh [Dreyfus <then> uh caused we shall say Zola uh to to emigrate to to England] [why] [because he was blamed for his intervention that was maybe a little too radical]

- 3) <ça>_{ag} [(c'est)_{SV} (la ville)_{SO}]^{urv}

<that> [that's the town]

Prosodic segmentation and annotation

- Step 1: Double blind prosodic segmentation by two experts, focusing on perceptive effects (~~syntax~~), multiple listening
- Step 2: Testing of semi-automatic procedure of different prosodic boundary levels → 2 levels
- Step 3: « Validation » of semi-automatic annotation

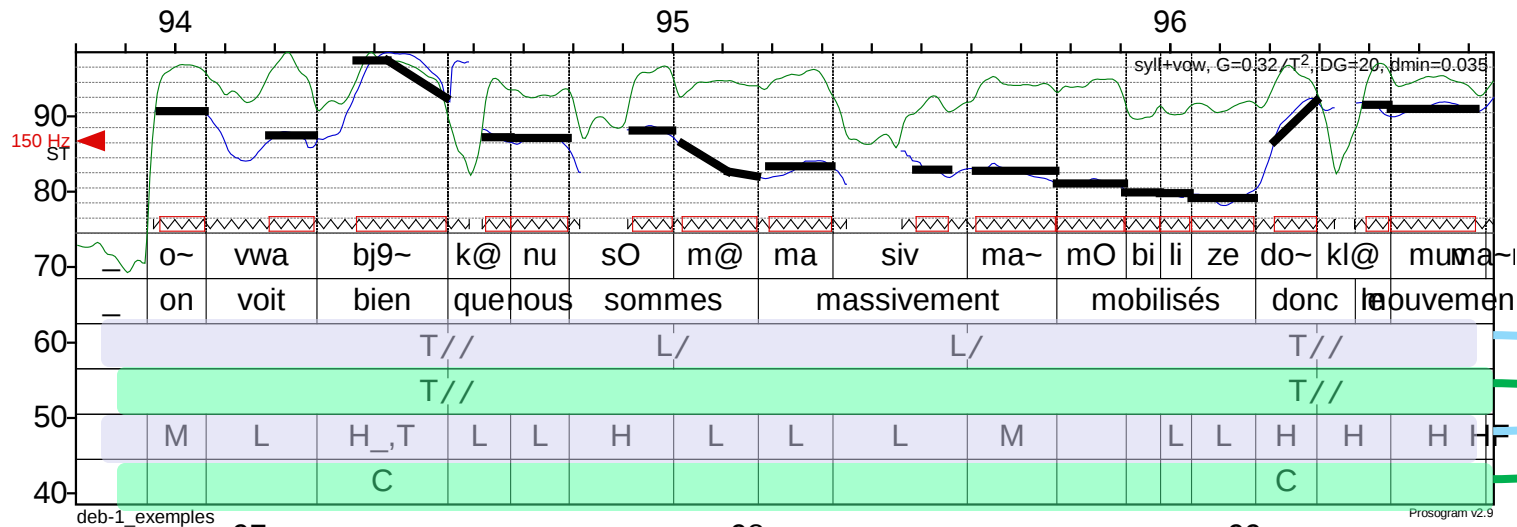
Major intonation units

- A major prosodic boundary (///) is established when one of the following cues is detected on the final syllable of a word (or penultimate when the final syllable is a schwa):
 - a subsequent silent pause longer than 250 ms ;
 - an extra-lengthening (the syllable is three times longer than the syllables in the context);
 - a sharp rise of f0 (intra-syllabic f0 rise superior to ten semi-tones), even when the f0 rise does not correlate with the lengthening of the syllable.
 - manual exclusion of a boundary which coincides with a hesitation mark ('euh' particle or vowel extra-lengthening with a level contour, and creaky voice).

Intonation contours

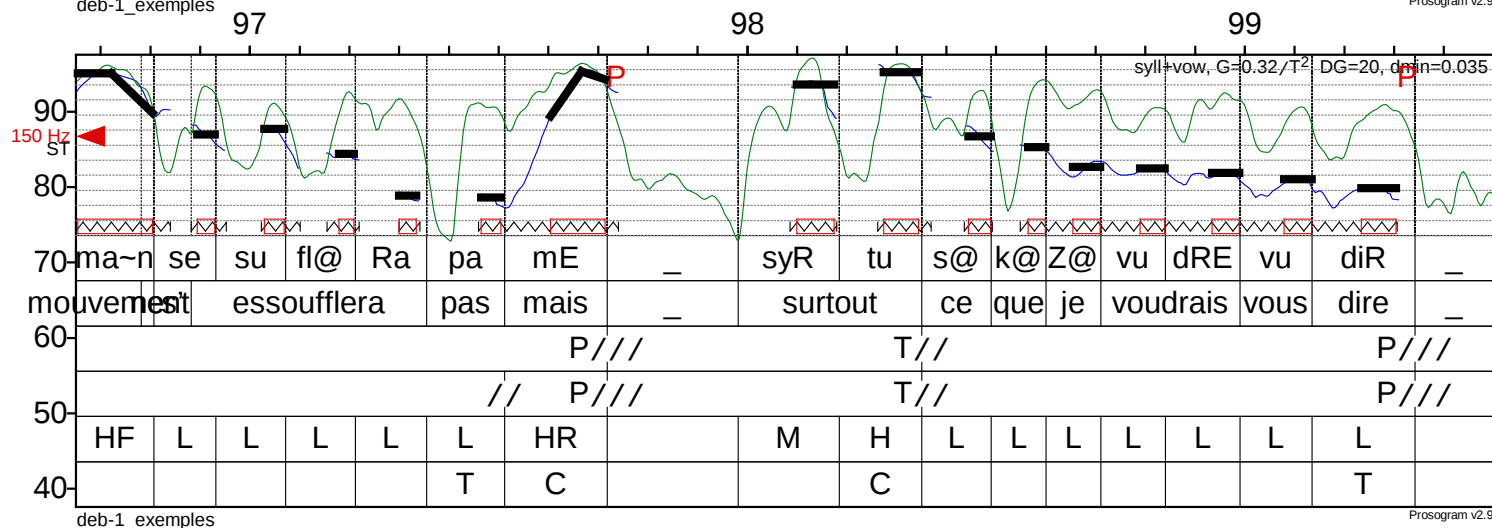
- Four alternatives exist for each prosodic boundary:
 - Continuation (rising f_0 movement),
 - Finality (falling or low f_0),
 - Focus (sharp falling from high to low contour)
 - Suspense (flat and lengthened contour).
- Two coders independently annotate intonation contours ,
and rest on automatic intonation annotation (Mertens 2012)
for resolving the disagreement cases.

