"Towards a Model for Discourse Marker Annotation in spoken French: From potential to feature-based discourse markers"

Bolly, Catherine ; Crible, Ludivine ; Degand, Liesbeth ; Uygur, Deniz

Abstract

Starting from the common observation that there is no recognized closed class of discourse markers (DMs) and that a number of linguistic markers may or may not count as DMs according to the definitions at stake (Schourup 1999: 228), we aim to present an empirical method for the identification and annotation of DMs in spontaneous spoken French (MDMA project). Central to our proposal is that DMs may be described as clusters of features that, in specific patterns of combination, allow distinguishing DM use from other uses. We proceeded in three steps: (i) using a very broad definition of DMs – i.e. items that "provide instructions to the hearer on how to integrate their host utterance into a developing mental model of the discourse in such a way as to make that utterance appear optimally coherent" (Hansen 2006: 25) – three analysts identified all potential DMs in an 800 words transcript; (ii) all types found were then extracted from a balanced 10,000 words corpus; and (iii) anal...

Document type : *Communication à un colloque (Conference Paper)*

Référence bibliographique

Bolly, Catherine ; Crible, Ludivine ; Degand, Liesbeth ; Uygur, Deniz. *Towards a Model for Discourse Marker Annotation in spoken French: From potential to feature-based discourse markers.*Pragmatic Markers, Discourse Markers and Modal Particles: What do we know and where do we go from here? (Università dell'Insubria, Como, du 16/10/2014 au 17/10/2014).

Towards a Model for Discourse Marker Annotation in spoken French:
From potential to feature-based discourse markers

Catherine Bolly

Ludivine Crible

Liesbeth Degand

Deniz Uygur-Distexhe

Identifying discourse markers in spontaneous speech

- Common ground:
  - Lack of consensus

    « 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)

  - Lack of satisfying models (see Uygur-Distexhe 2012 and Crible 2014 for exceptions)
  - Proliferation of case studies

➢ Need for contrastive, categorical, onomasiological studies

➢ Towards a comprehensive model of annotation

- From identification to description in context
- Corpus-based selective criteria for inclusion/exclusion of DMs

Discours VALIBEL Variation

UCL Université catholique de Louvain

Plan

1. DM Definition

2. MDMA Research group

3. Method : an annotation experiment

4. Results : features and clusters

5. Perspectives

DM definition

What are discourse markers?

"A pragmatic marker* is defined as a ❶ **phonologically short item** that is ❷ **not syntactically connected** to the rest of the clause (i.e., is parenthetical), and has ❸ **little or no referential meaning** but serves ❹ **pragmatic or procedural purposes**" (Brinton, 2008: 1)

DMs "provide **instructions** to the hearer on how to integrate their host utterance into a developing mental model of the discourse in such a way as to make that utterance appear optimally coherent" (Hansen 2006: 25)

- **Multifunctionality** within their "meaning potential" (Aijmer 2013) and in context
- **Metalinguistic** comments on the content/form of the linguistic utterance itself
- **Intersubjectivity** effects on interaction management

UCL
Université catholique de Louvain

Example

… **et** anorexique je ne parvenais pas à le retenir / j'ai / **alors** je pense à quelque ch/ je pensais à anus (rires) / **comme** c'est **quand même** le tube digestif **hein** qui est en bas (rires) **et ça va** depuis lors je n'oublie plus (rires) **et encore** l'autre jour **aussi** un mot / **tiens** je ne sais p/ **tu vois** / si / j'ai / j'oublie certains mots / **'fin** / je retombe dessus après **hein**…

*… **and** anorexic I couldn't memorize it / I / **so** I think of someth/ I thought of anus (laughing) / **since** it's **still** the digestive tube **right** which is at the bottom (laughing) **and it's ok** since then I don't forget anymore (laughing) **and again** the other day **too** a word / **see** I don't kn/ **you see** / yes / I / I forget certain words / **well** / I remember them afterwards **right**…*
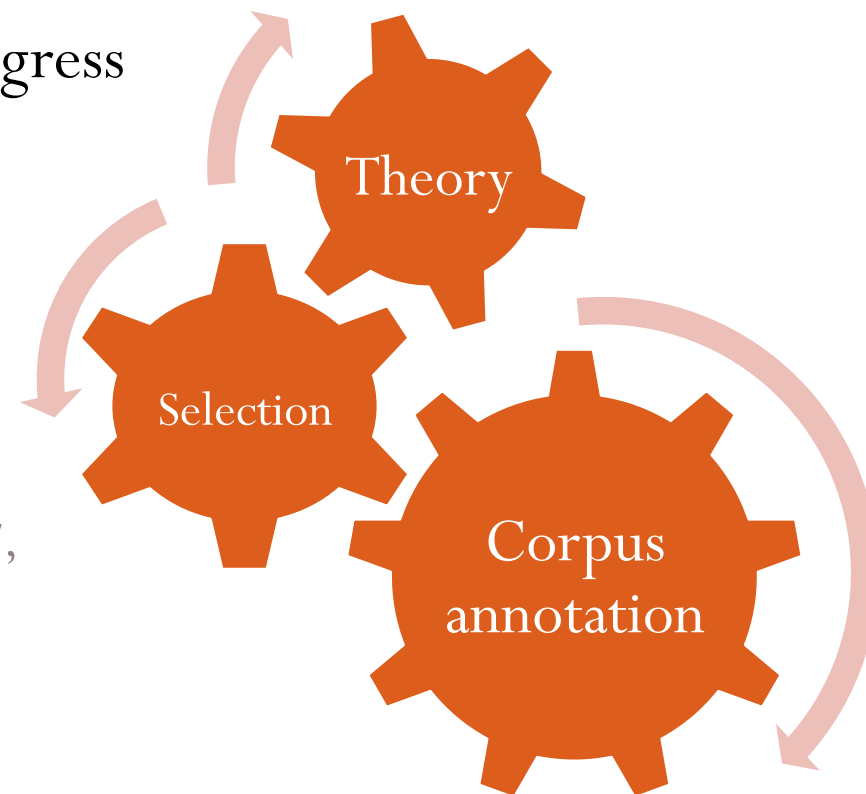
ID code: ageJM1; Pseudo: Jeanne; Age: 90; Source: Corpage 2012; Task 1 Line life; Time code: 1:13:10

