WORDS THAT MATTER
Towards a Swedish-Czech Colligational Lexicon of Basic Verbs

Silvie Cinková

ÚSTAV FORMÁLNÍ
A APLIKOVANÉ LINGVISTIKY

STUDIES IN COMPUTATIONAL AND THEORETICAL LINGUISTICS

Silvie Cinková

WORDS THAT MATTER
Towards a Swedish-Czech Colligational Lexicon of Basic Verbs

Published by Institute of Formal and Applied Linguistics
as the 2nd publication in the series
Studies in Computational and Theoretical Linguistics.

This book has been printed with the support of the project MSM0021620838 of The Ministry of Education of the Czech Republic.

ISBN 978-80-904175-3-3

Acknowledgements

Any merit that this work may have is dedicated to Karel Oliva (1927–2005), a great Czech linguist and the author of the largest Polish-Czech dictionary. He was a dear friend and mentor to many young people, myself included.

I would also like to express my gratitude to:

Anna Braasch, Kristín Bjarnadóttir, Ulrich Heid, Ann Lindvall, Markéta Lopatková, and Zdeněk Starý, who let me acquire up-to-date publications and who spent hours of their valuable time to share their knowledge with me,

Petr Pajas, Pavel Rychlý, and Zdeněk Žabokrtský for advising me on all technical matters and granting me access to relevant tools,

Jan Pomikálek, without whom I would never have been able to make my lemmatising rules comprehensible to any computer,

Pavel Straňák and Ondřej Bojar, who helped me with typographical issues,

Jan Hajič and Eva Hajičová, who gave me the opportunity and the joy of inspiring work at UFAL,

František Čermák and Patrick Hanks for reading my manuscript and adding valuable comments at various stages,

Vladimír Petkevič for becoming my supervisor, for leading and encouraging me through all the years,

Petra Tesařová, Michal Zikán, and Laura Janáčková for putting me back in the game

…and especially to my family, who have always kept me going and whom I owe so much.

This work was funded in part by the Companions project (`www.companions-project.org`) sponsored by the European Commission as part of the Information Society Technologies (IST) programme under EC grant number IST-FP6-034434, by the Czech Science Foundation (GA-405/06/0589), and by the Czech Ministry of Education (0021620823).

Contents

1
Introduction

Structure, or regularity, comes out of discourse and is shaped by discourse as much as it shapes discourse in an on-going process.

Paul Hopper: Emergent Grammar

1.1 The Importance of Collocations

Every human language has a grey area where grammar and lexicon overlap. It resists attempts at systematic description in grammar textbooks as well as in lexicons for foreigners; yet understanding the nature of this grey area is crucial, not only for effective language learning, but also for many other applications. It is necessary to seek language usage that is not only correct but also idiomatic. To do this, it is necessary to understand collocations.

In language production, collocational preferences in semantically transparent constructions rather than idioms pose problems for foreign speakers, [150]. As Sinclair notes, foreign speakers quite often abuse dictionaries to look up and combine rare and outdated words to express ordinary, everyday concepts instead of using common collocational blocks familiar to native speakers. English phrasal verbs provide nice examples of this issue: a foreign learner often gets lost in the jungle of these utterly unintelligible combinations of highly polysemous verbs with one or two preposition-like attachments. Therefore, a foreign learner often prefers e.g. *vomit* to *throw up* and *reconcile* to *make up*, although a native speaker would only use *vomit* or *reconcile* in formal contexts.

Textbooks and dictionaries for foreigners have been randomly treating diffuse uses of common lexemes either as grammar issues in text-books, or as 'idioms' and 'figurative senses' in lexicons, while many regularities have simply remained unnoticed by native speakers. Corpus linguistics has shown through analysis of vast amounts of data that the way in which native speakers believe words behave is often different from the way in which words actually behave. Manual excerpts of lexical evidence for dictionaries and grammars tend to highlight the unusual in the language, while missing important patterns (see [56]).

The motto of the following work is *grammaticalization* , understood as "movement towards structure" [65]. As Hopper puts it: "…the more useful a construction is, the more it will tend to become structuralized, in the sense of achieving cross-textual consistency, and serving as a basis for variation and extension." For the purpose of this study, 'grammaticalization' is understood in its broadest sense. It is basically used

interchangeably with 'lexicalization'. Structures that can combine with virtually any collocates (very generalized structures) like the English *come to + INF* (*come to know* etc.) will rather be called 'grammaticalized', while structures with stricter collocational restrictions will rather be called 'lexicalized'. Nevertheless, this study makes no attempt to draw a line between 'grammaticalization' and 'lexicalization'[1].

1.2 Motivation

The initial impulse for this work came from a previous experience with lexicographical processing of German lexical verbs. German, being famous for its abundant nominalizations, makes prominent use of *Funktionsverbgefüge* (light verb constructions), a way of integrating a nominalization into a sentence by means of a semantically-depleted lexical verb, such as *zur Verfügung stellen*, *eine Frage stellen* etc. The ultimate German grammar for foreign learners [62] even introduces a group of *Funktionsverben* (light verbs).[2] Examining Swedish light verbs seemed to be an interesting task. However, currently available Swedish grammars [153], [113] do not provide any explicit list of light verbs, and light verb constructions are only captured by the lexicon [2], sometimes under noun, sometimes under verb lemmas. The initial attempts to identify a generally accepted list of Swedish light verbs resulted in the insight that verbs, rather then *being* light verbs, *become* light verbs by being used in collocation with a predication-containing noun. Many lexical verbs belonging to the basic vocabulary have a shifting potential for occurring in collocation with predication-containing nouns. Some enter such constructions frequently and productively, while others only occur in one or few lexicalized cases, such as *bjuda* in *bjuda motstånd*. Only few lexical verbs occur almost exclusively in light verb constructions (i.e. *genomföra*). Most verbs with this ability have quite concrete meanings, i.e. *komma, ställa, hålla*, or *få*. Hence the field of inquiry expanded to semantic and syntactic changes in such verbs when they occur in combinations that reach beyond their most concrete meaning.

1.3 Objectives

Originally, this study sought only to explore ways of creating a systematic description of light verb constructions in a lexicon. However, the usage of a verb as a light verb is quite often only one instance of its grammaticalized uses. It is typically the basic spatial verbs and verbs denoting physical actions that are the most productive light verbs, and they typically exhibit a significant semantic complexity.

[1] Cf. Section 3.2.

[2] The English term *light verb* used in the present study was coined by Otto Jespersen [73]. A more accurate German equivalent for this term is **verblasste Verben**. Hanks et al. [59] argue that the semantic criterion of 'lightness' is preferable, as an identifier for these interesting verbs, to the syntactic criterion implied by *Funktionsverbgefüge*.

With the human lexicon user in mind, a thorough description of all types of grammaticalized uses with some common verbs seemed more sensible than focusing strictly on light-verb-like uses and defining Swedish light verbs as a group.

An experimental lexicon has been created, called **SWE-VALLEX/PNL**. The first part of the name refers to VALLEX, valency lexicon of Czech verbs [90] based on the valency theory of the Functional Generative Description (see Chapter 6). SWE-VALLEX is meant to be a Swedish counterpart of VALLEX. The abbreviation *PNL* stands for *Predicate Noun Lexicon*. The lexicon contains two lemma types: verb lemmas and noun lemmas. The noun entries present significant collocates of the verbs included, primarily predicate nouns in light verb constructions. Their types will be specified in more detail (see Section 16.4). The respective entry types are stored in two separate XML documents (SWE-VALLEX and PNL), which are interlinked.

To become eligible for lemmatizing in SWE-VALLEX/PNL, a verb must have been found in a large corpus to participate in at least one light verb construction. The selection of lemma candidates is described in Section 15.5.

These are the main features of the proposed lexicon:

- The lexicon is theory-bound and formalized in a way that makes it eligible for NLP tasks, e.g. as a valency lexicon for an FGD-based Swedish treebank.
- When enhanced with a browser, the lexicon is a user-friendly electronic lexical resource for human users.
- The description focuses on grammaticalizing verb-noun constructions. To increase the informative value of the lexicon, neither 'literal senses' nor idioms have been ignored, but they might be incomplete. This applies especially to idioms, evaluative expressions and proverbs. For a human user, the lexicon is by no means a substitute for large monolingual dictionaries, e.g. Svenskt språkbruk and Norstedts stora svenska ordbok, but it seeks to complement them by stressing the morphosyntactic information necessary for language production.
- The source language is Swedish, and the target language is Czech. Emphasis has been laid on the Swedish part. The Swedish part contains information on spelling, valency and morphosyntactic features, as well as numerous authentic examples in order to provide all knowledge relevant for improving the user's foreign language production. The Czech part comprises only equivalents. No further description of the Czech equivalents is provided.
- The data structure makes it possible to view the entries sorted according to the Czech equivalents. It can thus be used as a – very simplified – Czech-Swedish lexicon of verbs.
- The lexicon is corpus-driven in the sense of Hanks [56], but the results of the corpus-driven analysis are compared to large Swedish monolingual dictionaries (cf. Section 15.4.4).