Automatic annotation and manual validation



Automatic
annotation

Manual
validation



Mapping of prosody and syntax

- A BDU emerges from the mapping of coinciding syntactic and prosodic boundaries
 - Halford's 'talk units', i.e. "the maximal unit defined by syntax and intonation". (Halford, 1996: 43)
- Concretely, PRAAT script → 5 types of BDUs

Types of BDU

DC	[]	[a, md]	[] []	[]	[] [] []
MPU	[]	[]	[]	[] [] []	[] []
BDU	[bdu-c]	[bdu-a]	[bdu-i]	[bdu-s]	[bdu-x]

(Degand & Simon 2009a)

- **Congruent BDU-c** : one dependency clause corresponds to one major prosodic unit
- **BDU-s grouped by syntax**: several major prosodic units correspond to one dependency clause
- **BDU-i grouped by intonation**: several dependency clauses correspond to one major prosodic unit
- **Adjunct BDU-a** (or regulatory): a non governed element is autonomized in a major prosodic unit
- **Mixed BDU-x** : sequence of mismatching syntactic and intonation units

Types of BDUs: Examples

Congruent BDU-c

- [(dans la majorité politique libanaise) (beaucoup de voix) (s'étaient élevées) (contre sa venue)]^{urv} **///C**



in the lebanese political majority many voices raised against his coming ///

Syntax-bound BDU-s

- [(pourquoi) (ce rôle majeur) (n'est-il pas dévolu) **///S** (à la religion)]^{urv+} **///C**



why isn't this major role devoted /// to religion

Intonation-bound BDU-i

- euh [(on est devenu) (bien potes)]^{urv} [(tout le monde) (se connaissait)]^{urv} **///C**



uhm we've become good friends [urv] everybody knew one another [urv]

Regulatory BDU-a

- <mais> <bon> **///T**

but well ///

Types of BDUs: examples



'Mixed' BDU-x

- <sinon>_{md} //C [(les les exercices oraux)_{ss} //C <pr/ à proprement parler>_{ag} **///C** (sont minimales)_{sv} **]**^{urv+} **//T** [(ils existent)_{sv} //C (au niveau de la lecture //C bien sûr **///C** d'un **///S** déchiffrement //C ou défrichage //T comme vous voulez //C dans les deux sens //C en quatrième **///C** et certainement //C à propos de la lecture des poèmes //C en cinquième)_{SRd} **]**^{urv} **///C**

Corpus LOCAS-F

- **Louvain Corpus of Annotated Speech – French**
(collaboration with Anne Catherine Simon, Laurence Martin - Noalig Tanguy, Thomas Van Damme)
 - **14** « genres »
 - Communicative situations (degree of interactivity, degree of preparation, degree of broadcasting, number of speakers)
 - **48** audio samples
 - 3 hours 38 minutes
 - ca. 15 min/genre
 - **48** different speakers (12 speakers in more than one genre)
 - At least 3 different speakers per genre
 - **41 322** tokens

LOCAS-F Situational features

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 narration	10:22	2	0	0
Radio creation	13:24	0	1	2
Total	3:11:05			

LOCAS-F Situational features

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 narration	10:22	2	0	0
Radio creation	13:24	0	1	2
Total	3:11:05			

Functions of Basic Discourse Units

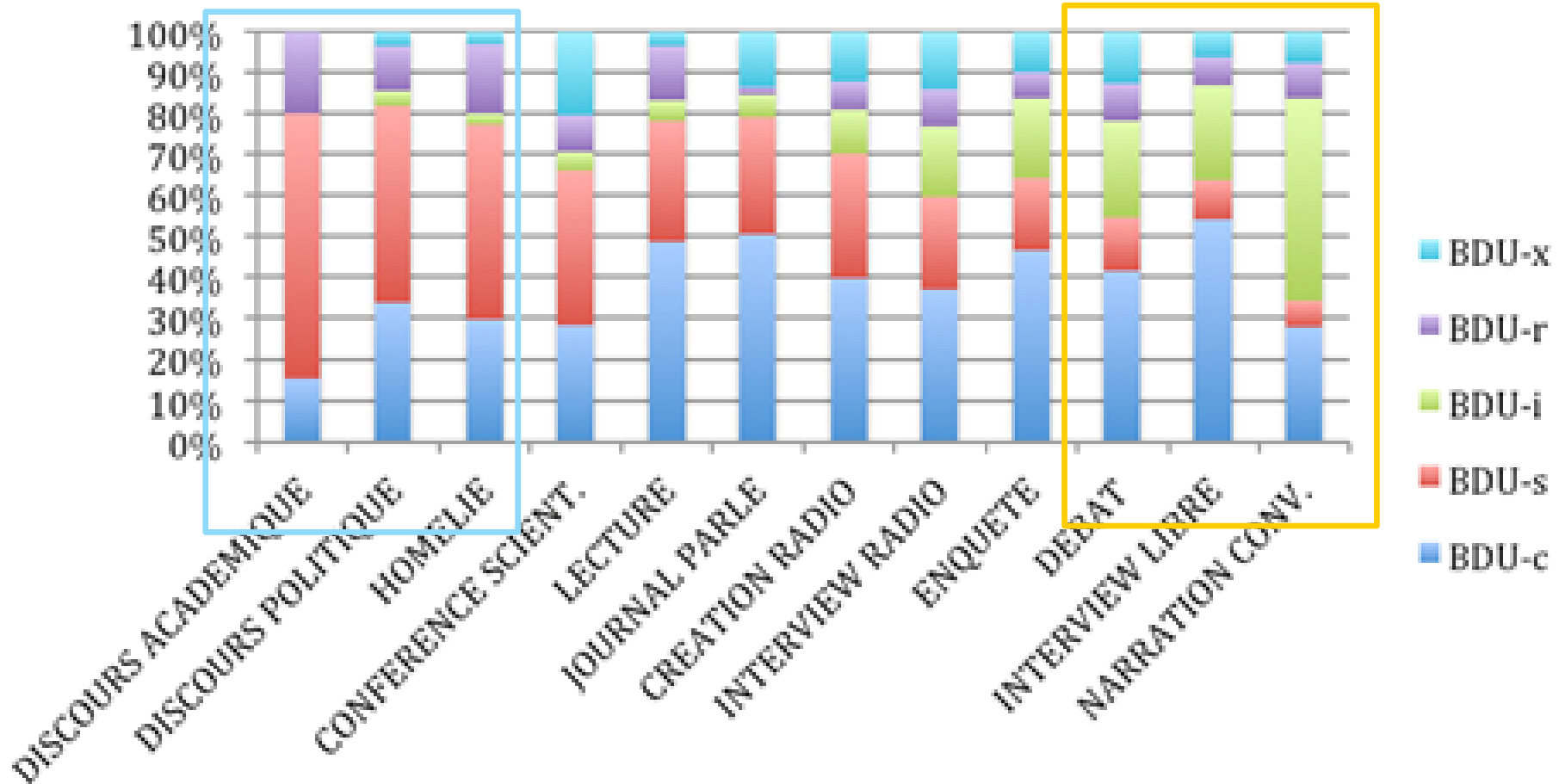
Hypothesis

- BDUs are “the segments speakers and hearers rely on to construct and interpret the ongoing discourse, viz. segments on the basis of which inferential processes can take place.” (Degand & Simon 2009a: 82)

Data-proof

- BDUs are discourse strategic units that vary with the communicative situation (Degand & Simon 2009ab, 2011; Simon & Degand 2011).
- Different genres call for different BDU distributions (40% congruent).