UCL
Université catholique de Louvain

MDMA Research group

- Check the relevance/reliability of identifying features of DMs
- Towards objective features to be automated (POS-tags, co-occurrence, position)

- Method overview :

1. Intuitive selection of potential DMs

2. Extraction of potential types in balanced corpus FR-BE

3. Elaboration of the coding scheme from theory

4. Annotation and revision of the scheme

5. Statistical analysis of « predictive » parameters

Discours VALIBEL Variation

UCL
Université
catholique
de Louvain

3. Elaboration of the coding scheme from theory

- What to code and how ?
- Selection of parameters to code from existing definitions
- Constant work-in-progress

(Schourup 1999, Schiffrin 1987, Brinton 2008…)

Theory

Selection

Corpus annotation

Discours VALIBEL Variation

UCL
Université catholique de Louvain

4. Annotation and revision of the scheme

- **Syntax**
- **Semantics**
- **Co-text**
- **Degree of certainty**

| TEXT L | ITEM 'token' | CONTEXT R | Certainty | Certainty | SYNTAX | | | | | SEMANTICS | | | LEXICAL | PROSODY |
Left context	Search item	Right context		MEAN	Category	Position	Turn	Mobility	Optionality	Coded meaning	Contextual m	Procedurality	Cooccurrenc	Pause
ben oui enfin bon	disons que	je (bruits de voix) / je	1	2,25	VP	INIT	other	yes	yes	Énoncer un prop	no	procedural	left	no
10 ouais ageNM1 $	donc	euh // moi j'ai retrou	3	3,00	CONJ	PRE	initial	no	yes	Exprime la consé	yes	procedural	right	no
ilrMFO mm ilrCP1	moi	il y a il y a plusieurs	1	1,75	PRON	PRE	initial	no	yes	Pronom personne	yes	concept-proced	no	no
vite // $ et alors là	aussi	/ on se rend compte a	1	2,00	ADV-P	INIT	other	no	yes	Exprime l'idée qu	yes	procedural	left	right
i / parce que oui /	ma foi	/ on r/ on réfléchit pl	2	2,50	NP	PRE	other	yes	yes	pour appuyer, as	yes	procedural	left	leftright
e créer euh ilrMFO	mm	ilrCP1 un document c	2	2,25	PART	AUTON	Turn	N/A	N/A	S'emploie pour m	no	procedural	no	no
stion `fin (si j' $ ai	bien	compris $) CG et sur	1	1,00	ADV-P	END	other	no	no	En rapport avec c	yes	conceptual	no	no
sur la nationalité	aussi	y a les [deux $ (inau	1	1,25	ADV-P	END	other	no	no	Exprime l'idée qu	yes	procedural	no	no
ilcMFO mm ilcDA1	mais	la hiérarchie doit rép	3	2,25	CONJ	INIT	initial	no	yes	1) "Mais" coord	yes	procedural	no	no
its bilingues peu/	parce que	ils ont devant eux des	3	2,50	CONJ	END	other	no	yes	marque une relat	yes	procedural	no	no
de voir les choses	euh	sa sa euh il y a une c	3	2,50	PART	END	other	yes	yes	Hésitation: 1) Ret	yes	procedural	right	no
(00:00:30:59) DIR	hein	$ / si encore vous rep	3	2,75	PART	?	initial	no	yes	Enoncé en réactic	no	procedural	no	right
nd compte / même	avec	les gosses / hein / $ c	1	1,25	PREP	END	other	no	no	Marque l'accomp	no	procedural	right	no
ein sur Macintosh	et cetera	$ / oui hein ces derni	1	1,50	ADV-P	POST	other	no	yes	Au terme d'une ér	yes	procedural	leftright	right

Discours VALIBEL Variation

UCL Université catholique de Louvain

4. Annotation and revision of the scheme

- **Syntax** :

  - Category

*« ben **c'est-à-dire que** bon comme bon quand je suis ... »* [« that is to say »] ➜ VP

  - Position in the utterance (Lindström 2001)

*« si j'ai **bien** compris »* [« correctly »] ➜ middle field

  - Position in turn

  - Mobility

*« il y avait d'autres problèmes **non** mais c'est vrai »* [« no but »] ➜ no

Discours VALIBEL Variation

UCL
Université catholique de Louvain

4. Annotation and revision of the scheme

- **Syntax**

- **Semantic**s :

  - Procedurality

*« non mais **faut dire qu**' en fait il a réfléchi »* [« the thing is »VS. « we need to say that »] ➔ conceptual-procedural

  - Coded meaning

*« j'aimerais **bien** travailler dans ce genre de matériel »* ➔ yes

4. Annotation and revision of the scheme

- **Syntax**

- **Semantics**

- **Co-text** :

  - Co-occurring DM

  - Pauses

4. Annotation and revision of the scheme

- **Syntax**
- **Semantics**
- **Co-text**
- **Degree of certainty** :

*« hmm **oui** je pensais que on fait avec ça »* ➜ 2, 3, 1, 1 ➜ 1,75 (mean)