One more word should be said about the decision to make the lexicon both human- and machine-oriented: the examples of the Czech VALLEX, PropBank and many other resources show that a user-friendly interface makes even a very formalized data

structure easily understood. Information overload with human users is not to be feared. The final layout of any lexicon can be adapted to any target user's needs, provided the relevant information is present and organized consistently enough to be retrieved – which has been the aim of this project from the very beginning, in full agreement with Bolshakov, Gelbukh and Haro [11]: "…the information should be accessible to both human users and other programs. A dictionary is so large piece of data, and its development is so expensive, that it is unacceptable to maintain, keep and use separate versions of dictionaries for humans and for the machine".

8
Lexical Functions

8.1 Brief Overview

The lexicon proposed by this study was significantly inspired by already existing collocational dictionaries that have paid systematic attention to LVCs, namely *The BBI Combinatory Dictionary of English Word Combinations*. [109], as well as *Tolkovo-kombinatornyj slovar sovremennovo russkovo jazyka* [68] and *Dictionnaire explicatif et combinatoire du français contemporain* [100], out of which the latter two are modeling 'institutionalized' lexical relations by the so-called Lexical Functions. A brief overview of these dictionaries has been given in Section 9.

Lexical Functions are part of the Meaning-Text-Theory developed by Igor Mel'čuk and his collaborators [99], [76]. There are two elementary types of LFs – paradigmatic and syntagmatic – and this study concerns only the latter. In terms of collocations, when two lexical units are collocates, one is usually the base that "selects" the other lexical unit to render a certain meaning together. MTT aims to capture it by the mathematical functional notation: LFi(X) = Y, where X is the keyword (the collocational base) and Y the value of the LFi (the collocate). LFs can assign one value or a set of values to a given keyword. The values stand in the same lexical relation to the keyword but they are not necessarily synonymous. The LFs describe the semantic relation between the keyword and the values. For further reference see [164].

As the data of PNL presents numerous instances of LF-combinations, most illustrative examples in this chapter will be English and directly taken from [164].

8.2 Lexical Functions and FGD

Lexical Functions are closely associated with valency. For light verbs and predicate nouns it is crucial to determine the participants in the light-verb frame as well as in the event that is denoted by the predicate noun. Often, but not necessarily, there is a corresponding lexical verb from which the noun had been derived (or vice versa), e.g. *ledning*: *leda*, *fråga*: *fråga*, etc., and it just inherits the frame of the verb. Othervise, the nominal event would have to be paraphrased with a verb that has no word-formation relation to the given noun. Mel'čuk calls the nominal event in a LVC "the underlying expression," which specifies the corresponding situation 'L': a full verbal form meaning 'L' ([101], p. 61).

LFs Oper and Labor use numerical indexes to indicate which of the participants of the nominal event ("the underlying expression") is identical with the subject of the

light verb. Meaning – Text Theory refers to participants as DSyntAs and labels them with numbers. As we are working within the FGD framework, we will have to relate what LFs use from MTT's valency theory to the valency theory of FGD.

In terms of FGD, the entire LVC would be drawn as a light-verb deep valency frame, in which the predicate noun is the deep frame slot with the deep frame slot filler CPHR (or sometimes DPHR). The predicate noun itself is a deep frame evoker. It has again its own deep frame slots whose fillers denote the cognitive roles of the participants. Now, one of the deep frame slots belonging to the predicate noun is in grammatical coreference with the Actor of the light verb. Its surface frame slot filler will get the substitutional t-lemma QCor (quasi-coreference).

The numerical index of the LF refers to the number of the particular deep frame slot of the predicate noun whose surface frame slot filler has got the substitutional t-lemma QCor. Deep frame slot fillers, unlike DSyntAs in MTT, are not labeled by numbers but they are assigned functors. Presumably, only functors of inner participants and no functors of free modifications will participate in describing LVCs. To be fully just to FGD, the LFs would actually have to carry functor names instead of numerical indexes when used within the FGD framework. However, the functors of inner participants can be easily displayed as numbers: ACT corresponds to 1 and PAT corresponds to 2. ADDR, ORIG and EFF correspond to 3, 4 and 5 respectively. As for the principle of shifting, any ADDR, ORIG or EFF 'shifted' to PAT is regarded as PAT and is to be assigned the label 2.

8.3 Basic Lexical Functions in LVC Description

Lexical Functions are, according to Mel'čuk ([101], p. 60), characterized by the following relation between the verb and the noun: "The support verbs serve to link, on the DSynt-level, (the name of) a DSyntA of L to L itself; they thus play an important semantico-syntactic role and can be loosely called 'semi-auxiliaries' ". The verbs "play an important communicative role: they are used to express the communicative organization, or perspective, of the sentence. […] Therefore, although they are semantically empty […], they are by no means asemantic. In addition, they carry all grammatical verbal caegories which must be expressed in a sentence (mood, tense, person and number, etc.)."

The following LFs are specific to LVCs; their keywords are the predicate nouns and their values are by definition verbs: Oper_i, $\text{Labor}_{i,j}$ and Func_i. For the purpose of PNL, Copul was added to be observed as well, though it does not belong to the LVC-describing Lexical Functions.

8.3.1 Oper_i

The Actor (and subject) of the light verb stands in grammatical coreference with the *i*-th inner participant of L, i.e. of the predicate noun in question. The predicate noun is

direct object or indirect object (if the verb cannot have a direct object) or the strongest prepositional object of the light verb (if the verb cannot take non-prepositional objects). In FGD, it gets the functor CPHR, which in this particular constellation substitutes for the Patient, which is nr. 2 in the hiearchy of FGD-inner participants. This corresponds to the original definition of MTT, saying that "The DSyntA I of this verb (and its SSynt-Subject) is the phrase that is described in the GP [*government pattern,* S.C.] of L as the *i*-th DSyntA of L, and Oper_i's DSyntA II (= its main SSynt-Object) is L itself."

Typically, Oper_i comes realized as Oper_1 or Oper_2. Constructions like *to give an order*, *to lend support* make good Oper_1 examples. Having a closer look at *to give an order* from the perspective of FGD, the construcion has the following features:

- The construction can be paraphrased as *X gives Y X's order to do z*. The Actor (at the same time subject on a-level) of the verb corefers with the Actor (at the same time subject on the a-level) of the event described by the noun *order*. (Let us assume the following deep valency frame in *order*: X's.ACT *order* to/for Y.ADDR to do z.PAT.
- The noun *order* is direct object of the verb *to give*. It has the functor CPHR – which has substituted PAT, the functor the noun would have got if it hadn't been a predicate noun in an LVC.

Oper_2 differs from Oper_1 in that the Actor (and subject on the a-level) of the verb corefers with the Patient (and object on the a-level) of the event denoted by the noun, e.g. in *to come under X's control*. Let us assume the following deep valency frame in the noun *control*: X's.ACT *control* over Y.PAT. The Actor (and subject on the a-level) of the verb *to come* corefers with the Patient (and object on the a-level) of the event denoted by the noun *control*. The construction can be paraphrased as *Y comes under X's control over Y*.

Though Mel'čuk [101] only mentions indexes 1 and 2 in connection with the Oper-LF, PNL introduces also Oper_3. Oper_3 is used for LVCs whose predicate noun's deep valency frame consists of ACT, ADDR and PAT. This would be e.g. the case of *to receive an order*. *Order* will have the following deep valency frame: X's.ACT order to Y.ADDR to do z.PAT. The entire LVC *to receive an order* could be paraphrased in this way: Y receives X's order to Y to do Z. The Actor (and subject on the a-level) of the verb *to receive* corefers with the Addressee (and object on the a-level) of the event denoted by the noun *order*.

So far, neither Oper_4 nor Oper_5 have occurred in PNL, but they are formally possible.