LOCAS-F: BDU distribution



$$\chi^2 = 602,84 ; dl = 44 ; p < 0,0001$$

LOCAS-F in figures

#	BDUs	Words	Length	Genres	Samples	DM
Total	2875	41322	3:38	14	48	1780



BDU-c: 43%
BDU-i: 17%
BDU-s: 23%
BDU-r: 9%
BDU-x: 8%

BDU Intermezzo

- What is the cognitive status of BDUs?

- Production units?
- Comprehension units?

Models of Discourse processing

Kintsch & van Dijk 1978, Oakhill et al. 1989, Dell & Brown, 1991, Bock, 1996 ...

- How to find out?

- Systematic (conversation) analysis of non-congruent BDUs?
- Systematic analysis of prosodic units vs. Syntactic units vs. BDUs?
 - One idea one unit vs. One clause constraint
- BDUs in interaction?
- BDUs in writing (keystroke logging?)
- Other ideas ...

THE PARADOX OF DISCOURSE MARKERS

The paradox of Discourse Markers

- Sentence-level: optional, not a (morpho-)syntactic category
- Discourse-level: Communicatively obligatory in ALL acts of human communication
 - written/spoken
 - formal/informal
 - CMC, even when space-constrained (texting, twitter, ...)

The paradox of Discourse Markers

- Human communication is not possible without a linguistic category that is itself vaguely defined, i.e. DMs.
- Indices of fundamental cognitive processing during (spoken) language production, and genuinely constitutive of human communication.
- DMs are both **speaker-oriented** traces and **hearer-oriented** signals (but how do we find out?)

Discourse Markers

- Linguistic expression 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 in LOCAS-F

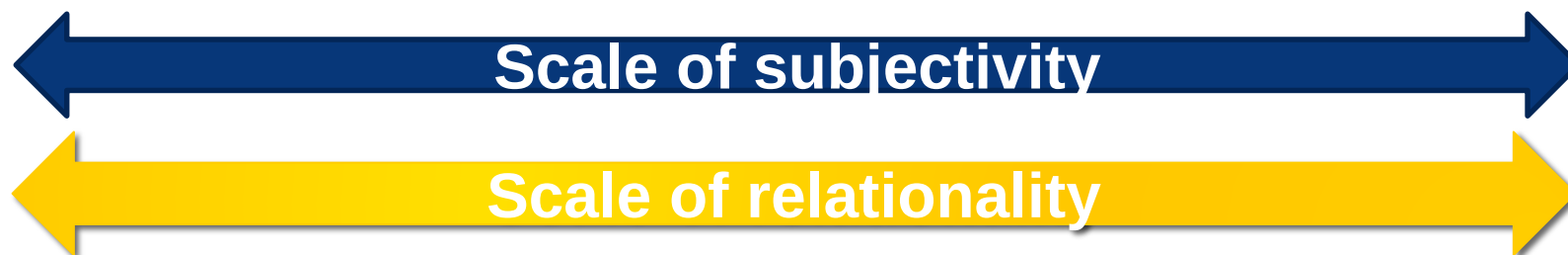
- Macro-syntactic constraints
 - It has to be syntactically detachable from a sentence (Schiffrin 1987)
 - They do not enter into the construction syntactically with other elements of the sentence (Sankoff et al. 2007)
 - [DMs are] “either outside the syntactic structure or loosely attached to it” (Brinton, 1996: 34).
- **Weak clause association** (outside the dependency clause)
- Identification of 1780 DM tokens (73 types)
 - From semasiology to onomasiology!

Variation in DM use (context!)

	+ interactive	- interactive
- prepared	42 types 649 tokens <i>après, au fond, aussi, bien que, bon, en même temps, encore que, par contre, quoique, sinon</i>	28 types 255 tokens <i>ensuite</i>
+ prepared	0	39 types 314 tokens <i>ainsi, au contraire, car, cependant, certes, de ce fait, de même, de plus, également, en effet, par ailleurs, par conséquent, pour autant, toutefois</i>

Functional annotation of DMs

Ideational	Rhetorical	Sequential	Interpersonal
linked to states of affairs in the world, semantic relations between real events .	pragmatic relations applying to subjective claims and implicit assumptions. .	linked to the structuring of discourse segments, explicit signals of the progressing steps of speech and thought.	linked to intersubjective and phatic functions in the management of the speaker-hearer relationship



Rounds of functional annotation

Round 1	Round 2	Round 3
Kappa: .43 + (abstract) coding scheme	Kappa: .59 + bias ideational > rhetorical > sequential > interpersonal	Kappa: .78 + bias + DM cues/paraphr.

- Double-blind coding
- Progressive procedure

Sense annotation: inter-dependent taxonomy

Ideational	Rhetorical	Sequential	Interpersonal
cause	motivation	punctuation	monitoring
consequence	conclusion	opening boundary	face-saving
concession	opposition	closing boundary	disagreeing
contrast	specification	topic-resuming	agreeing
alternative	reformulation	topic-shifting	elliptical
condition	relevance	quoting	
temporal	emphasis	addition	
exception	comment	enumeration	

Revised taxonomy: independence

Speaker
intensions

Ideational	Rhetorical	Sequential	Interpersonal
WHAT?	Cause/consequence		
	Addition		
	Contrast/Concession		
	Temporal		
	Condition		
	Alternative		
	Specification		
	Punctuation		
	Structuring (topic-shift, opening, ...)		