Discours **VALIBEL** Variation

UCL
Université
catholique
de Louvain

4. Annotation and revision of the scheme

- **Syntax** :
  - Category : VP, clause, NP, adv …
  - Position in utterance : pre-field, initial, middle, end, post-field
  - Position in turn : initial, medial, final, autonomous
  - Mobility : yes / no
  - Optionality : yes / no

- **Semantics** :
  - Procedural meaning : procedural, conceptual, conc-proc
  - Prototypical meaning or not : yes / no

- **Co-text** :
  - Co-occurrence : no / left / right / both
  - Pause : no / left / right / both

- **Degree of certainty** : 1 – 3

> Kappa-scores from
> 0,75 to 0,82

UCL
Université catholique de Louvain

From theory to data and back again

- ➤ Corpus-based annotation scheme, moulded by the annotation experiment
- ➤ Scope over both DM tokens and their non-DM forms
- ➤ Qualitative alternative to kappa-scores

Statistical processing :

- Relative relevance of variables
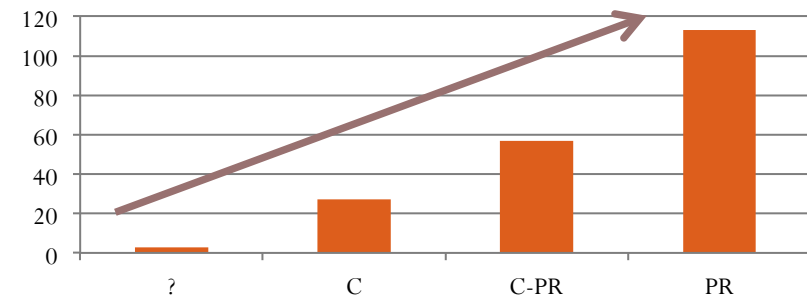- Clusters of co-occurring features

Discours VALIBEL Variation

UCL
Université catholique de Louvain

Overall distribution of some parameters

Syntactic category
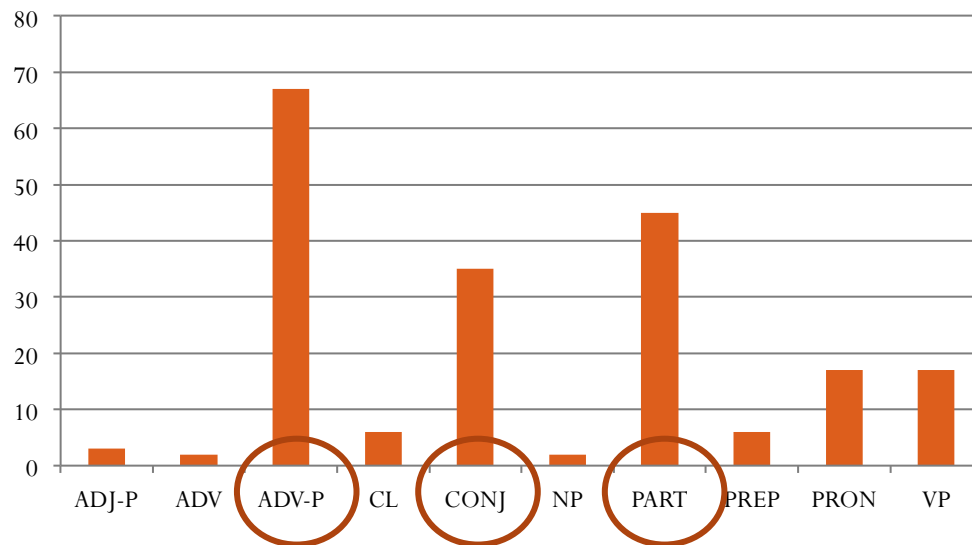
Degree of certainty (mean score)	% of tokens
1 - 1,25	34 %
1,5 - 2,5	46,5 %
2,75 - 3	19,5 %