8.3.2 $\text{Labor}_{i,j}$

In Labor, the predicate noun is the prepositional object of a transitive light verb. The Actor (and subject on the a-level) of the light verb stands in grammatical coreference with the *i*-th participant of the event L denoted by the predicate noun. The direct

object of the light verb stands in grammatical coreference with the *j*-th participant of the event L denoted by the predicate noun. This FGD-adopted definition originates in [101] saying that with $\text{Labor}_{i,j}$ "The DSyntA I f this verb (and its SSynt-Subject) is the *i*-th DSyntA of L, its DSyntA II (= its main SSynt-Object) is the *j*-th DSyntA of L, and its further DSyntA (= its second or third SSynt-Object) is L itself."

If *interrogation* combines the verb *to subject* as predicate noun, they belong to the $\text{Labor}_{1,2}$ LF as keyword and value. *Interrogation* is the prepositional object of *to subject*. In FGD, the light verb frame of the verb to subject would consist of ACT, PAT and CPHR. The event L denoted by the noun *interrogation* would have the following participants: X's.ACT interrogation of Y.PAT. The LF $\text{Labor}_{1,2}$ can be paraphrased as *X subjected Y to X's interrogation of Y*. The Actor (and subject on the a-level) of the light verb *to subject* stands in grammatical coreference with the Actor (and subject on the a-level) of the event L denoted by the predicate noun *interrogation* and, at the same time, the Patient (and direct object on the a-level) of the light verb *to subject* stands in grammatical coreference with the Patient of the event L denoted by the predicate noun *interrogation*.

8.3.3 Func_i

With the LF Func_i the predicate noun is the Actor (and subject on the a-level) of the light verb it joins in a LVC. The index refers to the number of the participant in the deep valency frame of the event L that is denoted by the predicate noun. in MTT's terms, "The DSyntA I of this verb (and its SSynt-Subject) is L itself, and its DSyntA II (= its main SSynt-Object) is the *i*-th DSyntA of L" ([101], p. 61), e.g. the event L denoted by the noun *blow* would have two participants, possibly an Actor (X's blow, blow from X) and a Patient (blow to Y). e.g. the sentence *The blow comes from X.* could be paraphrased as X's.ACT blow to Y.PAT comes from X.DIR1. It is the Actor of the event L denoted by the predicate noun *blow* that corefers with a member of the light verb's frame. Therefore the relation will get the index 1. In *The blow falls upon Y* it is the Patient of the event L denoted by the predicate noun *blow* that corefers with a member of the light verb's frame, and therefore this relation will be classified as Func_2.

The MTT's convention, which was fully adopted by PNL, says that the Func-index is 0 when the predicate noun joins an intransitive verb, as in Func_0(*the war*) =*is on*, or an intransitive LU of a verb, when a verb has intransitive as well as transitive LUs.

Unlike MTT, FGD does not regard the relation between a verb and a noun collocate as a LVC and the keyword of the Lexical Function Func keeps the functor ACT, not replacing it by CPHR.

8.3.4 Copul

The LF Copul captures verbs that acquire the meaning of a copula. In the Swedish data it is typically constructions as *ligga sjuk, sitta modell, gå vakt* etc. Though MTT

does not count such constructions as LVCs, and neither does FGD mark the nouns with the functor CPHR, PNL observes them together with LVCs, loosely associating them to LVCs.

8.3.5 Cross-linguistic Comparison of the Predicate Noun *disposal* in Swedish vs. in Czech

1. *ställa något till förfogande = dát něco k dispozici*
   give something at someone's disposal
2. *få/ha något till förfogande = dostat-mít něco k dispozici*
   get/have something at one's disposal
3. *något står till förfogande = něco je k dispozici*
   something is at someone's disposal

In both languages the deep valency frame of the noun is identical:

- *Y's.ACT förfogande över z.PAT = Y-ova.ACT dispozice se zPAT*

The first sentence is to be (language-specifically) paraphrased as follows:

- *X ställer z till Ys förfogande över z för Y.*[1]
- *X dává Y-ovi z k Y-ově dispozici se z.*

In both languages this LVC would be classified as CausLabor$_{2,3}$ (more about Caus see below, Section 8.5). The sentences can be paraphrased as *X causes that Y has z at Y's disposal*.

The second sentence would be classified as Labor$_{1,2}$. The sentences can be paraphrased as

- *Y.ACTfår/har z.PAT till Y's.ACT förfogande över z.PAT*
- Y.*ACT dostane/má z.PAT k Y-ově.ACT dispozici se z.PAT.*

The respective verbal Actors stand in grammatical coreference with the respective nominal Actors and the respective verbal Patients stand in grammatical coreference with the respective nominal Patients.[2]

[1] *för Y* is part of the verbal deep valency frame and is to be assigned the functor ADDR. It corresponds to Czech dative object.

[2] Note that Swedish does not allow for ADDR as direct object (or at least strongly prefers other forms of ADDR to direct object) in this LVC (no occurrences in PAROLE), which makes the coreferential relations quite difficult to resolve. The entire sentence gets the regular tectogrammatical representation if the Addressee is realized on the surface of the verbal frame, e.g. in the sentence *I en farmartjänst ställer de sina maskiner och kunskap til förfogande för industrier och kommuner för exempelvis grönyteskötsel*. Then the phrase *för industrier och kommuner* actually stands in grammatical coreference with the Actor of the predicate noun *förfogande*, i.e. the Actor of the predicate noun *förfogande* will never be realized on the surface together with the verbal ADDR, and it will obtain the substitutional tectogrammatical lemma QCor. A non-standard situation occurs when the nominal ACT is realized on the surface, e.g. in the sentence *I stället ställer Nato staber och militära resurser till Västeuropeiska unionen VEU:s förfogande vid kriser – operationer…* Then the verb is not allowed to realize the ADDR-participant on the surface, and the verbal ADDR gets the substitutional t-lemma QCor. The coreferential arrow pointing up in the tectogrammatical tree structure is hypothetically possible but has not occurred during the annotation of PDT, which implies that the issue has not been sufficiently explored yet.

The third sentence is to be classified as Oper_2 LF. The sentences can be paraphrased as

- *z.ACT står till Ys.ACT förfogande över z.PAT*
- *z.ACT je Y-ovi.ADDR k Y-ově dispozici se z.PAT.*

The Actors of the respective light verbs stand in grammatical coreference with the Patients of the respective predicate nouns.

8.4 Phasal Lexical Functions

There are three supplementary Lexical Functions that can specify the basic Lexical Functions Oper, Labor, Func and Copul:

- Incep (for inchoative/inceptive events)
- Cont (for continuing events)
- Fin (for finishing/ceasing events).

Examples:

1. *to open fire on X* = IncepOper_1(*fire*)
2. *to fall under the power of X* = IncepOper_2(*power*)
3. *to lose one's power over X* = FinOper_1(*power*)
4. *to get out of X's control* = FinOper_2(*control*)
5. *to retain one's power over X* = ContOper_1(*power*)

8.5 Causative Lexical Functions

Like phasal Lexical Functions, causative Lexical Functions specify basic Lexical Functions by observing additional semantic features that have to do with causativity. The causative Lexical Functions are:

1. Caus (for causative verbs)
2. Liqu (for verbs with the meaning of stopping an event)
3. Perm (for verbs with the meaning of permitting an event)

Examples:

1. *to lead X to an/the opinion* = CausOper_1*(opinion)*
2. *to stop aggression* = LiquFunc_0(*aggression*)
3. *to condone aggression* = PermFunc_0(*agression*)

They can even be used as basic LFs with verbs as keywords, even in Swedish: Caus(*falla*) = *fälla*, Caus(*sitta*)= *sätta*, etc.