Functional annotation: First results

- Annotation of 500 DMs, in context, with coding scheme

# DMs	Prepared	Semi-Prepared	Non Prepared	Total
Interactive	/	37	146	183
Semi-interactive	/	123	/	123
Non interactive	145	12	37	194
Total	183	145	172	500

Frequency of domains and senses

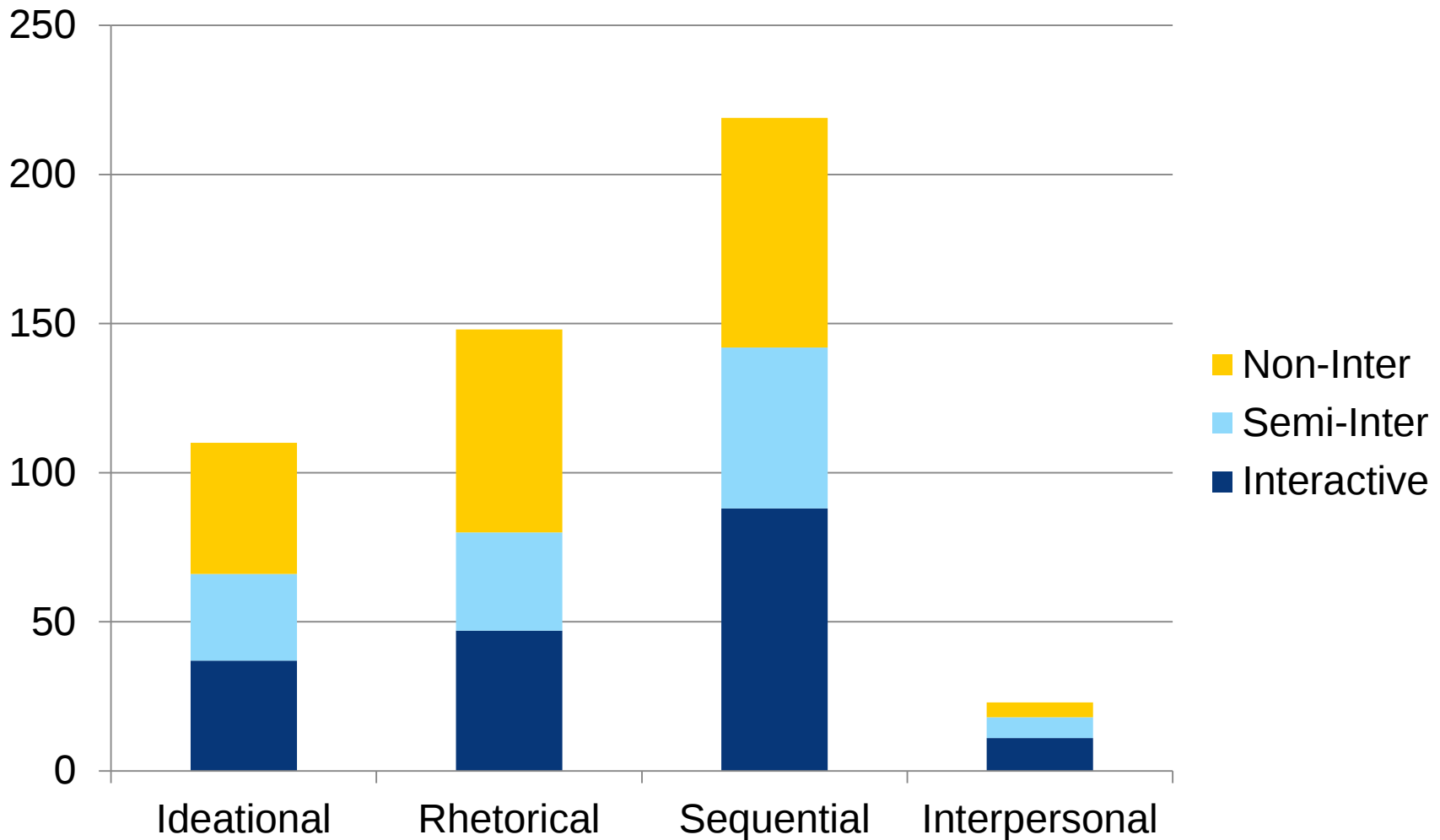
Sequential	219 (44%)	
Rhetorical	148 (30%)	
Ideational	110 (22%)	
Interpersonal	23 (4%)	

Addition	94 (19%)	
Contrast	81 (16%)	
Specification	56 (11%)	
Temporal	45 (9%)	
Consequence	45 (9%)	
Punctuation	41 (8%)	
Concession	30 (6%)	
Opening	28(6%)	
Alternative	19 (4%)	
Topic-shift	16 (3%)	
Resuming	13 (2%)	
Closing	11 (2%)	
Other	21 (5%)	

Independence between domains and senses

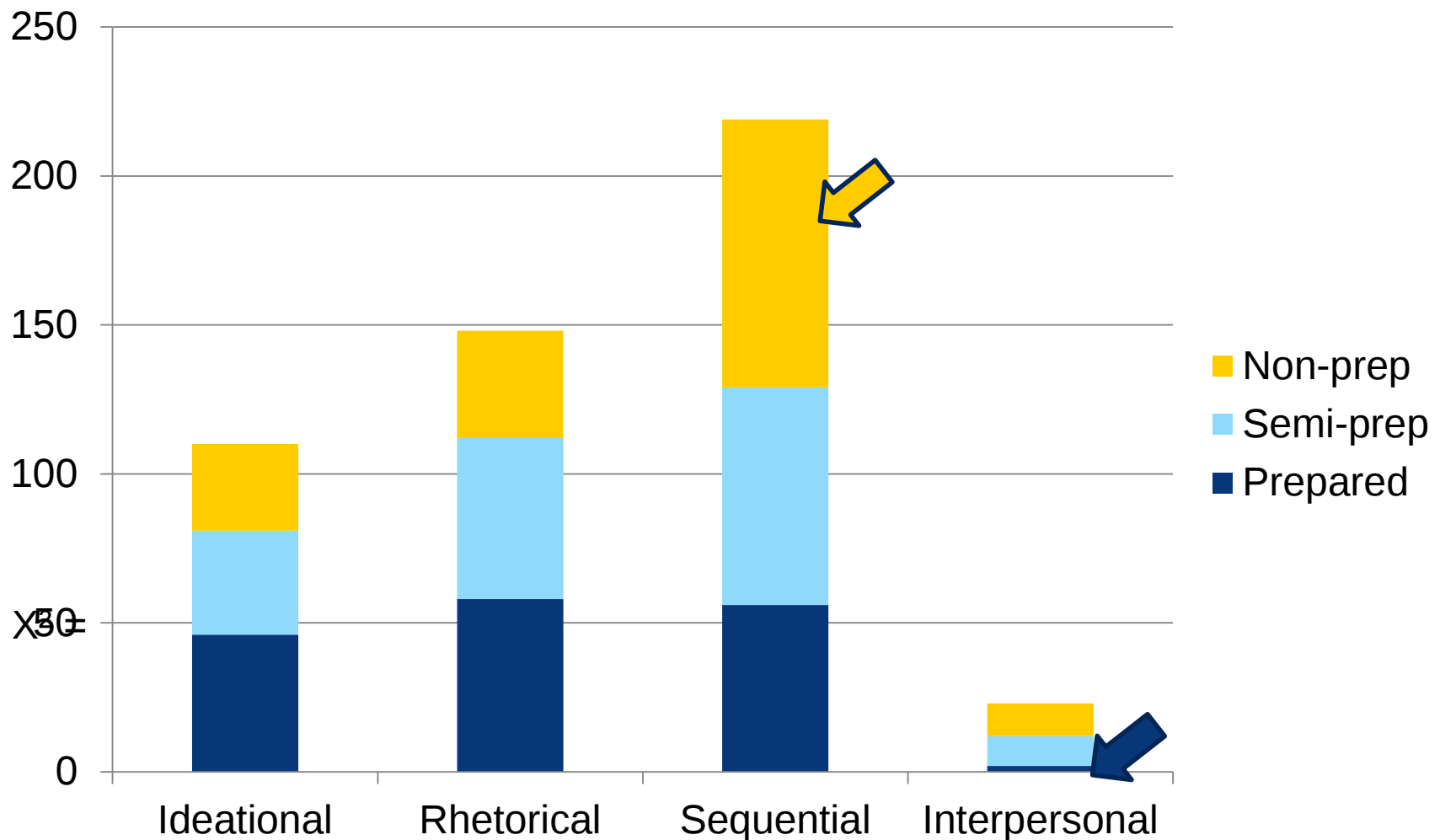
Ideational	Rhetorical	Sequential	Interpersonal
Contrast - Punctuation			
Addition – Specification – Consequence – Concession – Alternative – Topic-Shift			
Temporal		Temporal	
		Opening	
	Resuming – Closing - Condition		
Enumeration		Enumeration	
	Agreement		Agreement
	Cause		