Procedural meaning

➤ Both DMs & non-DMs

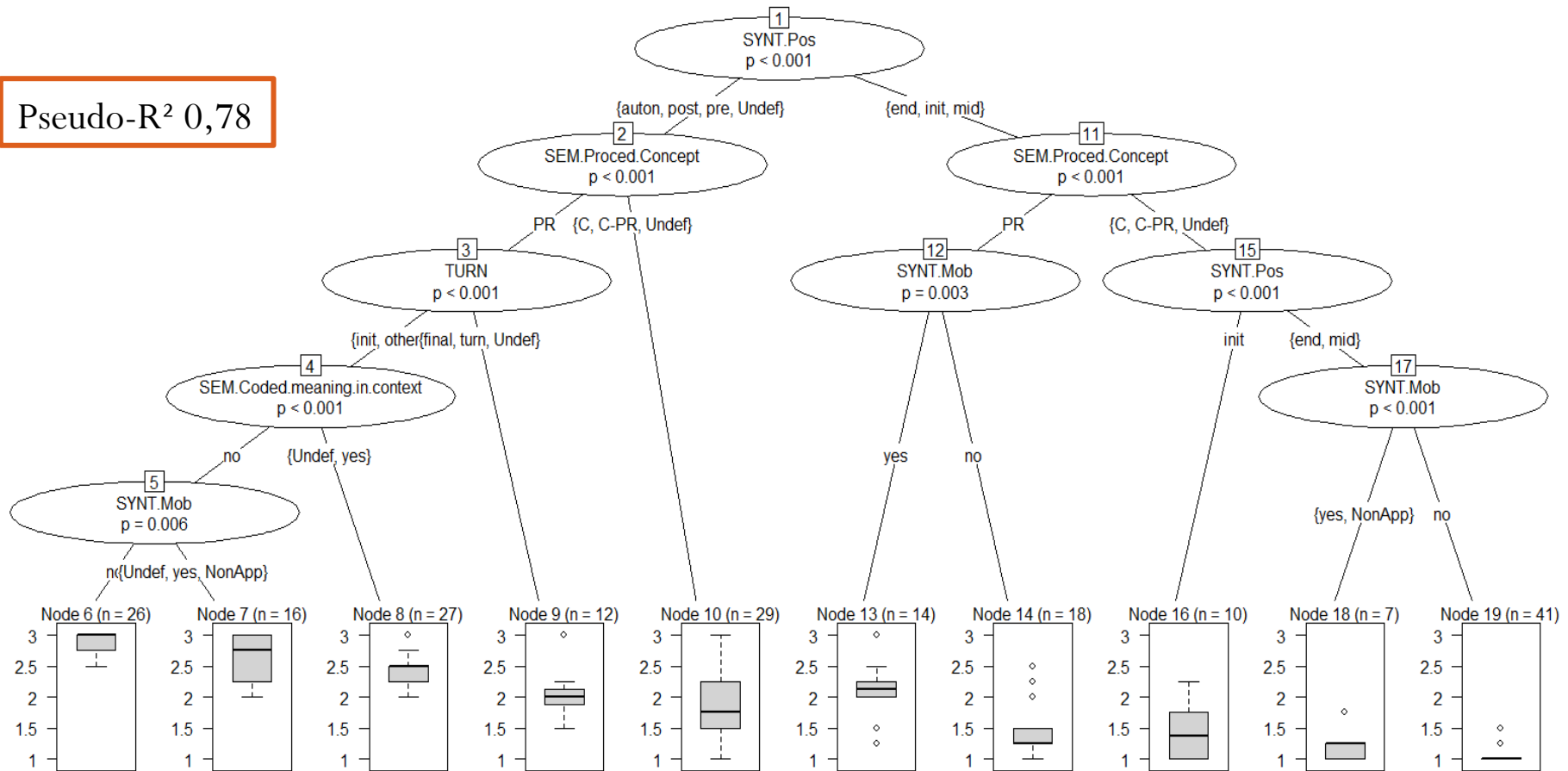# Ranking the variables