Bibliography

[1] Sture Allén. *Tiotusen i topp*. Almqvist & Wiksell, Stockholm, 1972. 10.1

[2] Sture Allén et al. *Norstedts stora svenska ordbok*. Norstedts, 1995. 1.2, 15.4.4, 17.1

[3] Sue Atkins. Tools for Computer-Aided Lexicography: The Hector Project. *Acta Linguistics Hungarica*, 41:5–72, 1993. 15.1

[4] Michael Barlow. Paraconc-beta, version 1.0 – build 269. software, 2004. 7

[5] Irène Baron and Michael Herslund. Support Verb Constructions as Predicate Formation. In *The Structure of the Lexicon in Functional Grammar*. John Benjamins, Amsterdam/Philadelphia, 1998. 1, 4.2, 4.6, 3

[6] Cyril Belica. Statistische Kollokationsanalyse und Clustering. Korpuslinguistische Analysemethode. Institut für Deutsche Sprache, Mannheim, Germany, 1995. 15.3

[7] Bernd Heine and Ulrike Claudi and Friederike Hünnemeyer. *Grammaticalization. A Conceptual Framework*. University of Chicago Press, 2001. 3.1, 4, 3.5, 13, 14, 3.5.1, 17, 3.5.2, 3.5.2, 19

[8] Kristín Bjarnadóttir. Verbal Syntax in an Electronic Bilingual Icelandic Dictionary: A Preliminary Study. *LexicoNordica. Tidsskrift om leksikografi i Norden utgitt av Nordisk forening for leksikografi i samarbeid med Nordisk språksekretariat*, 8:5–24, 2001. 17.6

[9] Tavs Bjerre. Event Structure and Support Verb Constructions. In *Proceedings of the ESSLLI Student Session 1999*, 1999. 1, 4.4, 9

[10] Frede Boje. Hvor finder man *finde anvendelse*? In Ásta Svavarsdóttir, Guðrún Kvaran, and Jón Hilmar Jónsson, editors, *Nordiske Studier i Leksikografi Rapport fra Konferanse om leksikografi i Norden, Reykjavík 7.-10. juni 1995*, volume 3 of *Skrifter utgitt av Nordiske forening for leksikografi*, pages 51–68, Reykjavík, 1995. 1, 4.3

[11] I.A. Bolshakov, A.F. Gelbukh, and S.N. Haro Galicia. Electronic Dictionaries: For both Humans and Computers. In *International Forum on Information and Documentation*. Federation Internationale de Documentation (FID), 1999. 1.3

[12] Anna Braasch. Formalised Representation of Collocations in a Danish Computational Lexicon. In *The ninth EURALEX International Congress Proceedings*, volume II., pages 475–488. Stuttgart, 2000. 9.8

[13] Anna Braasch and Sussi Olsen. Towards a Strategy for a Representation of Collocations – Extending the Danish PAROLE Lexicon. In *Second International Conference on Language Resources and Evaluation, Proceedings. vol. 2, Athen*, pages 1009–1064, 2000. 1, 9.8

[14] Anna Braasch and Sussi Olsen. Formalised Representation of Collocations in a Danish Computational Lexicon. In *The Ninth EURALEX International Congress, Proceedings, Vol. II, Stutztgart*, pages 475–488, 2000. 1

[15] Miriam Butt. The Light Verb Jungle. *Harvard Working Papers in Linguistics*, 9(Papers from the Harvard/Dudley House Light Verb Workshop), 2003. URL http://ling.uni-konstanz.de/pages/home/butt,quoted2007-01-19. 4.1, 4.3, 4.4, 4.8

[16] Joan Bybee. *Morphology: a study of the relation between meaning and form*, volume 9 of *Typological Studies in Language*. John Benjamins, Amsterdam/Philadelphia, 1985. 3.2, 3, 3.2, 3.3, 5, 3.3.5, 7, 9, 3.3.5, 3.4, 3.5.1, 3

[17] Ilse Cantell. Målspråkets verbkonstruktioner i en tvåspråkig produktionsordbok. *LexicoNordica. Tidsskrift om leksikografi i Norden utgitt av Nordisk forening for leksikografi i samarbeid med Nordisk språksekretariat*, 2:19–32, 1995. 17.6

[18] Tommaso Caselli, Nancy Ide, and Roberto Bartolini. A Bilingual Corpus of Inter-linked Events. In Nicoletta Calzolari, Khalid Choukri, Bente Maegaard, Joseph Mariani, Jan Odjik, Stelios Piperidis, and Daniel Tapias, editors, *Proceedings of the Sixth International Language Resources and Evaluation (LREC'08)*, Marrakech, Morocco, may 2008. European Language Resources Association (ELRA). ISBN 2-9517408-4-0. URL http://www.lrec-conf.org/proceedings/lrec2008/. 17.6

[19] František Čermák, Jiří Hronek, Jaroslav Machač, et al. *Slovník české frazeologie a idiomatiky – výrazy slovesné*, volume 3, chapter Introduction, pages 26–28. Academia, Praha, 1994. 9.3, 9.3, 9.3, 4

[20] František Čermák. Víceslovná pojmenování typu verbum – substantivum v češtině (příspěvek k syntagmatice tzv. abstrakt). *Slovo a slovesnost*, 4(35):287–306, 1974. 4.6

[21] František Čermák. Abstract Nouns Collocations: Their Nature in a Parallel English-Czech Corpus. In *Meaningful Texts: The Extraction of Semantic Information from Monolingual and Multilingual Corpora*. Barnbrook, Danielsson & Mahlberg, Birmingham, 1995. 4.2

[22] František Čermák. Syntagmatika slovníku: typy lexikálních kolokací. In Z. Hladká and P. Karlík, editors, *Čeština – Univerzália a specifika*, volume 3, pages 223–232. Brno, 2001. 2.2

[23] Silvie Cinková. Extraction of Swedish Verb-Noun Collocations from a Large Msd-Annotated Corpus. *The Prague Bulletin of Mathematical Linguistics 82*, pages 99–102, 2004. 15.5

[24] Silvie Cinková. From PropBank to EngValLex: Adapting the PropBank-Lexicon to the Valency Theory of the Functional Generative Description. In *LREC 2006 Proceedings*. Genoa, Italy, May 2006. 9.5.3

[25] Silvie Cinková and Veronika Kolářová. *Nouns as Components of Support Verb Constructions in the Prague Dependency Treebank*, pages 113–139. Veda Bratislava, Slovakia, 2005. ISBN 80-224-0880-8. 4.6

[26] Silvie Cinková and Jan Pomikálek. LEMPAS: A make-do lemmatizer for the swedish PAROLE-corpus. *Prague Bulletin of Mathematical Linguistics*, 86:47–54, 2006. ISSN 0032-6585. 15.4.1

[27] Silvie Cinková, Jan Hajič, Marie Mikulová, Lucie Mladová, Anja Nedolužko, Petr Pajas, Jarmila Panevová, Jiří Semecký, Jana Šindlerová, Josef Toman, Zdeňka Urešová, and Zdeněk Žabokrtský. Annotation of English on the Tectogrammatical Level. Technical Report 35, UFAL MFF UK, 2006. 6.3.2

[28] Silvie Cinková, Petr Podveský, Pavel Pecina, and Pavel Schlesinger. Semi-automatic Building of Swedish Collocation Lexicon. In *Proceedings of the 5th International Conference on Language Resources and Evaluation (LREC 2006)*, pages 1890–1893, Paris, France, 2006. ISBN 2-9517408-2-4. 15.5

[29] Silvie Cinková, Josef Toman, Jan Hajič, Kristýna Čermáková, Václav Klimeš, Lucie Mladová, Jana Šindlerová, Kristýna Tomšů, and Zdeněk Žabokrtský. Tectogrammatical Annotation of the Wall Street Journal. *Prague Bulletin of Mathematical Linguistics*, 92, 2009. ISSN 0032-6585. 1

[30] Ulla Clausén et al. *Svenskt Språkbruk – ordbok över konstruktioner och fraser*. Norstedts Ordbok and Svenska Språknämnden, 2003. 9.7, 15.4.4, 16.2

[31] Jan Cuřín, Martin Čmejrek, Jiří Havelka, Jan Hajič, Vladislav Kuboň, and Zdeněk Žabokrtský. Prague Czech-English Dependency Treebank Version 1.0, 2004. 9.5.3

[32] Czech National Corpus. Czech National Corpus – SYN2005. Institute of the Czech National Corpus, Praha, 2005. URL `http://www.korpus.cz`. 15.2

[33] Ela Dura. *Substantiv och stödverb*, volume 18 of *Meddelanden från Institutionen för Svenska Språket*. Göteborgs universitet, 1997. 1, 4.1, 4.5, 5.2, 5.7, 15.5

[34] Eva Ejerhed, Gunnel Källgren, and Benny Brodda. Stockholm-Umeå Corpus Version 2.0. Stockholm University, Dep. of Linguistics and Umeå Uni-versity, Dep. of Linguistics, 2006. 15.2

[35] Lena Ekberg. *Gå till anfall och falla i sömn. En strukturell och funktionell beskrivning av abstrakta övergångsfaser*, volume A 43 of *Lundastudier i nordisk språkvetenskap*. Lund University Press, Lund, 1987. 1, 4.5

