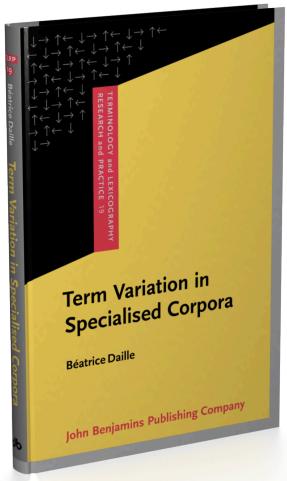
Terminology and Lexicography Research and Practice 19

Table of contents

https://doi.org/10.1075/tlrp.19.toc



Pages v–x of
Term Variation in Specialised Corpora: Characterisation, automatic discovery and applications
Béatrice Daille
[Terminology and Lexicography Research and Practice, 19]
2017. xii, 272 pp.

© John Benjamins Publishing Company

This electronic file may not be altered in any way. For any reuse of this material written permission should be obtained from the publishers or through the Copyright Clearance Center (for USA: www.copyright.com).

For further information, please contact rights@benjamins.nl or consult our website at benjamins.com/rights

Table of contents

Acknowledgements xi

CHAPTER 1
Introduction 1
1.1 Preliminary example 3
1.2 Variants and terminological analysis 4
1.3 The automatic detection of variants 6
1.4 Variants and applications 7
1.5 Typographical conventions 7

Part I. Characterisation

CHAPTER 2
Definitions 11
2.1 Term 11
2.2 Derivation 14
2.3 Compounding 15
2.3.1 Morphological compounds 16
2.3.2 Border between derivation and compounding 18
2.3.3 Syntagmatic compounds 19
2.3.4 Border between morphological and syntagmatic compounds 20
2.4 Borrowing 21
2.5 Term patterns 22
2.5.1 Simple term patterns 23
2.5.2 Morphological compound patterns 23
2.5.3 Syntagmatic compound patterns 24
2.5.4 Frequency of term patterns 27
2.6 Term variants 29
2.6.1 The definition of variant 29
2.6.2 Denominative variants 30
2.6.3 Conceptual variants 30
2.7 Border between terms and variants 31

CHAPTER 3
Conceptualisation of terminological variants 33
3.1 Description of variants 34
3.1.1 Organisation of variants 34
3.1.2 Mechanisms and linguistic operations 36
3.1.3 Properties of variants 37
3.2 Denominative variants 37
3.2.1 Synonymic substitution 38
3.2.2 Simplification 40
3.2.3 Exemplification 43
3.2.4 Competing patterns 45
3.3 Conceptual variants 46
3.3.1 Expansion 46
3.3.2 Anaphorical reduction 51
3.4 Linguistic variants 52
3.4.1 Graphics and spelling 52
3.4.2 Inflection 55
3.4.3 Derivation 56
3.4.4 Fullback-compounding 56
3.4.5 Modification 57
3.4.6 Coordination, disjunction and enumeration 58
3.5 Variants of register 62
3.5.1 Variation of scientification/popularisation 62
3.5.2 Variants of position 63
3.6 Borders between categories of variants 63
3.6.1 Denominative and linguistic variants 64
3.6.2 Denominative and conceptual variants 64
3.6.3 Conceptual and linguistic variants 64

CHAPTER 4
Semantics of conceptual variants 67
4.1 Structuring terms 67
4.1.1 Conceptual and semantic relations 68
4.1.2 Classic semantic relations 68
4.1.3 Collocation 70
4.1.4 Lexical functions 70
4.2 Fundamental relations between term and variant 73
4.2.1 Synonymy 73
4.2.2 Hierarchical relations 74

4.3 Complex relations between term and variant 75
4.3.1 Result 76
4.3.2 Plurality 77
4.3.3 Spatiality 77
4.3.4 Temporality 78
4.3.5 Quality 78
4.4 Other relations between term and variant 79
4.4.1 Predication 79
4.4.2 Instance 79

Part II. Automatic discovery

CHAPTER 5
Primitive exploration of variants using comparable corpora 83
5.1 Comparable corpora 83
5.1.1 Corpus 83
5.1.2 Properties 84
5.1.3 Collecting comparable corpora 86
5.1.4 Comparability 87
5.2 Comparable corpora used in this study 87
5.3 Looking for variants 89
5.3.1 Implementation 89
5.3.2 N-gram massive data 90
5.3.3 Unigrams 91
5.3.4 Skip-grams 99
5.3.5 Categories of variants facing data 101
5.4 Comparison according to communication levels 104
5.4.1 Unigrams 104
5.4.2 Skip-grams 105