Conditional importance of variables
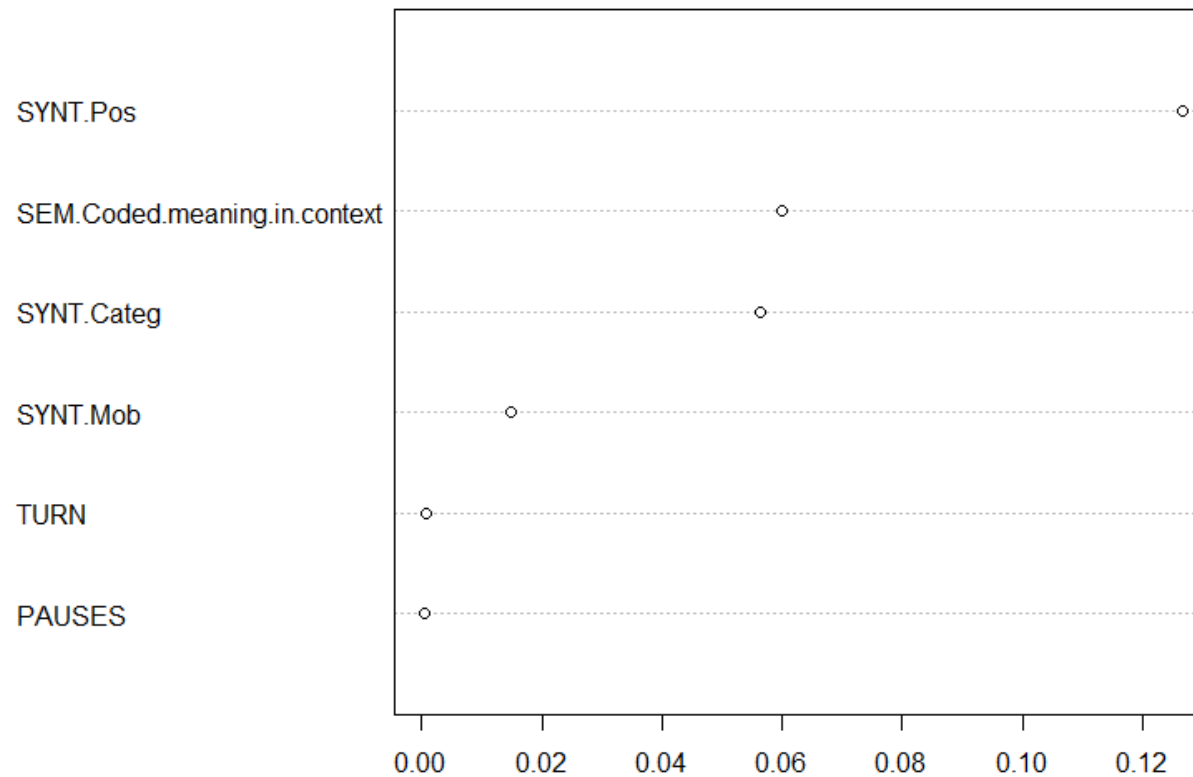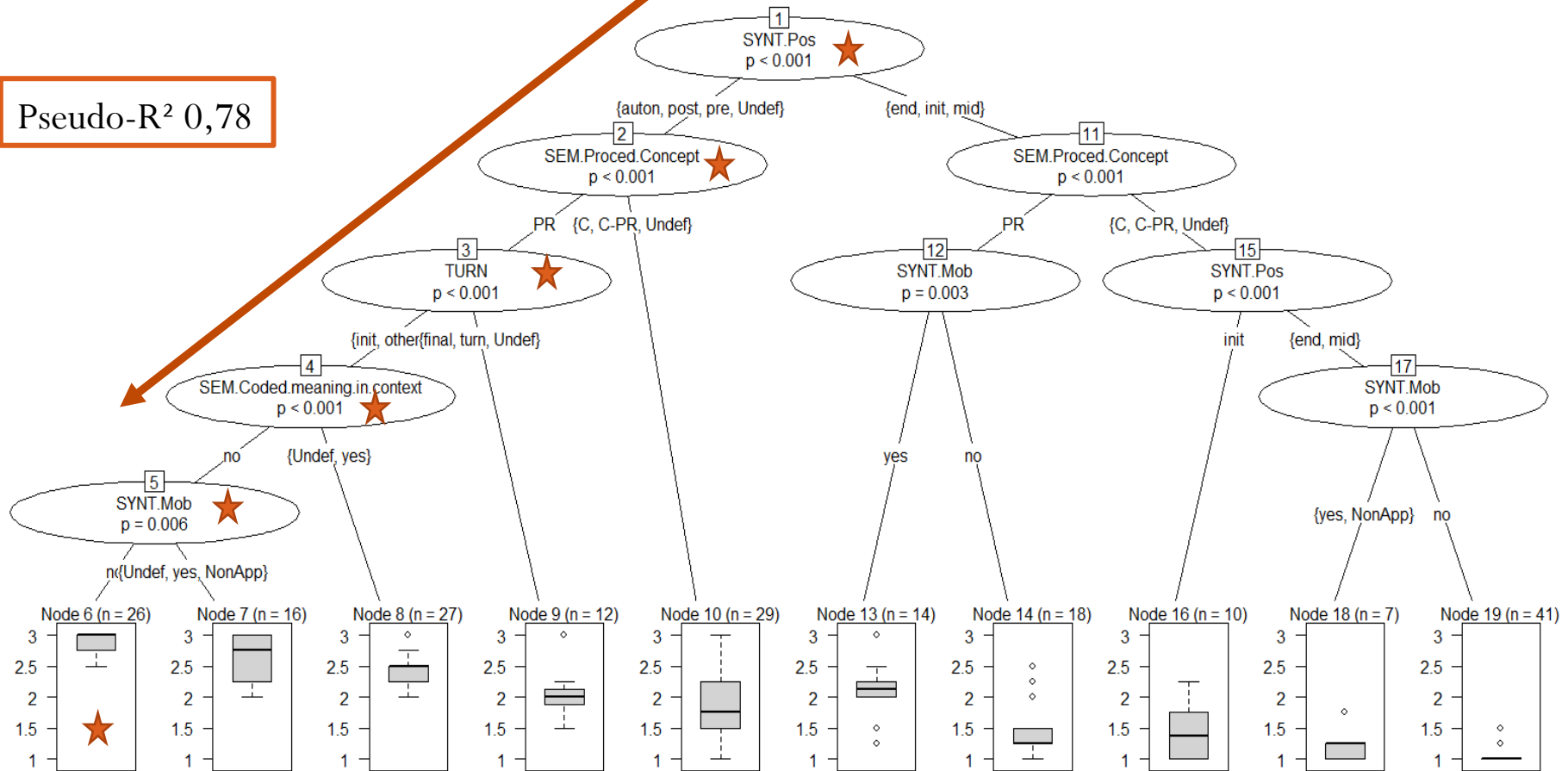