Domains and degree of interaction



$\chi^2 (6, N=500), = 8.01; p = 0.237$

Domains vs. Degree of preparation



$\chi^2 (6, N=500), = 23.94; p < 0.0005$

Polyfunctionality of DMs

	ideational	rhetorical	sequential	Inter-personal	Total
<i>et</i>	52	6	70		128
<i>mais</i>	8	55	26	1	91
<i>donc</i>	10	23	19		52
<i>alors</i>	7	4	22	1	34
<i>et puis</i>	15	1	13		29
<i>ben</i>			16	2	18
<i>hein</i>			17		17
<i>enfin</i>		6	10		16
<i>en fait</i>		8	6		14
<i>bon</i>			11		11
Total	92	103	193	22	410/500

DM = 10 most frequent → 82%

(Micro and Macro)Functions of DMs

- Highly frequent DMs are polyfunctional (domain & sense)
 - Good for the speaker, not for the hearer
 - Hell for annotation !
- Most discourse relations/senses (addition, contrast, ...) cross domains
 - Core meaning of (polyfunctional) DM?
 - *mais*: 72% contrast, 18% concession, + ... → specific signal!
 - *et* : 66% addition, 9% temporal, 7% specification, 12% « structuring », 4% contrast → underspecified signal/trace?
 - *donc*: 60% consequence, 25% specification, 10% opening or resuming, + ...
- (under)specified relations or (under)specified DMs??

(Macro)Functions of DMs

- Sequential function (topic-shift, closing, quoting, addition, ...)
 - Organization/Structuring of speaker's (ongoing) discourse flow
 - Good for the speaker, and? the hearer
- Rhetorical function (motivation, reformulation, comment, ...)
 - Expression of speaker's point of view
 - Good for the speaker, and? the hearer
- Interpersonal function (monitoring, face-saving, agreeing, ...)
 - Not very frequent, even in highly interactive contexts
- Variation with discourse situation
 - the less prepared, the more sequential; the more prepared, the less interpersonal

DMs as markers of coherence relations

- The analysis of discourse markers is part of the more general analysis of discourse coherence—how speakers and hearers jointly integrate forms, meaning, and actions to make overall sense out of what is said. (Schiffrin, 1987:49)
- ... the recognition of coherence relations by the hearer or reader enables them to assign coherence to a text. Discourse markers guide the text receiver in the recognition of those relations. (Taboada, 2006: 567-568)
- ...connectives certainly play a role in guiding the interpretation of the relation: the coherence relation that is assumed between the segments must be compatible with the meaning of the connective and with the meaning of the segments. (Sanders, Spooren & Noordman, 1993: 94)

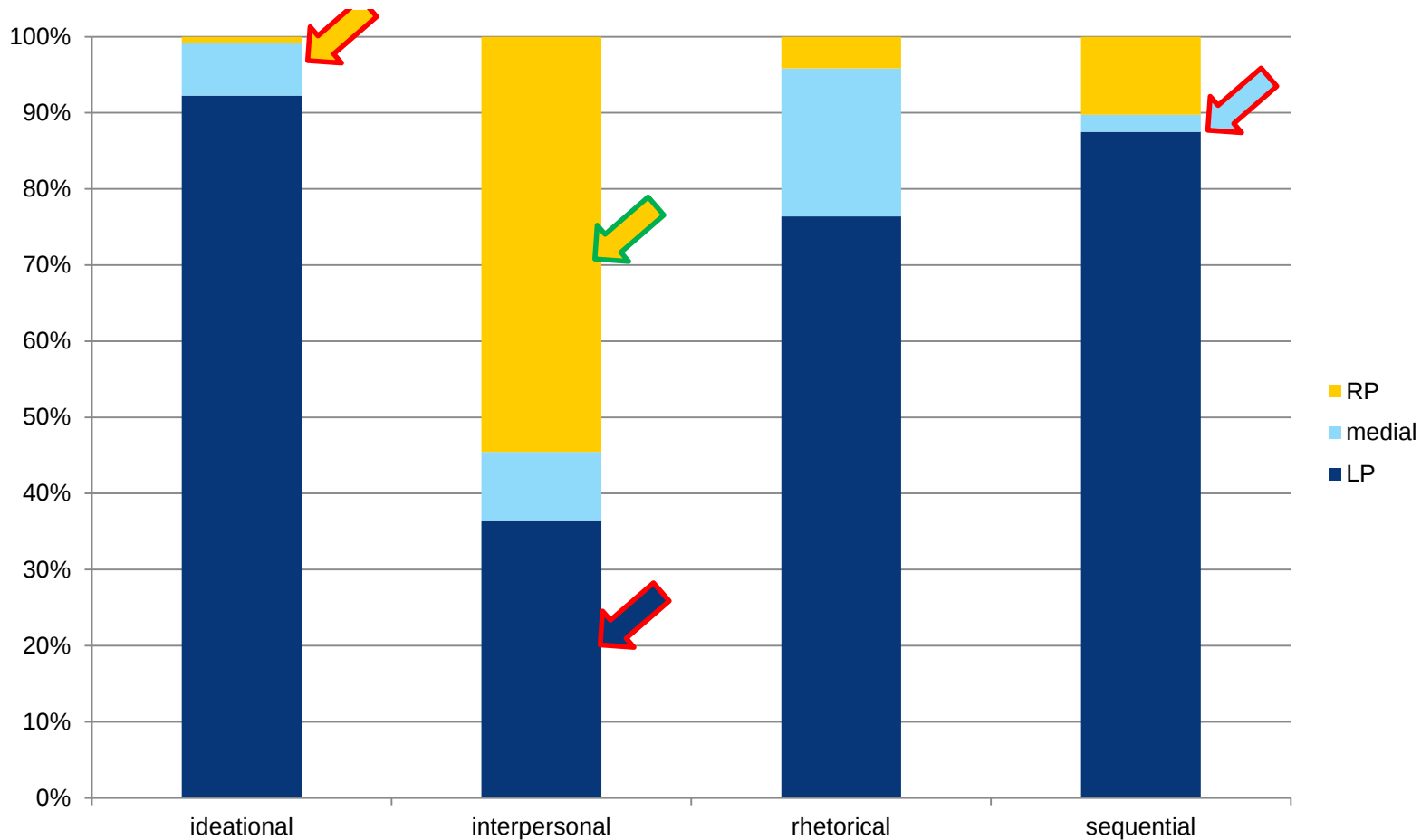
Or not?

