

1	Introduction	1
1.1	The Importance of Collocations	1
1.2	Motivation	2
1.3	Objectives	2
I	Theoretical Background	5
2	Key terms	7
2.1	Words, Lexemes, and Lexical Items	7
2.2	Collocation	7
3	Grammaticalization	11
3.1	The Notion of Grammaticalization	11
3.2	Emergent Grammar	12
3.3	The Most Grammaticalized Verb Categories	14
3.3.1	Voice and Tense	15
3.3.2	Mood	15
3.3.3	Number and Person	15
3.3.4	Valency	15
3.3.5	Aspect	16
3.4	Discovering Regularities in Verb Usage	17
3.5	Semantic Changes in Grammaticalizing Verbs	18
3.5.1	Metaphorical Extension from one Semantic Domain to Another	19
3.5.2	Context-induced Reinterpretation	20
4	Light Verb Constructions	23
4.1	The Notion of Light Verb Construction (LVC)	23

4.2	LVCs as Collocations	25
4.3	Semantic Aspects of LVCs	26
4.4	LVCs and Event Structure	27
4.5	Productivity vs. Lexicalization in LVCs	28
4.6	Aspects of Valency in LVCs	28
4.7	Communicative Aspects of LVCs	33
4.8	Conclusions	35
5	The Transitivity Hypothesis	37
5.1	What is Transitivity and (why) does it Matter?	37
5.2	Grammatical Interference in Lexicalized Collocations?	41
5.3	Transitivity Indicators	45
5.3.1	Participants	46
5.3.2	Kinesis	46
5.3.3	Aspect	47
5.3.4	Punctuality	47
5.3.5	Volitionality	47
5.3.6	Affirmation	47
5.3.7	Mode	47
5.3.8	Agency	47
5.3.9	Affectedness of the Patient	48
5.3.10	Individuation of the Patient	48
5.4	Transitivity as a Discourse Marker	48
5.5	Morphosyntactic Consequences of Transitivity Hypothesis	49
5.6	Relation between Aspect and Definiteness	51
5.7	Transitivity Hypothesis in Light Verb Constructions	52
II	Methods and Approaches	55
6	Valency Theory in Functional Generative Description	57
6.1	Functional Generative Description	57
6.2	Tectogrammatical Representation	57
6.3	Valency	59
6.3.1	Basic Notions	59

6.3.2	Inner Participants and Free Modifications	60
6.3.3	Obligatoriness and Dialog Test	63
6.3.4	Shifting	64
6.3.5	Quasi-Valency Complements	65
6.3.6	Discussion	66
6.3.7	Noun Valency in FGD	68
6.3.8	FGD-valency for Learners of Basic Verbs	69
7	Corpus Pattern Analysis	75
7.1	Theory of Norms and Exploitations (TNE)	75
7.2	Defining Lexical Sets	75
7.3	Applying CPA	78
8	Lexical Functions	79
8.1	Brief Overview	79
8.2	Lexical Functions and FGD	79
8.3	Basic Lexical Functions in LVC Description	80
8.3.1	Oper _{<i>i</i>}	80
8.3.2	Labor _{<i>i,j</i>}	81
8.3.3	Func _{<i>i</i>}	82
8.3.4	Copul	82
8.3.5	Cross-linguistic Comparison of the Predicate Noun <i>disposal</i> in Swedish vs. in Czech	83
8.4	Phasal Lexical Functions	84
8.5	Causative Lexical Functions	84
9	Examples of Valency Lexicons and Collocational Lexicons	85
9.1	BBI	85
9.2	Combinatorial Explanatory Dictionaries	87
9.2.1	Definitions in CEDs	88
9.2.2	Government Pattern	88
9.2.3	Lexical Combinatorics	89
9.2.4	Monolingual and Multilingual CEDs	89
9.3	Dictionary of Czech Phraseology and Idiomatics	89
9.4	PropBank, NomBank	90

9.5	VALLEX, PDT-VALLEX	92
9.5.1	Differences between VALLEX and PDT-VALLEX	92
9.5.2	Structure of VALLEX	92
9.5.3	Other Language Versions	94
9.6	FrameNet	95
9.7	Svenskt Språkbruk	95
9.8	PAROLE-SIMPLE	96
9.9	VerbaLex	96
III Grammar and Corpus - the Case of Swedish		99
10The Role of Verbs in the Vocabulary		101
10.1	Verb Statistics	101
10.2	The 20 Most Frequent Verbs in Swedish	102
11Komma att		107
11.1	Future	107
11.2	Accident, Coincidence	111
11.3	A Contrastive Corpus Analysis of the Non-Future Uses	112
11.3.1	Significant Collocates	112
11.3.2	Coincidence in Czech	113
11.3.3	<i>Kom att</i> in Result Clauses	119
11.3.4	Czech Equivalents of <i>komma att tänka</i>	123
12Hålla på ...		127
12.1	Valency Patterns	127
12.2	<i>X håller på med Y</i>	127
12.3	<i>X håller på (med) att verb-a</i>	128
12.4	Progressivity	129
12.5	Progressivity Hides Constancy	129
12.6	Tendentiality	133
13Pseudo-coordinations with <i>ligga, sitta, stå</i>		135
13.1	The Profile of <i>X står och verb-ar</i> in PAROLE	137
13.2	The Profile of <i>X sitter och verb-ar</i> in PAROLE	142

13.3	The Profile of <i>X ligger och verb-ar</i> in PAROLE	142
13.4	Deictic Markers in Pseudo-coordinations	144
14	Pseudo-coordinations with <i>ta</i>	147
14.1	The Profile of <i>X tar och verb-ar</i> in PAROLE	148
14.2	Czech Equivalents	149
IV	Implementation of SWE-VALLEX/PNL	151
15	Preparatory Work	153
15.1	Organizing Lexical Sources and Tools	153
15.2	Preparing the Corpora	153
15.3	Word Sketch Engine and Collocation Analysis	154
15.4	Adjusting the Word Sketch Engine for PAROLE	154
15.4.1	Lemmatization	155
15.4.2	Lemmatization Results Related to Generating Word Sketches	159
15.4.3	Computing Grammatical Relations	160
15.4.4	Empirical Evaluation of Word Sketches	162
15.5	Determining Entry Candidates	163
16	Data Structure	169
16.1	Usefulness of Word Sketches	169
16.2	Main Principles and Features	169
16.3	SWE-VALLEX	172
16.3.1	Macrostructure	172
16.3.2	Lemma	173
16.3.3	Patterns	173
16.3.4	Slot	174
16.3.5	Surface Form	175
16.3.6	FGD-Information	175
16.3.7	CPA-Information	176
16.4	Predicate Noun Lexicon	176
16.4.1	Macrostructure	176
16.4.2	Predicate Noun Lemma	177

16.4.3	Light Verb Unit	177
16.4.4	Noun Definiteness, Modifier Insertion	178
16.4.5	Slot	179
16.5	Linking	180
17	Discussion	185
17.1	Increasing Recall with Targeted Corpus Queries	185
17.2	Frequency Counts	189
17.3	Irrealis, Negation, and Semantic Definiteness	190
17.4	Undetected Information	190
17.5	Different GUI – New Corpus Annotation	192
17.6	Parallel Data	192
18	Conclusion	195
	Appendices	201
A	Data sample	201
B	A Sample of the Lemmatizer Script	215
C	Targeted Corpus Query Templates for PAROLE	219
	Summary	223
	Bibliography	225
	Index	237