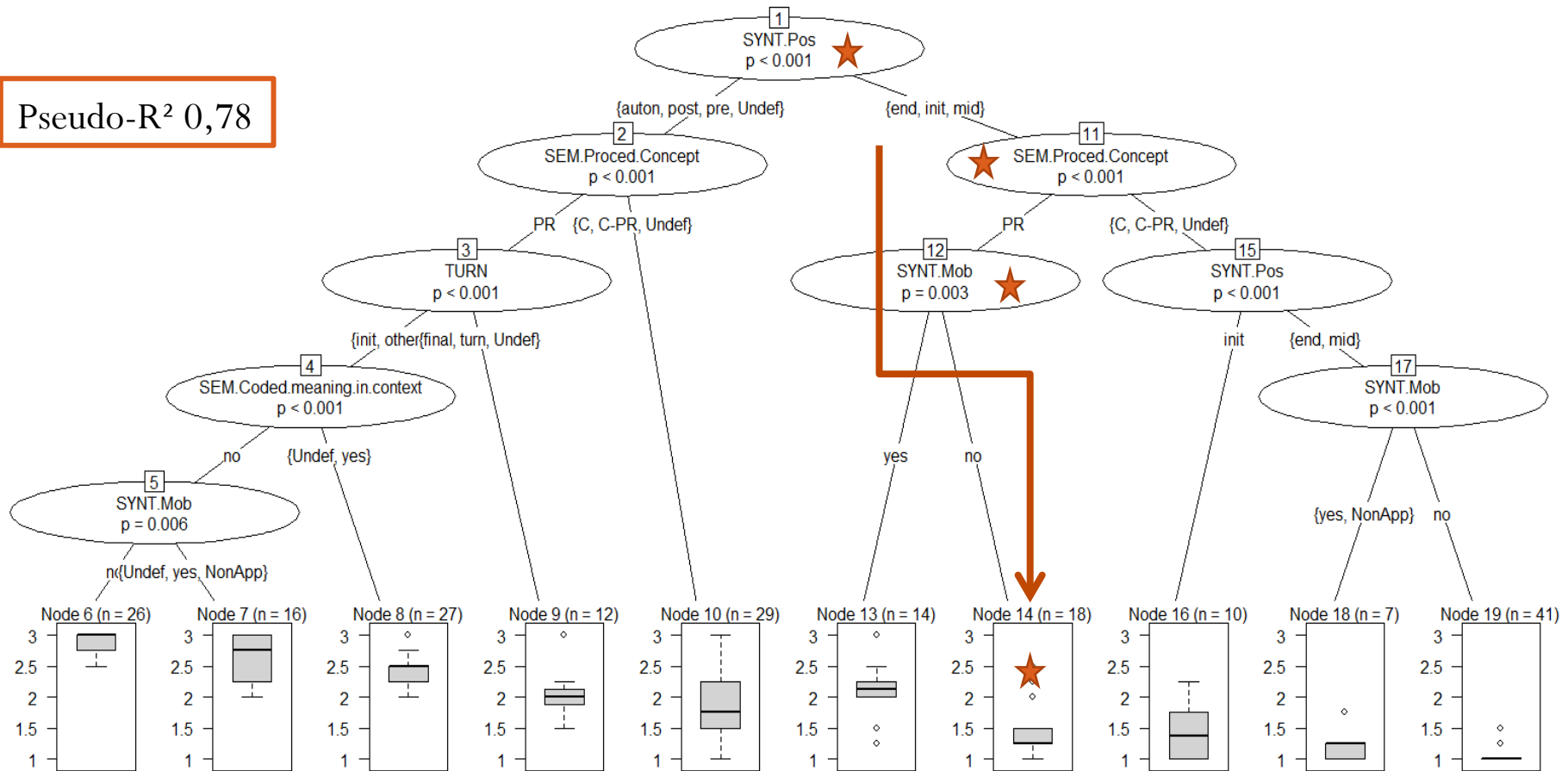Procedurality removed because circular

Predict DM status on feature patterns

Pseudo-R² 0,78
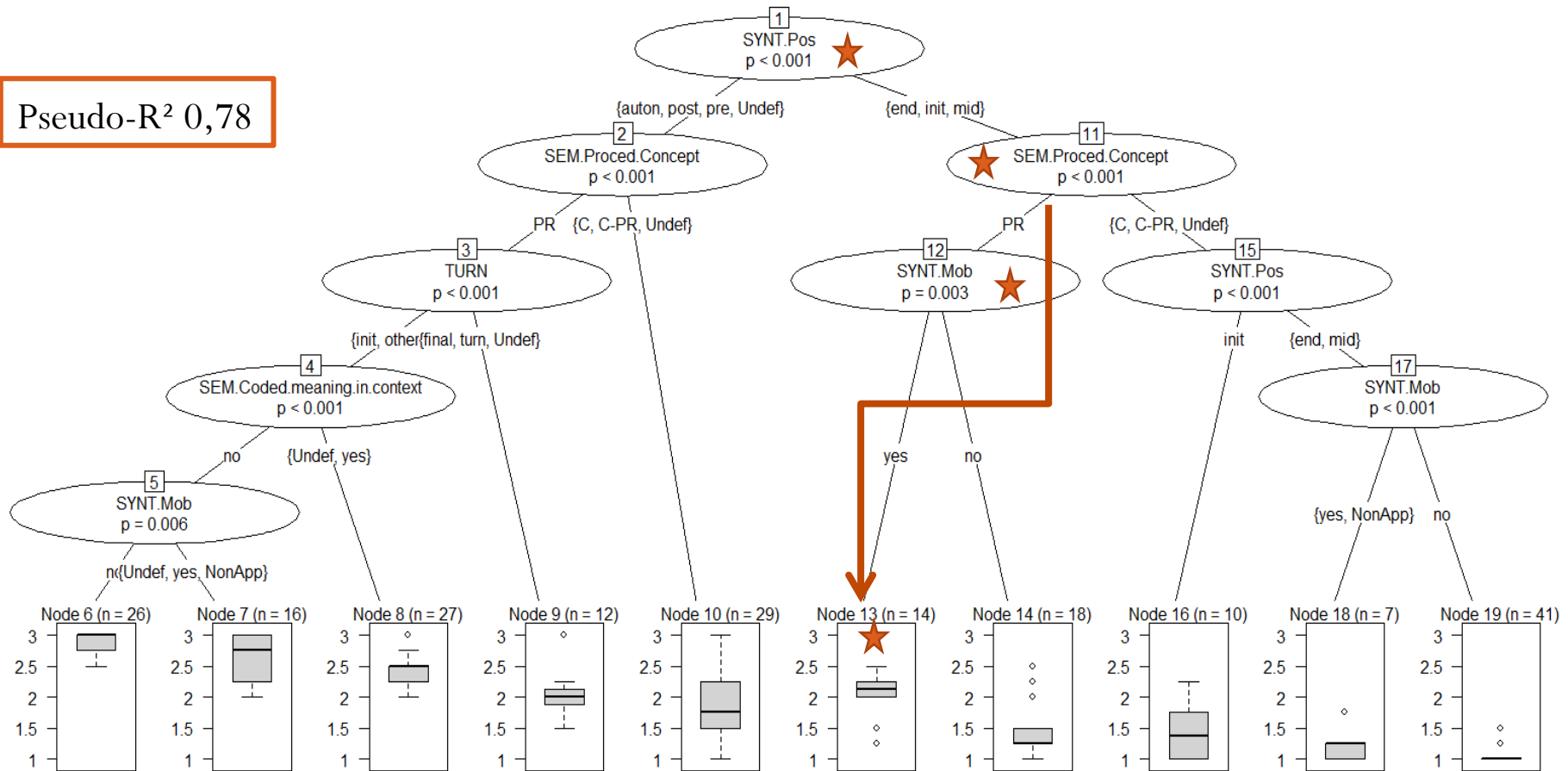
Predict DM status on feature patterns

Pseudo-R² 0,78

Predict DM status on feature patterns

Pseudo-R² 0,78

Predict DM status on feature patterns

Confirmed vs discarded DMs

Typical features of confirmed DMs :