[36] Lena Ekberg. Verbet *ta* i metaforisk och grammatikaliserad användning. *Språk och Stil*, 3: 105–139, 3. 4, 15.5

[37] Michael Rundell et al. *Macmillan English Dictionary for Advanced Learners*. Macmillan Publishers Ltd., 2002. 1, 15.3

[38] Stefan Evert and Hannah Kermes. Experiments on Candidate Data for Collocation Extraction. In *Companion Volume to the Proceedings of the 10th Conference of The European Chapter of the Association for Computational Linguistics*, pages 83–86, Budapest, Hungary, 2003. 2.2

[39] Stefan Evert and Birgitte Krenn. Exploratory Collocation Extraction. In *PHRASEOLOGY 2005 The many faces of Phraseology*. Université catholique de Louvain (Belgium), 13-15 October 2005. 2.2

[40] C. Fellbaum. *WordNet: An Electronical Lexical Database*. The MIT Press, Cambridge, MA, 1998. 9.4

[41] Katalin Fenyvesi-Jobbágy. Non-literal and non-metaphorical uses of Danish komme 'come': A case study. *Jezikoslovlje*, 2003. 4

[42] Charles J. Fillmore. Frame Semantics. In *Linguistics in the Morning Calm*, pages 111–137. Hanshin Publishing, Seoul, South Corea, 1982. 9.6

[43] Charles J. Fillmore, Christopher R. Johnson, and M. L. R. Petruck. Background to FrameNet. *FrameNet and Frame Semantics. International Journal of Lexicography – Special Issue*, 16:235–250, 2003. 4.4, 9.6

[44] John Rupert Firth. *Papers in Linguistics 1934-1951*. London: Oxford University Press, 1957. 2.2

[45] John Rupert Firth. *Papers in Linguistics 1934-1951*, chapter Modes of Meaning, pages 190–215. London: Oxford University Press, 1957. 5.2

[46] Thierry Fontenelle. Co-occurence Knowledge, Support verbs and Machine Readable Dictionaries. In *Papers in Computational Lexicography, COMPLEX'92, Budapest*, 1992. 4.2, 17.6

[47] Thierry Fontenelle. Using a Bilingual Computerized Dictionary to Retrieve Support Verbs and Combinatorial Information. *Acta Linguistica Hungarica*, 41(1-4):109–121, 1993. 17.6

[48] Thierry Fontenelle. Using a Bilingual Dictionary to Create Semantic Networks. *International Journal of Lexicography*, 10(4):274–303, 1997. 17.6

[49] Stefan Thomas Gries. Dispersions and Adjusted Frequencies in Corpora. *International Journal of Corpus Linguistics*, 13(4):403–437, 2008. 7.3

[50] Heide Günther and Sabine Pape. Funktionsverbgefüge als Problem der Beschreibung komplexer Verben in der Valenztheorie. In Helmut Schumacher, editor, *Untersuchungen zur Verbvalenz: eine Dokumentation über die Arbeit an einem deutschen Valenzlexikon*, Forschungsberichte/Institut für deutsche Sprache Mannheim, pages 92–128. Narr, Tübingen, 1976. 4.1

[51] Jan Hajič. SE030107x – A statistical tagger/lemmatizer based on the SUC corpus. Personal communication 2006-03-19, 2002. 3

[52] Jan Hajič, Jarmila Panevová, Zdeňka Urešová, Alevtina Bémová, Veronika Kolářová, and Petr Pajas. PDT-VALLEX: Creating a Large-coverage Valency Lexicon for Treebank Annotation. In *Proceedings of The Second Workshop on Treebanks and Linguistic Theories*, volume 9 of *Mathematical Modeling in Physics, Engineering and Cognitive Sciences*, pages 57–68. „ Vaxjö University Press, November 14–15, 2003 2003. ISBN 91-7636-394-5. GA405/03/0913, LN00A063. 6.3.1, 9.5.1

[53] Jan Hajič, Jarmila Panevová, Eva Hajičová, Petr Sgall, Petr Pajas, Jan Štěpánek, Jiří Havelka, Marie Mikulová, Zdeněk Žabokrtský, and Magda Ševčíková Razímová. Prague Dependency Treebank 2.0. Software prototype, Linguistic Data Consortium, Philadelphia, PA, USA, ISBN 1-58563-370-4, Jul 2006, 2006. URL `http://ufal.mff.cuni.cz/pdt2.0/`. 6.1, 1, 6.3.5

[54] Patrick Hanks. The Probable and the Possible: Lexicography in the Age of Internet. In Lee Sangsup, editor, *Asialex 2001 Proceedings, Seoul Korea*, 2001. 1, 17.2

[55] Patrick Hanks. Lexicography. In Ruslan Mitkov, editor, *The Oxford Handbook of Computational Linguistics*, chapter I. – Fundamentals, pages 48–70. Oxford University Press, Oxford, 2003. 3.1

[56] Patrick Hanks. *Norms and Exploitations: Corpus, Computing, and Cognition in Lexical Analysis*. MIT Press, forthcoming. Manuscript, obtained 2003 from the author. 1.1, 1.3, 3.1, 3.5.2, 4.2, 7.1, 1, 7.2, 7.3

[57] Patrick Hanks and Elisabetta Jezek. Shimmering Lexical Sets. In *Euralex XIII 2008 Proceedings*, Pompeu Fabra University, Barcelona, 2008. 7.2

[58] Patrick Hanks and James Pustejovsky. A Pattern Dictionary for Natural Language Processing. *Revue Francaise de linguistique appliquée*, 10(2), 2005. 7.2, 7.3, 16.3.7, 17.5

[59] Patrick Hanks, Anne Urbschat, and Elke Gehweiler. German Light Verb Constructions in Corpora and Dictionaries. *International Journal of Lexicography – Special Issue: Corpus-Based Studies of German Idioms and Light Verbs*, 19(4):439–458, 2006. 2, 2.2, 4.1, 4.7, 4.7, 4.8, 5.7

[60] Erik Hansen. *Stå, sidde, ligge*. *Mål & Mæle*, 1(2):26–32, 1974. 4

[61] Ulrich Heid. Towards a Corpus-based Dictionary of German Noun-verb Collocations. In Thierry Fontenelle, Philippe Hilligsmann, Archibald Michiels, AndréMoulin, and Siegfried Theissen, editors, *Actes EURALEX'98 Proceedings*, volume 1, pages 301–312, Liège, 1998. Universitéde Liège, Départements d'anglais et de néerlandai. 4

[62] Gerhard Helbig and Joachim Buscha. *Deutsche Grammatik. Ein Handbuch für den Ausländerunterricht*. Verlag Enzyklopädie, Leipzig, 1996. 1.2, 4.1, 4.3, 4.7, 4.7, 4.7, 4.7

[63] Dana Hlaváčková, Aleš Horák, and Vladimír Kadlec. *TSD 2006*, chapter Exploitation of the VerbaLex Verb Valency Lexicon in the Syntactic Analysis of Czech. Springer Verlag, 2006. 9.9

[64] Michael Hoey. *Lexical Priming. A new theory of words and language*. Routledge, 2005. 2.2

[65] Paul Hopper. Emergent Grammar. *Berkeley Linguistics Conference (BLS)*, 13:139–157, 1987. URL `http://eserver.org/home/hopper/emergence.html`. 1.1, 3.2, 3.4, 3.5.2, 10.2

[66] Hopper, Paul J. and Thompson, Sandra A. Transitivity in Grammar and Discourse. *Language*, 56(2):251–299, 1980. 9, 5.1, 5.1, 5.3, 3, 4, 5.3.10, 5.4, 5.5, 5.6

[67] Susan Hunston and Gill Francis. *Pattern Grammar: a Corpus-driven Approach to the Lexical Grammar of English*. John Benjamins Publishing Company, 200. 7.3

[68] Igor Mel'čuk, Alexander Žolkovskij. *Tolkovo-kombinatornyj slovar'sovremennovo russkovo jazyka*. Wiener Slawistischer Almanach, Sonderband 14, 1984. 8.1, 9.2.4

[69] Intercorp. Intercorp – projekt paralelních korpusů Filozofické fakulty Univerzity Karlovy v Praze. www.korpus.cz-intercorp, quoted 2009. 15.2, 17.6