- ... theories of discourse based on coherence relations are not adequate in general, and in particular not adequate as a framework for analysing discourse connectives. (Unger, 2006: 420)
- ... there is no evidence that discourse is hierarchically structured beyond the sentence level. (Unger, 2006: 424)
 - No appeal to hierarchical discourse structure is necessary. On the contrary, in a relevance theoretic framework there is no reason to assume discourse segments in the sense of formal entities or units. (Unger, 2006: 428)

Position of DMs in LOCAS-F

	Initial	Medial	Final	Isolated
BDU	697 (39%)	833 (47%)	163 (9%)	87 (5%)
Clause	1345 (83%)	88 (5%)	179 (11%)	/
Intonation Unit	715 (40%)	797 (45%)	181 (10%)	87 (5%)

Functional domain and position in clause



DM types vs. Position in Clause

- Initial (left periphery)
 - 1345 tokens vs. 48 types
 - Top DMs: *et* (467), *mais* (278), *donc* (133), *alors* (79), *ben/bien* (77)
- Medial (parenthetical)
 - 88 tokens vs. 28 types
 - Top DMs: *quand même* (17), *donc* (14), *d'ailleurs* (8)
- Final (right periphery)
 - 179 tokens vs. 29 types
 - Top DMs: *quoi* (42), *hein* (38), *donc* (12), *voilà* (11)

DM in the right periphery

- Unexpected uses in RP

L1: [je sais pas] // <moi> [je ne v/] [non] // [je di/ dis pas de chiffre] //
 <parce que j'en ai euh ///S une bonne dizaine en tête> // <mais> ///C

L2: [ah] // <donc> <quand même> // [une bonne dizaine] ///

L1: I don't know I no I don don't say a figure because I have at least a dozen in my head <mais> ('but') ///

L2: ah so at least a dozen

[j'ai fait des vols] // <enfin> [j'ai volé dans les petit mag/ des petits bonbons dans les magasins] // [j'ai fait mes petits trucs] [c'est jamais des grands trucs euh de grand gangster] // <mais> ///T [j'ai fait mes petites conneries] // <quoi> ///T

I have stolen // well I have stolen in small sh/ sweets in shops // I have done my little things never big gangster things // <mais> I did my stupid things you see

DM types between BDUs

- 87 tokens / 22 types
 - *alors, bon, et, mais, ...*

<mais>_{md} <avant>_{ag} [il me semble qu'il serait bon ///C de réviser de rappeler les courants principaux du dix neuvième siècle]urv ///T <bon>_{md} ///T [dans un premier temps ///C dans un premier temps ///S euh plantons je vais dire le décor ///C à la fois politique scientifique industriel ///C économique du dix neuvième siècle]urv+ ///T

but before this [it seems that it would be good to revise to recall the main currents of the nineteenth century] *bon* ('well; OK') [in the first place in the first place euh let's put I'd say the decor all in all political scientific industrial economical of the nineteenth century]

Functional domain and position in BDU

Domain	Position in BDU				Total
	initial	medial	final	Isolated	
ideational	35 (43%)	45 (55%)	1(1%)	1(1%)	82
rhetorical	30 (26%)	70 (60%)	7 (6%)	9 (8%)	116
sequential	80 (43%)	77 (42%)	14 (8%)	13 (7%)	184
interpersonal	2 (13%)	9 (56%)	4 (25%)	1 (7%)	16
Total	147	201	26	24	398

Functional domain and position in clause

Domain	Position in clause				Total
	LP	Medial	RP	/	
ideational	76 (93%)	2 (2%)	1 (1%)	3 (4%)	82
rhetorical	80 (69%)	22 (19%)	2 (10%)	12 (2%)	116
sequential	141 (77%)	3 (2%)	20 (11%)	20 (11%)	184
interpersonal	5 (31%)	1 (6%)	8 (50%)	2 (13%)	16
Total	147	201	26	24	398

→ DMs play a relational and structuring role at the local level

Conclusions

- Study of discourse segmentation and discourse markers to get insight into what makes discourse so crucial to human communication
- Basic Discourse Units are discourse strategic units reflecting projection of syntactically and prosodically complete units
- Discourse markers are lexical devices functioning as (listener-oriented) signals but also as (hearer-oriented) traces of the discourse production process.
- From case studies to « category » studies
 - Loss of fine-grainedness
 - Gain insight into the system
 - Cross-linguistic robustness of the taxonomy

Conclusions

- The two level annotation in domains and senses is promising
 - Distinction between general discourse functions from specific DM senses
 - New insight into polyfunctionality of DMs → signal and/or trace of discourse production
 - Speaker intentions (illocution / perlocution) → contextual implicatures
 - Core (procedural ?) meanings → explicatures
 - Weak representation of interpersonal domain even in highly interactive contexts
 - Importance of the sequential domain in spoken language

Research perspectives

- What is the cognitive status of BDUs?
 - Units of projection during language production?
 - Fit with models of language/discourse production?
- What makes DMs to signals or symptoms?
 - Cognitive load during language production?
 - Frequency of polyfunctional / underspecified DMs
- How can we cognitively make sense of the difficulty of (discourse) annotation?
 - Semantic underspecification?
 - Relational underspecification?