- Conjunctions > particles > adverbs
- Pre-field
- Not coded meaning
- Procedural meaning

*et alors c' est très compliqué de pouvoir régler ces problèmes **donc** c' est même je dirais c'est presque plus du français*

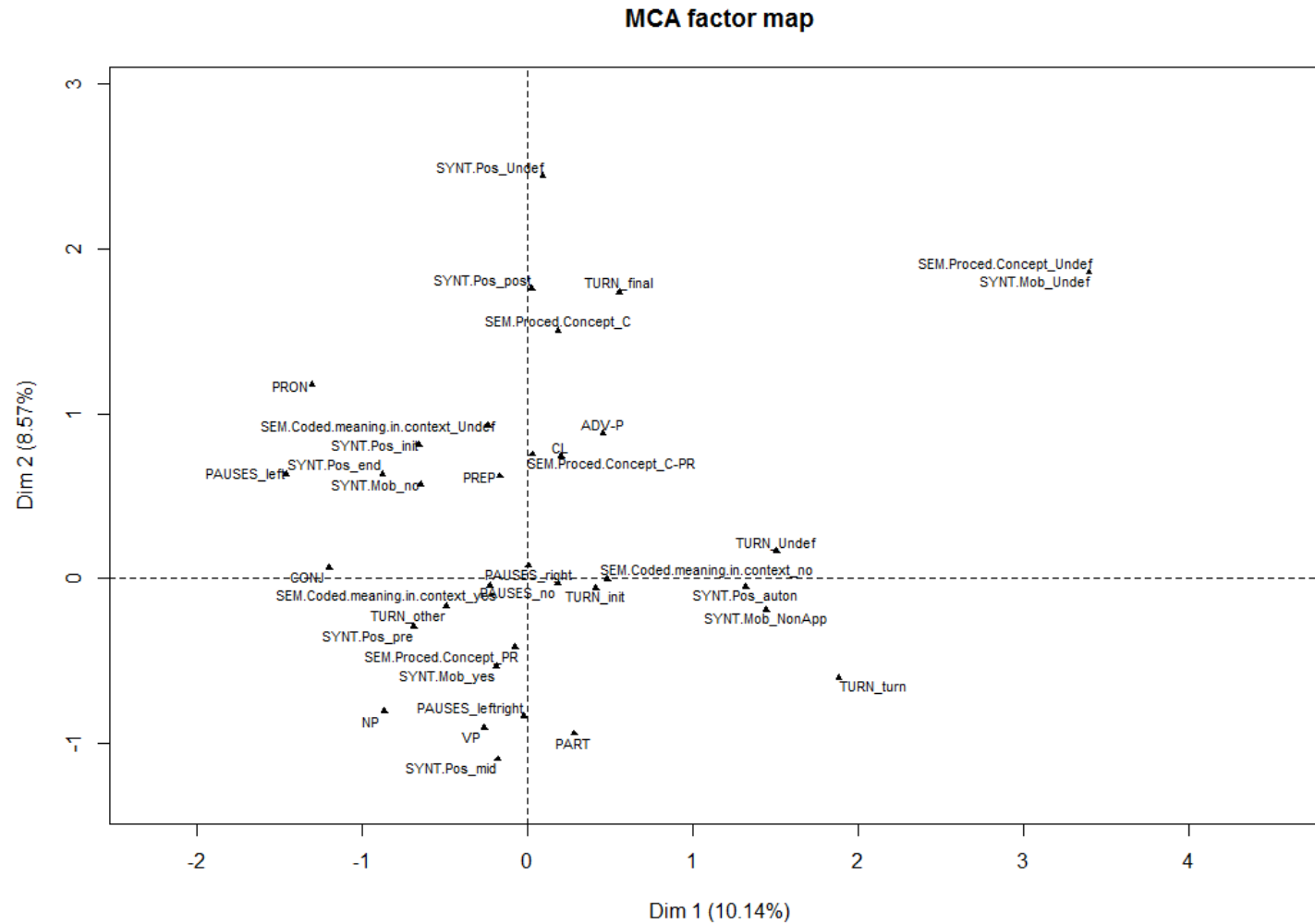## Typical features of discarded DMs :