[70] Ulrike Jakobsson. Familjelika betydelser hos STÅ, SITTA och LIGGA. En analys ur den kognitiva semantikens perspektiv. Technical report, Lunds universitet. Institutionen för nordiska språk, Lund, 1996. 4, 2, 15.5

[71] M. Jelínek. O verbonominálních spojeních ve spisovné češtině. In *Přednášky a besedy z XXXVI. běhu LŠSS*, pages 37–51. MU Brno, 2003. 4.1, 4.3, 4.7, 15, 16

[72] Torben Juel Jensen. Kan man 'ligge' i et mentalt rum? In *Nydanske studier & almen kommunikationsteori. Artikler om partikler.* , pages 73–100. Københavns Universitet. Institut for Nordisk Filologi, København, 2000. 4

[73] Otto Jespersen. *A Modern English Grammar on Historical Principles*, volume 6. London: George Allen & Unwin & Copenhagen: Ejnar Munksgaard., 1954. 2, 4.1

[74] Joan Bybee and Revere Perkins and William Pagliuca. *The Evolution of Grammar. Tense, aspect, and modality in the languages of the world*. The University of Chicago Press, Chicago & London, 1994. 3.2, 2, 3.2, 3.3, 3.4, 3.5, 13, 15, 10.2, 1

[75] Christopher Johnson and Charles Fillmore. The FrameNet tagset for framesemantic and syntactic coding of predicate-argument structure, 2000. URL `citeseer.ist.psu.edu/johnson00framenet.html`. 9.6

[76] Sylvain Kahane. The Meaning-Text Theory. In *Dependency and Valency. An International Handbook on Contemporary Research*. de Gruyter, Berlin, 2003. 8.1, 9.2

[77] Fred Karlsson. SWETWOL: a comprehensive morphological parser for Swedish. *Nordic Journal of Linguistics*, 15:1–45, 1992. 2

[78] Adam Kilgarriff, Pavel Rychlý, Pavel Smrž, and David Tugwell. The Sketch Engine. In *Proceedings of the Eleventh EURALEX International Congress. Lorient, France*, pages 105–116. Universite de Bretagne-Sud, 2004. 15.2, 15.3

[79] Karin Kipper, Martha Palmer, and Owen Rambow. Extending PropBank with VerbNet Semantic Predicates. In *Workshop on Applied Interlinguas, held in conjunction with AMTA-2002*. Tiburon, CA, October 2002. 9.4

[80] Veronika Kolářová. *Valence deverbativních substantiv v češtině (PhD thesis)*. PhD thesis, ÚFAL MFF UK, Prague, 2005. 4.6

[81] Veronika Kolářová. Valency of Deverbal Nouns in Czech. *Prague Bulletin of Mathematical Linguistics*, 86:5–20, 2006. ISSN 0032-6585. 6.3.1

[82] George Lakoff. *Women, Fire and Dangerous Things. What categories reveal about the mind*. Chicago University Press, Chicago, 1987. 3.1, 4, 3.5.1

[83] George Lakoff and Mark Johnson. *Metaphors We Live By*. Chicago University Press, 1980. 4

[84] Mats Larsson. translation of B. Hrabal's text "Taneční hodiny pro starší a pokročilé". personal communication, September 18 2006. 3, 6

[85] Beth Levin. *English Verb Classes and Alternations*. University of Chicago Press, 1993. 9.1, 9.4

[86] Ann Lindvall. *Transitivity in Discourse. A Comparison of Greek, Polish and Swedish.*, volume 37 of *Travaux de l'Institut de Linguistique de Lund*. Lund University Press, 1998. 5.1, 5.1, 5.1, 5.6

[87] Ann Lindvall. Swedish verb particles and Polish aspect marking. In Arthur Holmer, Jan-Olof Svantesson, and Åke Viberg, editors, *Proceedings of the 18th Scandinavian Conference of Linguistics, Lund 18-20 May 2000*, volume 2, 2001. 5.6

[88] Markéta Lopatková. Valency in the Prague Dependency Treebank: Building the Valency Lexicon. *Prague Bulletin of Mathematical Linguistics*, 79–80:37–60, 2003. 9.5.2

[89] Markéta Lopatková and Jarmila Panevová. *Recent Developments of the Theory of Valency in the Light of the Prague Dependency Treebank*, pages 83–92. Veda Bratislava, Slovakia, 2005. ISBN 80-224-0880-8. 6.3.1, 6.3.5

[90] Markéta Lopatková, Zdeněk Žabokrtský, Karolina Skwarska, and Václava Benešová. VALLEX 1.0 Valency Lexicon of Czech Verbs. Technical Report TR-2003-18, UFAL/CKL MFF UK, Prague, 2003. 1.3, 9.9

[91] Markéta Lopatková, Zdeněk Žabokrtský, Karolina Skwarska, and Václava Benešová. VALLEX 1.0 Valency Lexicon of Czech Verbs. Technical Report TR-2003-18, UFAL/CKL MFF UK, Prague, 2003. 6.3.6

[92] Markéta Lopatková, Zdeněk Žabokrtský, and Karolina Skwarska. Valency Lexicon of Czech Verbs. In *LREC 2006 Proceedings*. Genoa, Italy, May 2006. 9.5.2

[93] Markéta Lopatková, Zdeněk Žabokrtský, Václava Kettnerová, Karolina Skwarska, Eduard Bejček, Klára Hrstková, Michaela Nová, and Miroslav Tichý. VALLEX 2.5 – Valency Lexicon of Czech Verbs, version 2.5. Software prototype, 2007. 16.3.1

[94] Markéta Lopatková, Zdeněk Žabokrtský, and Václava Kettnerová. *Valenční slovník českých sloves*. Nakladatelství Karolinum, Praha, 2008. ISBN 978-80-246-1467-0. 6.3.6

[95] Eva Macháčková. Analytické predikáty. Substantivní názvy dějů a statických situací ve spojení s funkčními slovesy. *Jazykovědné aktuality*, 3, 4(10):122–176, 1983. 4.6, 12

[96] C. Macleod. Lexical Annotation for Multi-word Entries Contatining Nominalizations. In *Proceedings of Third International Conference on Language Resources and Evaluation (LREC 2002); Las Palmas, Canary Islands, Spain*, pages 943–948, 2002. 9.4

[97] Sven-Göran Malmgren. *Begå eller ta självmord? Om svenska kollokationer och deras förändringsbenägenhet 1800-2000*. Rapporter från ORDAT. Göteborgs universitet. Institutionen för svenska språket., Göteborg, 2002. 4.2, 4, 15.5

[98] Zdeněk Martínek. Winconcord. software, 1996. Developed by Zdeněk Martínek at the University of West Bohemia, Pilsen, Czech Republic, in close Collaboration with Les Siegrist from the Technische Hochschule Darmstadt, Germany. Version 2.0 (July, 1996). 15.5

[99] Igor Mel'čuk. *Dependency Syntax: Theory and Practice*. State University of New York Press, 1988. 8.1, 9.2.3

[100] Igor Mel'čuk. *Dictionnaire explicatif et combinatoire du français contemporain*. Les Presses de l'Université de Montréal, 1984. 8.1, 9.2.4

[101] Igor A. Mel'čuk. Lexical Functions: A Tool for the Description of Lexical Relations in a Lexicon. In Leo Wanner, editor, *Lexical Functions in Lexicography and Natural Language Processing*, pages 37–105. John Benjamins, Amsterdam/Philadelphia, 1996. 8.2, 8.3, 8.3.1, 8.3.2, 8.3.3

[102] Igor A. Mel'čuk and Alain Polguère. A Formal Lexicon in the Meaning-Text Theory (or How to Do Lexica with Words). *Computational Linguistics*, 13(3–4):261–275, 1987. 9.2, 9.2.1, 9.2.2

[103] Adam Meyers, R. Reeves, C. Macleod, R. Szekely, V. Zielinska, B. Young, and R. Grishman. Annotating Noun Argument Structure for NomBank. In *Proceedings of LREC-2004*, Lisbon, Portugal, 2004. 9.4

[104] Adam Meyers, R. Reeves, C. Macleod, R. Szekely, V. Zielinska, B. Young, and R. Grishman. The NomBank Project: An Interim Report. In *HLT-NAACL 2004 Workshop: Frontiers in Corpus Annotation*, pages 24–31, Boston, Massachusetts, USA, May 2 – May 7 2004. Association for Computational Linguistics. 9.4

[105] Adam Meyers, Ruth Reeves, and Catherine MacLeod. *NomBank v 1.0*. Linguistic Data Consortium, Philadelphia, 2008. 6.3.1, 6.3.7

[106] Marie Mikulová, Alevtina Bémová, Jan Hajič, Eva Hajičová, Jiří Havelka, Veronika Kolářová, Markéta Lopatková, Petr Pajas, Jarmila Panevová, Magda Razímová, Petr Sgall, Jan Štěpánek, Zdeňka Urešová, Kateřina Veselá, Zdeněk Žabokrtský, and Lucie Kučová. Anotace na tektogramatické rovině Pražského závislostního korpusu. Anotátorská příručka. Technical Report TR-2005-28, ÚFAL MFF UK, Prague, Prague, 2005. 4.6, 4.6, 4.6

[107] Marie Mikulová, Alevtina Bémová, Jan Hajič, Eva Hajičová, Jiří Havelka, Veronika Kolářová, Lucie Kučová, Markéta Lopatková, Petr Pajas, Jarmila Panevová, Magda Razímová, Petr Sgall, Jan Štěpánek, Zdeňka Urešová, Kateřina Veselá, and Zdeněk Žabokrtský.

