Constructicography

Constructicon development
across languages

edited by

Benjamin Lyngfelt
Lars Borin
Kyoko Ohara
Tiago Timponi Torrent

Constructicography

Constructional Approaches to Language

The series brings together research conducted within different constructional models and makes them available to scholars and students working in this and other related fields.

The topics range from descriptions of grammatical phenomena in different languages to theoretical issues concerning language acquisition, language change, and language use. The foundation of constructional research is provided by the model known as Construction Grammar (including Frame Semantics). The book series publishes studies in which this model is developed in new directions and extended through alternative approaches. Such approaches include cognitive linguistics, conceptual semantics, interaction and discourse, as well as typologically motivated alternatives, with implications both for constructional theories and for their applications in related fields such as communication studies, computational linguistics, AI, neurology, psychology, sociology, and anthropology.

This peer reviewed series is committed to innovative research and will include monographs, thematic collections of articles, and introductory textbooks.

For an overview of all books published in this series, please see
http://benjamins.com/catalog/cal

Editors

Jan-Ola Östman
University of Helsinki,
Finland

Kyoko Ohara
Keio University, Japan

Advisory Board

Peter Auer
University of Freiburg, Germany

Hans C. Boas
University of Texas at Austin, USA

William Croft
University of New Mexico, USA

Charles J. Fillmore†
International Computer Science
Institute, Berkeley, USA

Mirjam Fried
Charles University, Prague,
Czech Republic

Adele E. Goldberg
Princeton University, USA

Martin Hilpert
University of Neuchâtel, Switzerland

Seizi Iwata
Kansai University, Japan

Paul Kay
University of California, Berkeley,
USA

Knud Lambrecht
University of Texas at Austin, USA

Michael Tomasello
Duke University, USA

Arnold M. Zwicky
Stanford University, USA

Volume 22

Constructicography. Constructicon development across languages
Edited by Benjamin Lyngfelt, Lars Borin, Kyoko Ohara
and Tiago Timponi Torrent

Constructicography

Constructicon development across languages

Edited by

Benjamin Lyngfelt
University of Gothenburg

Lars Borin
University of Gothenburg

Kyoko Ohara
Keio University

Tiago Timponi Torrent
Federal University of Juiz de Fora

Table of contents

Preface

This book has been long in the making, but now having the result under our eyes, we believe that it was worth the wait. The idea of compiling a volume collecting the experiences of the various constructicon initiatives going on around the world was born in the context of an international collaboration between the universities at Gothenburg in Sweden and Juiz de Fora in Brazil, and the excellent opportunities to interact and learn from each other's experiences afforded by both research visits and the international FrameNet workshops organized jointly by the Swedish and Brazilian teams, together with the FrameNet group in Berkeley, California: IFNW 2013 in Berkeley, IFNW 2016, collocated with ICCG9 in Juiz de Fora, and the upcoming IFNW 2018 with the special theme *Multilingual FrameNets and Constructicons*, collocated with LREC in Miyazaki, Japan. Moreover, profitable discussions relevant to the works presented in this book took place in the special sessions *Cognitively grounded lexica, constructicons, and metaphor repositories*, at ICLC12 in Edmonton, Canada, in 2013, and *Constructionist resources – a workshop in honor of Charles J. Fillmore*, at ICCG8 in Osnabrück, Germany, in 2014.

During these events, most – if not all – authors of the chapters in this volume had the chance to share their points of view, positions and questions on the development of constructionist resources. Beyond the group of authors whose contributions make this book, we'd like to thank our – and their – interlocutors.

The volume editors would like to express their gratitude to the anonymous reviewers for their thorough and insightful comments and suggestions, to the series editor Jan-Ola Östman for his final vetting of the volume, and to Esther Roth and Susan Hendriks at John Benjamins, always helpful and unerringly professional, under whose watchful eyes the volume went from idea to finished product. Last but

not least, we are grateful to the person without whom none of all this would have happened: the late Charles J. Fillmore. For inspiring us all, for being a core element in our research frame(works), for having guided our ways into the development of constructicons, and for being such a great guy, we dedicate this volume to Chuck.

<div align="right">

Benjamin Lyngfelt
Lars Borin
Kyoko Ohara
Tiago Timponi Torrent

</div>

Introduction

Constructicons and constructicography

Benjamin Lyngfelt

A constructicon is on the one hand a theoretical conception of language as a structured inventory of constructions, and on the other hand a collection of construction descriptions, essentially a practical instantiation of the former concept. In this introductory chapter, we review the role of these notions in constructionist theory and practice, and relate it to the closely connected development of Frame Semantics into FrameNet. Practical constructicon development is characterized as a combination of construction grammar and lexicography, for which we introduce the term constructicography. Central issues in constructicography are introduced, setting the stage for the explorations in the following chapters.