Research perspectives

- What is the (cognitive) nature of discourse production? Is (spontaneous) spoken discourse speaker-oriented?
“adults routinely process language egocentrically” (Keysar, Barr, and Horton, 1998: 46)

Acknowledgements

- Part of this research was carried out in collaboration with:
 - Anne Catherine Simon, Laurence J. Martin, Ludivine Crible (University of Louvain)
- Financial support from grants:
 - ARC-project « A Multi-Modal Approach to Fluency and Disfluency Markers» granted by the Fédération Wallonie-Bruxelles (grant nr.12/17-044)
 - FRFC-project F 6/15 - MCF/OL - 20.472, Convention nr 2.4524.11 from the Belgian Science Foundation (FRS-FNRS).
- More information on the COST Action TextLink
 - <http://textlink.ii.metu.edu.tr>

References

- Altenberg, Bengt. 1987. *Prosodic Patterns in Spoken English: Studies in the Correlation Between Prosody and Grammar for Text-to Speech Conversion*. Lund University Press.
- Auer, Peter. 2009. "On-Line Syntax: Thoughts on the Temporality of Spoken Language." *Language Sciences* 31 (1): 1–13.
- Auchlin, Antoine, and Anna Ferrari. 1994. "Structuration Prosodique, Syntaxe, Discours : évidences et Problèmes." *Cahiers de Linguistique Française* 15: 187–217.
- Berrendonner, Alain. 2002. "Les Deux Syntaxes." *Verbum* 24 (1-2): 23–36.
- Bilger, Mireille, and Estelle Campione. 2002. "Propositions Pour Un étiquetage En 'Séquences Fonctionnelles'." *Recherches Sur Le Français Parlé* 17: 117–36.
- Blanche-Benveniste, Claire. 2002. "Phrase et Construction Verbale." *Verbum* XXIV (1-2): 23–36.
- Blanche-Benveniste, Claire, Mireille Bilger, C. Rouget, and Karel Van den Eynde. 1990. *Le Français Parlé: études Grammaticales*. Paris: Editions du CNRS.
- Bögels, Sara, and Francisco Torreira. 2015. "Listeners Use Intonational Phrase Boundaries to Project Turn Ends in Spoken Interaction." *Journal of Phonetics* 52 (September): 46–57.
- Carlson, Lynn, Daniel Marcu, and Mary Ellen Okurowski. 2001. "Building a Discourse-Tagged Corpus in the Framework of Rhetorical Structure Theory." In *Proceedings of the Second SIGdial Workshop on Discourse and Dialogue - Volume 16*, 1–10. SIGDIAL '01. Stroudsburg, PA, USA: Association for Computational Linguistics.
- Chafe, Wallace L. 1994. *Discourse, Consciousness, and Time*. Chicago: University of Chicago Press.
- Chafe, Wallace. 2001. "The Analysis of Discourse Flow." In *The Handbook of Discourse Analysis*, edited by Deborah Schiffrin, Deborah Tannen, and Heidi Ehernberger Hamilton, 673–87. Oxford: Blackwell Publishers.
- Couper-Kuhlen, Elizabeth. 2012. "Turn Continuation and Clause Combinations." *Discourse Processes* 49 (3-4): 273–99.
- Crible, Ludivine. in press. "Towards an Operational Category of Discourse Markers: A Definition and Its Model." In *Discourse Markers, Pragmatics Markers and Modal Particles: New Perspectives*, edited by Chiara Fedriani and Andrea Sanso. Amsterdam: Benjamins.

- Crible, Ludivine, Liesbeth Degand, and Gaëtanelle Gilquin. in press. "The Clustering of Discourse Markers and Filled Pauses: A Corpus-Based French-English Study of (dis)fluency." *Languages in Contrast*
- Croft, William. 1995. "Intonation Units and Grammatical Structure." *Linguistics* 33 (5): 839–82. doi:10.1515/ling.1995.33.5.839.
- Degand, Liesbeth. 2014. "'So Very Fast Very Fast Then' Discourse Markers at Left and Right Periphery in Spoken French." In *Discourse Functions at the Left and Right Periphery: Crosslinguistic Investigations of Language Use and Language Change*, edited by Kate Beeching and Ulrich Detges, 151–78. Studies in Pragmatics, Volume 12. Leiden ; Boston: Brill.
- Degand, Liesbeth, Laurence Martin, and Anne-Catherine Simon. 2014. "LOCAS-F : Un Corpus Oral Multigenres Annoté." In *CMLF 2014 - 4 ème Congrès Mondial de Linguistique Française*, 2613–26. Berlin. doi:10.1051/shsconf/20140801211.
- Degand, Liesbeth, and Anne Catherine Simon. 2009. "On Identifying Basic Discourse Units in Speech: Theoretical and Empirical Issues." *Discours* 4: on – line:URL : <http://discours.revues.org/index5852.html>.
- Degand, Liesbeth, and Anne Catherine Simon. 2009. "Mapping Prosody and Syntax as Discourse Strategies: How Basic Discourse Units Vary across Genres." In *Where Prosody Meets Pragmatics: Research at the Interface*, edited by Anne Wichmann, Dagmar Barth-Weingarten, and Nicole Dehé, 79–105. Bingley, UK: Emerald Group Publishing.
- Dell, G.S, and P.M. Brown. 1991. "Mechanisms for Listener-Adaptation in Language Production Limiting the Role of the 'model of the Listener.'" In *Bridges between Psychology and Linguistics*, edited by Donna Jo Napoli and Judy Anne Kegl, 105–29. New York: Academic Press.
- Deulofeu, José. 2003. "L'approche Macrosyntaxique En Syntaxe : Un Nouveau Modèle de Rasoir D'occam Contre Les Notions Inutiles ?"," *Scolia* 16: 77–95.
- Ford, Cecilia, and Sandra Thompson. 1996. "Interactional Units in Conversation : Syntactic, Intonational, and Pragmatic Resources for the Management of Turns." In *Interaction and Grammar*, edited by Elinor Ochs, Emanuel A. Schegloff, and Sandra A. Thompson, 134–84. Studies in Interactional Sociolinguistics 13. Cambridge: Cambridge University Press.
- Graesser, Arthur C, Keith K. Millis, and Rolf A. Zwaan. 1997. "Discourse Comprehension." *Annual Review of Psychology* 48: 163–89.
- Grosz, Barbara J, and Candace L Sidner. 1986. "Attention, Intentions and the Structure of Discourse." *Computational Linguistics Journal* 12 (3): 175–204.
- Givón, Talmy. 1983. *Topic Continuity in Discourse: A Quantitative Cross-Language Study*. Amsterdam/Philadelphia: John Benjamins Publishing.
- Halford, Brigitte. 1996. *Talk Units: The Structure of Spoken Canadian English*. Tübingen: Narr.
- Holsánová, Jana. 2008. *Discourse, Vision, and Cognition*. Human Cognitive Processing 23. Amsterdam: John Benjamins Publishing.