Annotation on the Tectogrammatical Level in the Prague Dependency Treebank. Annotation Manual. Technical Report 30, ÚFAL MFF UK, Prague, Czech Rep., 2006. 6.3.1, 6.3.6

[108] P. M. Mitchell et al. Building a Large Annotated Corpus of English: The PennTreebank. *Computational Linguistics*, 1993. 9.4

[109] Morton Benson and Evelyn Benson and Ilson, Robert F. *BBI Dictionary of English Word Combinations*. John Benjamins, 1997. 2.2, 8.1, 9.1

[110] Alexander Nakhimovsky. A Case of Aspectual Polysemy, with Implications for Lexical Functions. In Leo Wanner, editor, *Lexical Functions in Lexicography and Natural Language Processing*, pages 169–179. John Benjamins, Amsterdam/Philadelphia, 1996. 5.3.3

[111] Sanni Nimb and Bolette Sandford Pedersen. Treating Metaphoric Senses in a Danish Computational Lexicon – Different Cases of Regular Polysemy. In *EURALEX 2000, Proceedings*,, pages 679–692. Stuttgart, 2000. 9.8

[112] Joakim Nivre, Jens Nilsson, and Johan Hall. Talbanken05: A Swedish Treebank with Phrase Structure and Dependency Annotation. In *Proceedings of the fifth international conference on Language Resources and Evaluation (LREC2006)*, May 2006. URL http://www.msi.vxu.se/users/nivre/papers/talbanken05.pdf. 15.4.4

[113] Elisabeth Nylund and Britta Holm. *Deskriptiv svensk grammatik*. Stockholm, 1993. 1.2, 15.4.1

[114] Karel Pala and Pavel Ševeček. Valence českých sloves. In *Sborník prací FFBU*, 1997. 9.9

[115] Martha Palmer, Daniel Gildea, and Paul Kingsbury. The Proposition Bank: An Annotated Corpus of Semantic Roles. *Computational Linguistics*, 31(1):71–106, 2005. URL http://www.cs.rochester.edu/~gildea/palmer-propbank-cl.pdf. 9.4

[116] Martha Palmer, Paul Kingsbury, Olga Babko-Malaya, Scott Cotton, and Benjamin Snyder. Proposition Bank I. LDC2004T14, ISBN: 1-58563-304-6, Sep 01, 2004. 9.4

[117] Jarmila Panevová. On Verbal Frames in Functional Generative Description I. *Prague Bulletin of Mathematical Linguistics*, 22:3–40, 1974. 6.3.1, 6.3.7

[118] Jarmila Panevová. On Verbal Frames in Functional Generative Description II. *Prague Bulletin of Mathematical Linguistics*, 23:17–52, 1975. 6.3.4, 3, 6.3.6

[119] Jarmila Panevová. Valency Frames and the Meaning of the Sentence. In Philip A. Luelsdorff, editor, *The Prague School of Structural and Functional Linguistics*, pages 223–243. John Benjamins Publishing Company, Amsterdam, Philadelphia, 1994. 6.3.1

[120] Jarmila Panevová. Poznámky k valenci podstatných jmen. In *Čeština – univerzália a specifika 2. Sborník konference ve Šlapanicích u Brna, 17.-19.11.1999 (ed. Zdeňka Hladká, Petr Karlík)*, pages 173–180. Masarykova Univerzita v Brně, ISBN 80-210-2262-0, 2000. 6.3.1, 6.3.7, 6.3.7

[121] Jarmila Panevová. Sloveso: centrum věty, valence: centrální pojem syntaxe. In *Proceedings of Aktuálne otázky súčasnej syntaxe*, pages 73–77. Budmerice, Slovakia,, Nov. 7-8 2002. 9.5.2

[122] PAROLE. The PAROLE Corpus at The Swedish Language Bank. University of Gothenburg, 1997. URL http://spraakbanken.gu.se/parole. 15.2

[123] Pavel Pecina and Pavel Schlesinger. Combining Association Measures for Collocation Extraction. In *Proceedings of the 21th International Conference on Computational Linguistics and 44th Annual Meeting of the Association for Computational Linguistics (COLING/ACL 2006), Poster Sessions*, Sydney, Australia, July 2006. 15.5

[124] Bolette Sandford Pedersen. Den danske SIMPLE-ordbog – en semantisk, ontologibaseret ordbog. In *Datalingvistisk Forenings Rrsmøde 1999, Proceedings.*, pages 71–80. København, 1999. 9.8

[125] Bolette Sandford Pedersen and Sanni Nimb. Semantic Encoding of Danish Verbs in SIMPLE – Adapting a Verb-framed Model to a Satellite-framed Language. In *Proceedings from 2nd Conference on Language Resources and Evaluation, LREC.*, pages 1405–1412. Athens, 2000. 9.8

[126] Susanne Nøhr Pedersen. The Treatment of Support Verbs and Predicative Nouns in Danish. In Jörgen Pind and Eiríkur Rögnvaldsson, editors, *Papers from the Seventh Scandinavian Conference of Computational Linguistics*, pages 208–218, Reykjavík, 1990. Institute of Lexicography. 4.6

[127] Ingemar Persson. *Das System der kausativen Funktionsverbgefüge. Eine semantisch-syntaktische Analyse einiger verwandter Konstruktionen*. Liber, Malmö, 1975. PhD thesis. 4.1, 4.2

[128] Ingemar Persson. Das kausative Funktionsverbgefüge (FVG) und dessen Darstellung in der Grammatik und im Wörterbuch. *Deutsche Sprache*, 20:153–171, 1992. 4.1, 4.2, 4.5

[129] Sven Pihlström. *Hålla på* och *hålla på och*. *Språkvård*, 2:8–10, 1988. 4, 12.2, 2, 13

[130] Petr Piťha. On the Case Frames of Nouns. *Prague Studies in Mathematical Linguistics*, 1981. 6.3.1, 6.3.7

[131] Peter von Polenz. Funktionsverben im heutigen Deutsch. *Sprache in der rationalisierten Welt. Wirkendes Wort*, Beiheft 5, 1963. 4.1, 4.7

[132] Alain Polguère. Towards a theoretically motivated general public dictionary of semantic derivations and collocations for French. In *Proceedings of EURALEX 2000*, pages 517–527, 2000. 9.2

[133] projektet OSA – Svenska Akademiens Ordbok i databasform. Svenska akademiens ordbok, quoted 2006-09-17. URL URL: <http: //g3. spraakdata. gu. se/saob/index. html>. 2

[134] PropBank Annotation. PropBank Annotation Guidelines. Version 3., February 22 2002. URL http: //verbs. colorado. edu/~mpalmer/projects/ace/PBguidelines. pdf. [quoted 2009-02-23]. 9.4

[135] James Pustejovsky. The Syntax of Event Structure. *Cognition*, 41:47–81, 1991. 4.4

[136] James Pustejovsky. *The Generative Lexicon*. MIT Press, Cambridge, MA, 1995. 9.8

[137] James Pustejovsky, Catherine Havasi, Jessica Littman, Anna Rumshisky, and Marc Verhagen. Towards a Generative Lexical Resource: The Brandeis Semantic Ontology. In *Proceedings of the Fifth Language Resource and Evaluation Conference*, 2006. 7.2