Keywords: constructicon, construction, constructicography, construction grammar, lexicography, FrameNet

1. Introduction[1]

One of the central ideas in construction grammar (CxG) is the conception of a language, or at least its lexico-grammar, as a structured inventory of constructions: a *constructicon*. Although this notion has been at the heart of CxG since its early development in the 1980s (cf. Fillmore, 1988, p. 37; Jurafsky, 1991, p. 18), it remains one on the less explored features of constructionist theory and the internal structure of the constructicon is still largely uncharted territory.

In recent years, however, another kind of constructicon is emerging: a repository of construction descriptions, basically a dictionary of constructions. The idea was first introduced by Fillmore (2008), who initiated the development of an English constructicon (Fillmore, Lee-Goldman & Rhomieux, 2012; Lee-Goldman & Petruck, this volume) as a complement to the English FrameNet. This initiative

---

1. I am grateful to Steffen Höder, Jan-Ola Östman and two anonymous reviewers for valuable comments on earlier versions of this chapter.

has been followed by constructicon projects for Brazilian Portuguese (Torrent et al., this volume), German (Boas & Ziem, this volume), Japanese (Ohara, this volume), Russian (Janda et al., this volume), and Swedish (Lyngfelt, Bäckström et al., this volume). Thus, 'constructicon' now exhibits the same kind of polysemy as the related notions 'grammar' and 'lexicon': a theoretical notion of a linguistic system, on the one hand, and a corresponding descriptive resource, on the other. This volume is devoted to constructicons in the latter sense.

Practical constructicon development may be characterized as a blend between construction grammar and lexicography, which we label constructicography. The present volume is an introduction to constructicography in general and to the constructicon resources currently under development in particular. The bulk of the volume (Chapters 2–7) consists of language-particular presentations of each individual constructicon: English, Swedish, Brazilian Portuguese, Japanese, Russian, and German. In addition, there is a paper on the relation between linguistics and language technology in constructicon development (Borin, Dannells & Grūzītis, this volume) and one on multilingual constructicography, i.e., on connecting constructicons across languages (Lyngfelt, Torrent et al., this volume).

In this introductory chapter, I first give a brief introduction to constructions and construction grammar (Section 2), followed by a discussion of 'constructicon' as a theoretical conception (Section 3). Section 4 introduces frame semantics and FrameNet, which most of the constructicon projects presented in this volume are closely connected to, albeit in different ways. Section 5 addresses constructicography as such, discussing the conditions for constructicon development in light of the preceding sections, with a focus on conflicts between the traditions of lexicography and syntax. Finally, Section 6 briefly presents the chapters to come.

2.   Constructions and construction grammar

The central units of description in a constructicon are, naturally, constructions (cxns). These may be defined as "conventional, learned form-function pairings at varying levels of abstraction and complexity" (Goldberg, 2013, p. 17). Thus, they roughly correspond to signs, in the Saussurean sense. These pairings may be words (lexical cxns), idioms, phrasal patterns, clause types, conversational practices – basically any linguistic structure where a formal pattern is conventionally associated with a certain meaning or function.[2] Examples from English include general

---

2.   Whether morphemes are to be considered cxns is a matter of some debate, however, despite their being conventional pairings of form and meaning. According to Booij's (2010, 2013) influential approach to construction morphology, morphemes are not cxns in their own right but

structures such as noun phrase and imperative clause, argument structure configurations such as ditransitive and passive, so-called coerced structures such as count-to-mass (*There was cat all over the driveway*, e.g. Langacker, 2008, p. 144), information packaging structures such as clefts, idioms such as *pull someone's leg*, morphological categories such as compound and deverbal noun, and partially schematic structures such as the examples in (1):[3]

(1) a. [the X-er the Y-er]   (e.g. Fillmore, 1987; Culicover & Jackendoff, 1999)
        the more the merrier, the harder they come the harder they fall
     b. [What's X doing Y?]   (Kay & Fillmore, 1999)
        What's that fly doing in my soup?

Notably, a good number of cxns are not purely syntactic, lexical, or morphological patterns, but combine features from different "levels". For instance, the cxn in (1b) requires the word *what*, the specific word form *doing*, and a copular verb on the one hand, and an interrogative clause structure on the other; the variable X is a subject noun phrase, whereas Y is a locative adverbial; pragmatically, the whole cxn is an expression of incredulity, despite its interrogative syntax.

In constructionist models such as Construction Grammar (CxG), (the lexico-grammar of) a language is perceived as an inventory of cxns in the above sense. The same notion of cxn also forms the base for the constructicon (ccn) resources presented in this volume.

In a general dictionary, by contrast, 'construction' is typically defined as in (2):

(2) "the way in which words are put together in a sentence, phrase etc."
                                              (Longman English Dictionary Online)

This notion differs from most CxG definitions in at least two central aspects: On the one hand, it only concerns syntagmatic word combinations, disregarding both other linguistic structures and non-syntagmatic properties of the remaining configurations. On the other hand, it makes no distinction between conventionalized patterns and utterance-specific features; and thereby it applies equally well to so-called constructs, i.e. concrete instantiations of constructions.[4]

---

rather constituent parts of (morphological) cxns. For example, instead of treating the derivational suffix *-able* as a construction, Booij (2013, p. 256) views it as part of the cxn [Vtr -able], where Vᴛʀ represents the transitive verb with which *-able* is combined to form a derived adjective with a certain meaning.