CHAPTER 6
Processing methods for the detection of variants from corpora 109
6.1 Linguistic-based methods 110
6.1.1 Morphological analysis 110
6.1.2 Syntagmatic and paradigmatic analysis 111
6.1.3 Syntactic analysis 112
6.1.4 Distributional analysis 113
6.2 Algorithms on strings 114
6.2.1 Distance computed from common substrings 115
6.2.2 Edit distances 117

6.3 Statistical methods **120**
6.4 Typology of variant occurrences **121**
6.4.1 Isolated variant occurrences **121**
6.4.2 Inter-mixed term and variant occurrences **122**
6.4.3 Separated term and variant occurrences **123**
6.5 Relationship between processing methods and types of occurrences **124**

CHAPTER 7
Grammar of variants **125**
7.1 Specifications and properties **125**
7.1.1 Expressivity of the syntagmatic rules **126**
7.1.2 Core operations **129**
7.1.3 Ambiguity of the syntactic analysis **130**
7.2 Generic grammar of recognition of variants **132**
7.2.1 Competing structures **133**
7.2.2 Augmented/reduced structures **134**
7.2.3 Contextual structures **135**
7.2.4 Function words **135**
7.2.5 Ad-hoc rules **135**
7.3 Variant grammars for specific languages **136**
7.4 Cross-lingual observations **138**
7.4.1 Coverage **138**
7.4.2 Precision **140**
7.5 Summary of observations **146**

CHAPTER 8
Synonymic variants **147**
8.1 Distributional analysis **147**
8.1.1 Modelling of the distributional methods **148**
8.1.2 Observations in specialised domains **151**
8.2 Compositional method **152**
8.3 Semi-compositional method **153**
8.4 Cross-lingual and cross-method observations **154**
8.4.1 Reference lists of synonyms **155**
8.4.2 Experimental setup parameters **157**
8.4.3 Evaluation measures **158**
8.4.4 Results **159**
8.5 Towards the detection of antonymic variants **165**

Part III. Applications and tools

CHAPTER 9
Terminology extraction 171
9.1 The core of terminology extraction 172
9.2 Collecting candidate terms 173
9.2.1 Patterns 173
9.2.2 Generic rules 173
9.2.3 Borders 176
9.2.4 Lexical expansion 176
9.3 Filtering and sorting candidate terms 177
9.3.1 Frequency 178
9.3.2 Association measures 178
9.3.3 Specificity measures 179
9.3.4 Filtering by removing nested terms 182
9.3.5 Contextual filtering 183
9.3.6 Supervised learning methods 185
9.4 Evaluation 186
9.4.1 References 187
9.4.2 Measures 189
9.5 Comparing term extraction without and with variant recognition 190
9.6 Experimental setting 190
9.6.1 Corpora 190
9.6.2 Our integrated terminology extraction 192
9.6.3 Comparison protocol 194
9.6.4 Maximum recall 194
9.6.5 Observations with a posteriori RTL 195
9.6.6 Observations with a priori RTL 199
9.7 Summary of observations 201

CHAPTER 10
End-user applications and tools 205
10.1 Machine-aided indexing and FASTR 205
10.2 Thematic cartography and TermWatch 206
10.3 TermSuite 208
10.3.1 Architecture 208
10.3.2 Token Regex 208
10.3.3 Compost 212
10.3.4 Variant grouping 216
10.3.5 Ranking by termhood 219
10.3.6 Performance 220
10.3.7 Release 221

Part IV. Conclusions

CHAPTER 11
Term variants and their discovery **225**
11.1 Summary of the present study **225**
11.1.1 A unified typology of term variants **225**
11.1.2 A variety of methods for the discovery of variants **226**
11.1.3 A terminology-resource building application **228**
11.2 Remaining issues and direction for further research **229**
11.2.1 Semantic analysis of variations **229**
11.2.2 Distributional analysis at the morpheme level **230**
11.2.3 Recognition of other variants **231**
11.3 Implications for related studies **231**
11.3.1 Variants and paraphrases **231**
11.3.2 Variants and translation **232**

Bibliography **235**

APPENDIX A
Notation **247**
A.1 Examples **247**
A.2 Specialised domains **247**
A.3 Specialised corpora **247**

APPENDIX B
Multext categories **249**

APPENDIX C
Search with AntConc **251**
C.1 Parameters **251**
C.2 Collection of n-grams **251**
C.3 Results of n-grams **253**

APPENDIX D
GGRV **257**
D.1 French **257**
D.2 English **260**
D.3 Spanish **263**
D.4 German **266**
D.5 Russian **268**

Index **271**