[138] Randolph Quirk, Sidney Greenbaum, Geoffrey Leech, and Jan Svartvik. *A Comprehensive Grammar of the English Language*. Longman, London ; New York :, 1985. ISBN 0582517346 0582517346 0582517346 0582517346 0582517346. 3

[139] Mikael Reuter. Lägg ribban högt. *Reuters Ruta*, Forskningscentralen för de inhemska språken 1986. URL http: //www. kotus. fi/svenska/reuter/ Kotimaistenkieltentutkimuskeskus. quoted2003- 04- 16. 4, 15.5

[140] Veronika Řezníčková. Czech Deverbal Nouns: Issues of Their Valency in Linear and Dependency Corpora. In Petya Simov, Kiril//Osenova, editor, *Proceedings of the Workshop on Shallow Processing of Large Corpora (SProLaC 2003)*, pages 88–97, Lancaster, 2003. UCREL, Lancaster University. ISBN 1-86220-134-X. 4.6

[141] Annely Rothkegel. *Feste Syntagmen. Grundlagen, Strukturbeschreibung und automatische Analyse*. Linguistische Arbeiten. Niemeyer, Tübingen, 1973. 4.1, 4.2, 4.3

[142] Ruth Feil. Funktionsverber i det danske sprog. In *Nordiske Studier i Leksikografi. Rapport fra Konferanse om leksikografi i Norden, Reykjavík 7.-10. juni 1995*, volume 3 of *Skrifter utgitt av Nordiske forening for leksikografi*, pages 137–148, Reykjavík, 1995. 1

[143] Pavel Rychlý and Pavel Smrž. Manatee, Bonito and Word Sketches for Czech. In *Proceedings of the Second International Conference on Corpus Linguisitcs*, pages 124–131, 2004. ISBN 5-288-03531-8. URL http://www.fit.vutbr.cz/research/view_pub.php?id=7700. 15.2

[144] Jan Schroten. Light Verb Constructions in Bilingual Dictionaries. In *From Lexicology to Lexicography*, pages 83–94. University Utrecht. Utrecht Institute of Linguistics OTS., Utrecht, 2002. 4.2, 4.3, 8

[145] Jiří Semecký and Silvie Cinková. Constructing an English Valency Lexicon. In *Proceedings of the Workshop on Frontiers in Linguistically Annotated Corpora 2006*, pages 94–97, Sydney, Australia, July 2006. Association for Computational Linguistics. URL http://www.aclweb.org/anthology/W/W06/W06-0612. 9.5.3

[146] Gilles Sérasset. A Generic Collaborative Platform for Multilingual Lexical Database Development. In *COLING 2004, Multilingual Linguistic Resources*, pages 73–79, Geneva, Switzerland, Aug. 2004. 9.2.4, 17.6

[147] Petr Sgall. *Generativní popis jazyka a česká deklinace*. Prague:Academia, 1967. 6.1

[148] Petr Sgall. Underlying Structure of Sentence and its Relation to Semantics. In T. Reuther, editor, *Wiener Slawisticher Almanach. Sonderband 33*, pages 273–282. Institut für Slavische Philologie, Universität München, 1992. 6.1

[149] Petr Sgall, Eva Hajičová, and Jarmila Panevová. *The Meaning of the Sentence in Its Semantic and Pragmatic Aspects*. Dordrecht:Reidel Publishing Company and Prague:Academia, 1986. 6.1

[150] John Sinclair. *Corpus, Concordance, Collocation*. Oxford University Press, 1993. 3rd edition, 1st edition 1991. 1.1, 2.2, 3.1, 4.2, 10.1

[151] Otakar Smrž, Viktor Bielický, Iveta Kouřilová, Jakub Kráčmar, Jan Hajič, and Petr Zemánek. Prague Arabic Dependency Treebank: A Word on the Million Words. In *Proceedings of the Workshop on Arabic and Local Languages (LREC 2008)*, pages 16–23, Marrakech, Morocco, 2008. 1

[152] Språkbanken. Språkbanken vid Göteborgs universitet, since 1975. URL http://spraakbanken.gu.se. 3

[153] U. Teleman, S. Hellberg, and E. Andersson. *Svenska Akademiens grammatik*. Svenska Akademien/Norstedts, Stockholm, 1999. 1.2, 10.2, 2, 3, 4, 6, 8, 11.3.3, 12.2, 12.4, 12.6, 12.6, 13

[154] Maria Toporowska-Gronostaj and Karin Warmenius. SWEDISH SIMPLE – Lexicon Documentation. Technical report, Språkdata, University of Gothenburg, 2000. URL `http://www.ub.es/gilcub/SIMPLE/reports/simple/D312_GOTrev.html`. [quoted 2003-03-10, S.C.]. 9.8

[155] Harald Trost. Morphology. In Ruslan Mitkov, editor, *The Oxford Handbook of Computational Linguistics*, chapter I. – Fundamentals, pages 25–47. Oxford University Press, Oxford, 2003. 6

[156] Åke Viberg. Fysiska kontaktverb i svenskan. En skiss. *Svenskans beskrivning*, 14:174–185, 1984. 15.5

[157] Åke Viberg. Svenskans lexikala profil. *Svenskans beskrivning*, 17:391–408, 1990. 10.1, 15.5

[158] Åke Viberg. The Meanings of Swedish *dra pull*: a Case Study of Lexical Polysemy. In *EURALEX'96 Proceedings*, volume 1, pages 293–308, Göteborg, 1996. Dept. of Swedish, Göteborg University. 15.5

[159] Åke Viberg. Cross-linguistic lexicology. The case of English *go* and Swedish *gå*. In *Languages in Contrast. Papers from a Symposium on Text-based Cross-linguistic Studies*, volume 88 of *Lund Studies in English*, pages 157–186, Göteborg, 1996. Dept. of Swedish, Göteborg University. 15.5

[160] Åke Viberg. Polysemy and Differentiation in the Lexicon. Verbs of Physical Contact in Swedish. In Jens Alwood and Peter Gärdenfors, editors, *Cognitive Semantics. Meaning and Cognition*. John Benjamins, Amsterdam, 1998. 15.5

[161] Åke Viberg. *A Man of Measure: Festschrift in Honour of Fred Karlsson on his 60th Birthday*, volume 19, chapter What One Verb can Do: The Swedish Verb *göra* in a Cross-linguistic Perspective, pages 243–257. The Linguistic Association of Finland, Turku, special supplement to sky journal of linguistics edition, 2006. 10.1

[162] Åke Viberg. Polysemy and Disambiguation Cues across Languages. The Case of Swedish *få* and English *get*. In S. Granger and B. Altenberg, editors, *Lexis in Contrast*, pages 119–150. John Benjamins, 2002. 10.1, 13, 15.5

[163] P. Vossen, editor. *EuroWordNet: A Multilingual Database with Lexical Semantic Networks.* Kluwer Academic Publishers, 1998. 9.9

[164] Leo Wanner, editor. *Lexical Functions in Lexicography and Natural Language Processing*, volume 31 of *Studies in Language Companion Series (SLCS)*, Amsterdam-Philadelphia, 1996. John Benjamins. 8.1

[165] Alison Wray. *Formulaic Language and the Lexicon*. Cambridge University Press, 2002. 5.2

[166] XMLmind. XMLmind XML Editor Personal Edition 3.7.0. free software version, www.xmlmind.com/xmleditor/, 2000–2007. 16.2

[167] Zdeněk Žabokrtský. *Valency Lexicon of Czech Verbs (PhD thesis)*. PhD thesis, Charles University, Prague, Czech Rep., 2005. 6.2, 9.5.2

[168] Zdeněk Žabokrtský and Markéta Lopatková. Valency Frames of Czech Verbs in VALLEX 1.0. In Adam Meyers, editor, *HLT-NAACL 2004 Workshop: Frontiers in Corpus Annotation*, pages 70–77, Boston, 2004. Association for Computational Linguistics. 6.3.6

[169] Heike Zinsmeister and Ulrich Heid. Collocations of Complex Words: Implications for the Acquisition with a Stochastic Grammar. In *Proceedings of the International Workshop on Computational Approaches to Collocations Vienna*, Vienna, Austria, 22., 23. July 2002 2002. Austrian Research Institute for Artificial Intelligence (ÖFAI). URL `http://www.ofai.at/~brigitte.krenn/colloc02/index.html`. 15.4.4

Index

E

F

G

H

I

K

L

M

N

T

U

V

W

X

Z