

UNIVERZITA KARLOVA V PRAZE
FILOZOFICKÁ FAKULTA
ÚSTAV TEORETICKÉ A KOMPUTAČNÍ LINGVISTIKY

Czech syntactic lexicon

Hana Skoumalová

2001

filologie — matematická lingvistika
vedoucí práce — prof. PhDr. Jarmila Panevová, DrSc.

Acknowledgments

I would like to thank Prof. Jarmila Panevová, for her tutoring. She introduced me to the valency theory, she led me in my work, and she pointed out work to me that has been done by generations of linguists when I tried to reinvent the wheel.

I also want to thank my colleagues at work, Niki Petkevič, Saša Rosen and Milena Hnátková who were exposed to my crazy ideas and who helped me to judge some strange sentences and syntactic constructions.

My gratitude belongs to my parents—my mother Zdena Skoumalová and my late father Jan Skoumal—who trained me in linguistics since my childhood. It was my father who showed me that linguistics is fun, and my mother who showed me that linguistics is serious work. In spite of their dissuasion, I continue the family tradition and I hope that I will not thereby harm it.

My thanks go also to Karel Oliva who read a part of the dissertation and gave me several useful remarks, and Clare Britton who corrected my English.

Last but not least, I must thank the Research Support Scheme, which supported my work with grant no. 72/94, the Grant Agency of the Czech Republic, which supported the work by grant no. 405/96/K214, and the Research and Development Council of the CR Government, which supported the work by Research program MSM 112100002.

Contents

Acknowledgments	ii
1. Introduction	1
1.1. Terminological remarks	2
2. Theoretical background	3
2.1. An overview of FGD	3
2.2. Comparing FGD with other theories	6
2.2.1. Government-Binding Theory	6
2.2.2. Lexical-Functional Grammar	7
2.2.3. Head-Driven Phrase Structure Grammar	7
2.2.4. Comparison with FGD	9
3. Using existing sources	10
3.1. Source data	11
3.1.1. The attributes used in the lexicon and their values	11
4. Content of the lexicon	14
4.1. Format of a lexical entry	14
4.1.1. Voice	15
4.1.2. Reflexivity	16
4.1.3. Subject	16
4.1.4. Functor	17
4.1.5. Grammatemes	17
4.1.6. Diatheses	18
4.2. Reflexivity	21

4.2.1.	True reflexive with <i>se</i>	21
4.2.2.	True reflexive with <i>si</i>	23
4.2.3.	Reciprocal verbs with <i>se</i>	23
4.2.4.	Reciprocal verbs with <i>si</i>	27
4.2.5.	Reflexive tantum with <i>se</i>	28
4.2.6.	Derived reflexive verbs with <i>se</i>	28
4.2.7.	Reflexive tantum with <i>si</i>	28
4.2.8.	Derived reflexive verbs with <i>si</i>	29
4.2.9.	Reflexive with optional <i>se</i>	29
4.2.10.	Reflexive with optional <i>si</i>	30
4.2.11.	Reflexive passive	31
4.2.12.	Mediopassive	31
4.2.13.	Homonymy of reflexive verbs	31
4.3.	Diatheses	33
4.3.1.	Diatheses encoded in the lexicon	40
4.3.2.	Periphrastic passive	41
4.3.3.	Reflexive passive	44
4.3.4.	Mediopassive	46
4.3.5.	Constructions with <i>mít</i> and <i>dostat</i>	47
4.3.6.	Resultative construction with <i>mít</i>	49
4.4.	Verbs with the infinitive in their frames	49
4.4.1.	Raising verbs	55
4.4.2.	Equi verbs	59
5.	Algorithm for processing the surface frames	66
5.1.	Identifying and merging frames, marking the obligatoriness	66
5.2.	Assigning functors	68
5.3.	Marking diatheses	73
5.4.	Usage of the final lexicon	73
5.4.1.	Generating frame instances from frames	74
5.4.2.	Extracting subcat lists	76
6.	Conclusions	78

6.1. Verb grouping	78
6.2. Further perspectives	80
Bibliography	81
Subject index	86
Verbs used in examples	88
A. Abbreviations	90
B. Symbols used in the dictionary	92
B.1. Voice	92
B.2. Reflexivity	92
B.3. Subject	93
B.4. Functors	93
B.5. Grammatemes	94
B.6. Obligatoriness	96
B.7. Passive and other diathesis	96
C. Possible functors assigned to grammatemes	97
C.1. Abbreviations used in lists of possible functors	97
C.2. Lists of functors attached to every surface realization	98
D. Algorithm for assigning functors	102
D.1. Prototypical and less typical surface forms	102
D.2. Assigning non-prototypical frame	103
D.3. Results	103
D.3.1. Verbs processed fully automatically	103
D.3.2. Verbs with ambiguous frames	108
E. Classification of Czech frames	115
E.1. Automatically processed frames	115
E.2. Ambiguous frames	116
F. Experiment with LFG	121

F.1. Verb lexicon	121
F.2. Templates	122
F.3. Lexical rules	123
F.4. Grammar	125
F.5. Test sentences	126
G. Web interface to the lexicon	132

List of Tables

4.1. Taxonomy of reflexive verbs	21
4.2. Three types of reciprocal verbs	24
4.3. Reciprocal verbs with <i>si</i>	27
4.4. Subject diatheses	39
4.5. Subject diatheses revisited	40
5.1. Identifying single frames	67
5.2. Merging frame variants	67
5.3. Prototypical frames	70
5.4. Non-prototypical frames	70
5.5. Merging frame of the verb <i>čertit se</i> (be angry)	71
6.1. Classification of verbs	78
6.2. Classification of verbs with adjuncts simplified	79

List of Figures

- 4.1. Three-level system 36
- 4.2. Three-level system revisited 37

- 5.1. Mapping between TL and ML in active voice 69
- 5.2. Mapping between TL and ML for verbs with at least three actants 69

- D.1. The algorithm for assigning functors to non-prototypical frame 104

- F.1. Simple grammar in LFG 125
- F.2. Testing sentences 126
- F.3. C-structure of sentence 140a 127
- F.4. F-structure of sentence 140a 127
- F.5. C-structure of sentence 140b 128
- F.6. F-structure of sentence 140b 128
- F.7. C-structure of sentence 140c 129
- F.8. F-structure of sentence 140c 129
- F.9. C-structure of sentence 140d 130
- F.10. F-structure of sentence 140d 131

- G.1. Main window of the web interface 133
- G.2. File with all frames containing hPTc2 134
- G.3. Frames processed fully automatically, with ambiguous free modifications 135