

# Table of contents

List of figures	IX
List of tables	XI
List of abbreviations and acronyms	XIII
Acknowledgments	XV
CHAPTER 1	
Introduction	1
1.1 Fluency in time and space	1
1.2 Background and objectives	4
1.3 Preview of the book	5
CHAPTER 2	
Definitions and corpus-based approaches to fluency and disfluency	9
2.1 Disfluency or repair? Levelt's legacy	10
2.2 Holistic definitions of fluency	13
2.3 Componential approaches to fluency and disfluency	14
2.3.1 Qualitative components of perception	14
2.3.2 Quantitative components of production	16
2.3.3 Götz's qualitative-quantitative approach	20
2.4 Synthesis: Definition adopted in this work	22
2.5 A usage-based account of (dis)fluency	23
2.5.1 Key notions in usage-based linguistics	24
2.5.2 From schemas to sequences of fluencemes	24
2.5.3 Variation in context(s)	26
2.5.4 Accessing fluency through frequency	28
2.6 Summary and hypotheses	30
CHAPTER 3	
Definitions and corpus-based approaches to discourse markers	33
3.1 From connectives to pragmatic markers: Defining the continuum	34
3.2 Discourse markers in contrastive linguistics	37
3.3 Models of discourse marker functions	40
3.3.1 Discourse relations in the Penn Discourse TreeBank 2.0	40
3.3.2 The many scopes of DM functions	43

- 3.4 “Fluent” vs. “disfluent” discourse markers 47
  - 3.4.1 DM features and (dis)fluency 47
  - 3.4.2 Previous corpus-based accounts of DMs and disfluency 48
- 3.5 Summary and hypotheses 52

#### CHAPTER 4

### Corpus and method 55

- 4.1 The *DisFrEn* dataset 55
  - 4.1.1 Source corpora 55
  - 4.1.2 Comparable corpus design 57
  - 4.1.3 Corpus structure in situational features 59
- 4.2 Discourse marker annotation 61
  - 4.2.1 Identification of DM tokens 62
  - 4.2.2 Functional taxonomy 64
  - 4.2.3 Three-fold positioning system 66
  - 4.2.4 Other variables 69
  - 4.2.5 Annotation procedure 70
- 4.3 Disfluency annotation 71
  - 4.3.1 Simple fluencemes 72
  - 4.3.2 Compound fluencemes 73
  - 4.3.3 Related phenomena and diacritics 75
  - 4.3.4 Annotation procedure 76
  - 4.3.5 Macro-labels of sequences 78
- 4.4 Summary 79

#### CHAPTER 5

### Portraying the category of discourse markers 81

- 5.1 Distribution across languages and registers 81
  - 5.1.1 General frequency 82
  - 5.1.2 The status of tag questions 83
  - 5.1.3 Register variation 83
  - 5.1.4 A greater effect of register over language? 85
  - 5.1.5 DM expressions in contrast 85
  - 5.1.6 Diversity hypothesis 87
- 5.2 Position of DMs: Initiality in question 89
  - 5.2.1 Clause-initial DMs 89
  - 5.2.2 Utterance-initial DMs 90
  - 5.2.3 Turn-initial DMs 91
  - 5.2.4 Non-initial DMs 93
  - 5.2.5 Interim summary on position 97

5.3	Domains and functions: Frequency and diversity	98
5.3.1	Single domains	98
5.3.2	Single functions	107
5.3.3	Double domains and functions	111
5.4	Integrating syntax and pragmatics	113
5.5	Co-occurrence of DMs	119
5.5.1	Co-occurrence across languages and registers	120
5.5.2	Co-occurrence across positions	122
5.5.3	Integrated statistical model of co-occurrence	124
5.6	Summary	125
5.7	Interim discussion: The potential of bottom-up research	126

## CHAPTER 6

### Disfluency in interviews 129

6.1	Data	129
6.2	Fluenceme rates in English and French	130
6.2.1	Number of tags	130
6.2.2	Number of tokens	131
6.2.3	Radio vs. face-to-face interviews	133
6.3	Clustering tendencies	136
6.3.1	Isolation vs. combination	136
6.3.2	Most frequent clusters	137
6.3.3	DMs in clusters	138
6.4	Fluency as frequency	139
6.4.1	Frequency and structural complexity	139
6.4.2	Frequency and sequence length	142
6.5	Summary	146

## CHAPTER 7

### The (dis)fluency of discourse markers 149

7.1	Sequence types across registers	149
7.1.1	"Cluster"	150
7.1.2	"Sequence category"	152
7.1.3	"Internal structure"	156
7.1.4	Sequence-specific DMs	158
7.2	Sequence types across DM features	159
7.2.1	Disfluency and functional domain	159
7.2.2	Disfluency, domain and position	162
7.2.3	Synthesis of variables	165
7.3	Potentially Disfluent Functions	166

7.3.1	PDFs across registers	167
7.3.2	PDFs and sequence types	169
7.3.3	PDFs and sequence structure	171
7.4	Summary	174
7.5	Interim discussion: The “silence” of corpora	175
CHAPTER 8		
Discourse markers in repairs		177
8.1	Previous approaches to repair	178
8.1.1	Reformulation and its markers: The French classics	178
8.1.2	Contrastive perspectives on reformulation markers	180
8.1.3	From reformulation to repair: Levelt’s (1983) typology of repair	184
8.1.4	Research questions and hypotheses	186
8.2	Data and method	187
8.2.1	Selection criteria	188
8.2.2	Repair category	188
8.2.3	Relation to annotated fluencemes	190
8.2.4	Intra-annotator agreement	191
8.3	Repair categories across languages	191
8.4	DMs in repairs	193
8.4.1	Position of the DMs	193
8.4.2	DM lexemes	195
8.4.3	Potentially Disfluent Functions in repairs	196
8.4.4	Specification and enumeration	198
8.5	DMs and modified repetitions	200
8.6	Summary	201
8.7	Interim discussion: Low quantity, high quality?	203
CHAPTER 9		
Conclusion		207
9.1	Summary of the main findings	207
9.2	General discussion	210
9.3	Implications and research avenues	212
Bibliography		215
Appendices		
Appendix 1. Discourse markers by register		233
Appendix 2. List of discourse markers in <i>DisFrEn</i> and their functions		235
Appendix 3. List of functions in <i>DisFrEn</i> and their discourse markers		245
Appendix 4. Top-five most frequent functions by register in <i>DisFrEn</i>		249
Index		251