- Adverbs > pronouns > VP

- Middle field, end field
- Coded meaning
- Conceptual-procedural

*j' aimerais **bien** travailler dans ce genre de matériel*

High-profile clusters of features

MCA factor map
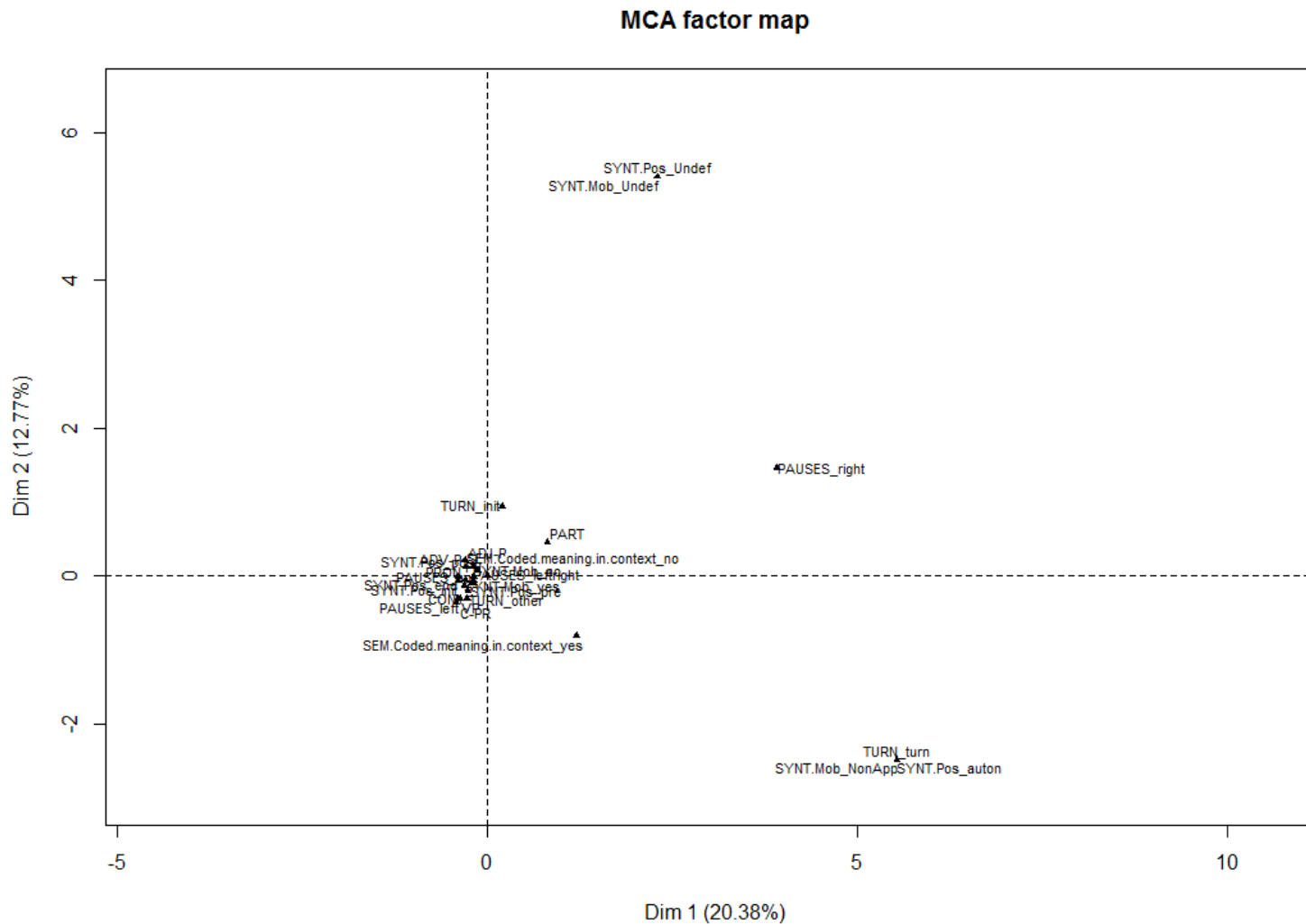
Borderline clusters of features

MCA factor map

Borderline cases

From 1.5 to 2.5 (included) degree of certainty

➤ Other PMs ?

- Modal adverbs : *vraiment, encore, bien sûr*

- Response markers : *c'est ça, c'est juste, oui, non, voilà*

- Hesitation particles : *euh*

➤ Typical features :

- Autonomous position

- Procedural meaning

Conclusion : a confirmatory method for operational annotation scheme

- Novelty of the approach to both DMs and their propositional (non-DMs) equivalents

- Provides corpus-based validation of the annotation model and its parameters

- Visualization of clusters of co-occurring features

Discours VALIBEL Variation

UCL
Université
catholique
de Louvain

DMs profiling

- Corpus-based profiles of typical features
  - Intuition-based potential DMs either confirmed or discarded
  - Borderline cases : different types of pragmatic markers

- Overall, syntactic position stands out as most predictive
  - Pre-field, i.e. not integrated : main difference with MPs

Discours VALIBEL Variation

UCL
Université
catholique
de Louvain

Perspectives

- Functional annotation (forthcoming)
  - same method but for confirmed and potential DMs only (>1,5)

- Test-bed for application of our corpus-based scheme to other data types :
  - Other corpora, genres, languages, modalities

- Semi-automatic annotation ?
  - Surface features : more objective
  - Still difficult to automate

Discours VALIBEL Variation

UCL
Université
catholique
de Louvain

References

Aijmer, K. 2013. *Understanding pragmatic markers. A variational pragmatic approach*. Edinburgh University Press.

Bolly, C. & Crible, L. (submitted). « From context to functions and back again: disambiguating pragmatic uses of discourse markers ». Paper submitted to the 14th International Pragmatics Conference (IPrA), July 26-31, Antwerp, Belgium.

Brinton, L. 2008. *The comment clause in English: Syntactic origins and pragmatic development*. Cambridge : CUP.

Crible, L. 2014. « Selection and functional description of discourse markers in French and English: Towards crosslinguistic and operational categories for contrastive annotation ». Paper presented at the International Workshop "Pragmatic Markers, Discourse Markers and Modal Particles: What do we know and where do we go from here?" October 16-17, Como, Italia.

Degand, L., Cornillie, B., Pietrandrea, P. 2013. *Discourse markers and modal particles. Categorization and description*. Amsterdam : John Benjamins.

Hansen, M.-B. M. 2006. « A dynamic polysemy approach to the lexical semantics of discourse markers (with an exemplary analysis of French *toujours*) ». In K. Fischer (ed.), *Approaches to discourse particles*. Amsterdam : Elsevier. 21-41.

Lindström, J. 2001. « Inner and outer syntax of constructions: the case of the *x och x* construction in Swedish ». Paper presented at the 7th International Pragmatics Conference, July 9-14, Budapest, Hungary.

Uygur-Distexhe, D. 2012. « *Lol, mdr* and *ptdr*. An inclusive and gradual approach to discourse markers. » *Linguisticae Investigationes* 35/2, 389-413.

Discours **VALIBEL** Variation

| Introduction | DM definition | MDMA | Method | Results | Conclusion |

UCL
Université
catholique
de Louvain