- Kelly, John, and John Local. 1989. "On the Use of General Phonetic Techniques in Handling Conversational Material - Research Database, The University of York." In *Conversation: An Interdisciplinary Perspective*, edited by Derek Roger and Peter Bull, 197–212. Clevedon: Multilingual Matters.
- Keysar, Boaz, Dale J. Barr, and William S. Horton. 1998. "The Egocentric Basis of Language Use Insights From a Processing Approach." *Current Directions in Psychological Science* 7 (2): 46–49. doi:10.1111/1467-8721.ep13175613.
- Kintsch, Walter, and Teun A van Dijk. 1978. "Toward a Model of Text Comprehension and Production." *Psychological Review* 85 (5): 363–94.
- Mann, William, and Sandra Thompson. 1988. "Rhetorical Structure Theory: Toward a Functional Theory of Text Organization." *Text* 8 (3): 243–81.
- Mertens, Piet, and Anne Catherine Simon. 2013. "Towards Automatic Detection of Prosodic Boundaries in Spoken French." In *Proceedings of the Prosody-Discourse Interface Conference 2013 (IDP-2013)*, 81–87. Leuven.
- Monschau, Jacqueline, Rolf Kreyer, and Joybrato Mukherjee. 2004. "Syntax and Semantics at Tone Unit Boundaries." *Anglia - Zeitschrift Fur Englische Philologie* 121: 581–609. doi:10.1515/ANGL.2003.581.
- Oakhill, Jane, Alan Garnham, and Wietske Vonk. 1989. "The on-Line Construction of Discourse Models." *Language and Cognitive Processes* 4 (3-4): S1263–86. doi:10.1080/01690968908406370.
- Pawley, Andrew, and Frances Hodgetts Syder. 2000. "The One-Clause-at-a-Time Hy..." In *Perspectives on Fluency*, edited by Heidi Riggenbach, 163–99. Ann Arbor, USA: University of Michigan Press. <https://researchers.anu.edu.au/publications/7325>.
- Polanyi, Livia. 1988. "A Formal Model of the Structure of Discourse." *Journal of Pragmatics* 12: 601–38.
- Prasad, Rashmi, Bonnie Webber, and Aravind Joshi. 2014. "Reflections on the Penn Discourse Treebank, Comparable Corpora, and Complementary Annotation." *Computational Linguistics* 40 (4): 921–50. doi:10.1162/COLI_a_00204.
- Roulet, Eddy. 1999. *La Description de L'organisation Du Discours*. Langues et Apprentissage Des Langues. Didier.
- Roulet, Eddy, Laurent Filliettaz, Anne Grobet, and Marcel Burger. 2001. *Un Modèle et Un Instrument D'analyse de L'organisation Du Discours*. Berlin: Peter Lang.
- Sanders, Ted, and Wilbert Spooren. 2007. "Discourse and Text Structure." In *Handbook of Cognitive Linguistics*, D. Geeraerts & H. Cuykens, 916–43. Oxford: Oxford University Press.

- Sanders, T. J. M., W. P. M. S. Spooren, and L. G. M. Noordman. 1992. "Towards a Taxonomy of Coherence Relations." *Discourse Processes* 15: 1–35.
- Sanders, T. J. M., W. P. M. Spooren, and L. G. M. Noordman. 1993. "Coherence Relations in a Cognitive Theory of Discourse Representation." *Cognitive Linguistics* 4 (2): 93–133.
- Schiffrin, Deborah. 1987. *Discourse Markers*. Cambridge: Cambridge University Press.
- Simon, A.C., and L. Degand. 2011. "L'analyse En Unités Discursives de Base: Pourquoi et Comment?" *Langue Française*, no. 2: 45–59.
- Swerts, Marc, and Ronald Geluykens. 1994. "Prosody as a Marker of Information Flow in Spoken Discourse." *Language and Speech* 37 (1): 21–43. doi:10.1177/002383099403700102.
- Taboada, Maite. 2006. "Discourse Markers as Signals (or Not) of Rhetorical Relations." *Journal of Pragmatics* 38 (4): 567–92. doi:10.1016/j.pragma.2005.09.010.
- Tanguy, Noalig, Thomas Van Damme, Liesbeth Degand, and Anne-Catherine Simon. 2012. "Projet FRFC ' Périphérie gauche des unités de discours ' - Protocole de codage syntaxique." <https://halshs.archives-ouvertes.fr/halshs-00762866/document>.
- Thompson, Sandra A., and Elizabeth Couper-Kuhlen. 2005. "The Clause as a Locus of Grammar and Interaction." *Discourse Studies* 7 (4-5): 481–505. doi:10.1177/1461445605054403.
- Unger, Christoph. 1996. "The Scope of Discourse Connectives: Implications for Discourse Organization." *Journal of Linguistics* 32 (02): 403–38. doi:10.1017/S0022226700015942.



BDU intermezzo: conversational narration

BDU-I	<moi>ag [(c'est)SV (la fête)SO //C (parce que mon parrain //C c'est comme si c'était mon frère)SRd]urv //C euh<et>md [(comme)SAdv-I]urv-I ///S
BDU-C	[(c'est un peu)SV (les seules la seule occasion //C d'avoir un bébé //C très proche de moi)SO]urv //C <quoi>md ///T
BDU-C	<puisque j'ai pas de frère>ad //C [(j'aurais jamais eu)SV (de neveux //C ou de nièces)SO]urv ///C
BDU-R	<et>md <voilà>md ///C
BDU-X	<puis>md euh<bon>md //C [(quelques mois après)SRg euh (une une fois que je vais chez eux)SRg //C (je les entends)SV ///C (en train de discuter)SO]urv //C [(ah)int]int [(comment)SRg (est-ce qu')mi (on va préparer)SV]urv //S [(comment)SRg (est ce qu')mi (on va demander)SV //C (à //C une copine de de la femme de mon parrain)SO ///C (pour qu'elle soit marraine)SRd]urv //C <et>md [(je les entends)SV //C (préparer toute une surprise //C pour elle //C et tout)SO]urv //C <et>md [(je me dis)SV]urv //C [(ah)int]int ///C

BDU intermezzo: scientific conference



BDU- <alors>md //S euh<le phonostyle>ag //C [(c'est un un objet d'étude
C particulier)SO]urv ///C

BDU- euh [(pour paraphraser ce qui a été dit ce matin euh par Ioannis Kanellos)SRg (on
S peut dire)sv (que le phonostyle comme l'expressivité ///C c'est une question de
norme de réception)so]^{urv} ///C

BDU- <donc>md <évidemment>ag //S [(un phonostyle)SS //S (va être interprété)SV ///S
X euh (par un auditeur)SO]urv //C <et>md //C [(on ne peut pas toujours //C savoir)SV
(à l'avance)SRd //C (quelles sont les caractéristiques du signal de parole qui vont
///S euh qu'il va associer à un phonostyle particulier)SO]urv ///C

BDU- <et>md <la fonction du phonostyle>ag //C [(c'est une fonction euh extra euh
I phonologique)SO]urv <si on veut>ad //C <puisque elle est destinée à montrer //C où
à faire exister une communauté d'appartenance>ad [(j'ai dit)SV (tout à l'heure)SRd
//C (une profession //C ou des choses comme ça)SO]urv ///T

BDU- euh<mais>md <pourtant>md ///C
R