3. Names of constructions, as well as frames (Section 4), are written in a sans serif font (Consolas).

4. In usage-based approaches to language (e.g. Bybee, 2010), according to which the status of linguistic patterns depends on their frequency, salience, etc., the distinction between construct and construction is a matter of degree.

There is also a more specific notion of construction, according to which constructions correspond to valence patterns, as in the OED definition in (3). This can be viewed as a special case of the Longman definition in (2), and is typical of how the term *construction* is employed in lexicography.

(3)   "The syntactical connection between verbs and their objects or complements, adjectives and their extensions, prepositions and objects, etc."

<div align="right">(Oxford English Dictionary)</div>

While both these notions of 'construction' are markedly different from the CxG concept, they are similar enough for the concepts to risk being confused with each other. This is all the more the case since the notions in (2) and (3) are not only characteristic of non-technical, everyday language, but they are also somewhat prevalent in (non-constructionist) linguistics. Therefore it is worth emphasizing that the sense of *construction* relevant here is: 'conventionalized pairing of form and meaning/function'.

From cxns, we now turn the focus to construction grammar (CxG), the theoretical base for constructicography. CxG was developed in the 1980s (Fillmore, Kay & O'Connor, 1988; Lakoff, 1987; and others), along with other constructionist approaches such as cognitive grammar (Langacker, 1987, 2008). It may be distinguished by the following five tenets, as summarized by Goldberg (2013):

1.   Grammatical constructions
2.   Surface structure
3.   A network of constructions
4.   Crosslinguistic variability and generalization
5.   Usage-based

The first four of these are shared by all constructionist approaches and the fifth by the majority.

Grammatical constructions: While assuming grammatical cxns as the primary units of investigation may seem obvious to the point of being trivial, this is actually not the case. First, not all linguistic theories acknowledge the existence of constructions; in some models (e.g. Chomsky, 2000) they are merely considered epiphenomenal results of the interaction between more abstract principles. Second, the assumption is not merely that there are cxns in language but that lexico-grammar in its entirety consists of cxns (plus relations between them and principles for combining them).

A core argument for this assumption is the intermingling of levels displayed in (1) (see also Michaelis, 2012, p. 56). Language abounds with such patterns combining lexical, morphological, pragmatic etc. properties, which would be harder to account for by a more modular approach. Furthermore, a methodological benefit

with treating abstract grammatical patterns, specific lexical items and anything in between as the same kind of linguistic unit is that they can all be represented with the same kind of descriptive machinery (Fillmore, Kay & O'Connor, 1988, p. 534).

Surface structure: Cxns are essentially surface generalizations, associating meaning directly with (surface) form, in explicit opposition to more derivational approaches (such as the Minimalist Program; Chomsky, 1995).

A network of constructions: The cxns of a language are not an unordered set but presumably organized in a network: a constructicon. This idea, which is the foundation behind the work presented in this volume, will be introduced in more detail in the next section.

Crosslinguistic variability and generalization: While the CxG literature exhibits the same bias towards widespread and culturally dominant languages as linguistics in general does, there is also a strong emphasis on crosslinguistic variability (e.g. Croft, 2001; Fried & Östman, 2004; Boas, 2010). Cxns are to a large extent language specific, as can be expected given that they are conventions, presumably acquired through generalization over instances. Hence, instead of assuming and trying to categorize a universal category of, say, passive, the corresponding CxG approach would be to compare "passive" cxns in different languages to discern what similarities and differences they exhibit.[5] While cross-linguistic generalizations are indeed a desirable research goal, attested similarities are typically attributed to language contact, similar functional motivation, etc., rather than innate universals or the like.

From the viewpoint of constructicography, this issue plays out somewhat differently. In this context, the main purpose of contrastive work is towards the development of bi- or multilingual constructicon applications, i.e. connecting constructicon resources across languages. Thus, a central objective is to establish the closest corresponding cxns (or constructicon entries) for different languages (cf. Bäckström, Lyngfelt & Sköldberg, 2014; Laviola, 2015), much in the same manner as in multilingual lexicography. The development of multilingual constructicography is treated in Lyngfelt, Torrent et al. (this volume).

Usage-Based: While constructionist approaches in general are usage-based in the pre-theoretical sense of being empirically grounded, not all conform to usage-based theory, i.e. the view that linguistic patterns are gradually established by generalization over instances, based on e.g. frequency and salience (Langacker, 1987; Tomasello, 2003; Bybee, 2010; and others). This view of gradual establishment of a cxn is most often discussed in terms of entrenchment in the mind of the individual language user, but also applies to its corresponding conventionalization in the speech community (cf. Schmid, 2015). It is rather the latter perspective that

---

5.   In this context, the label "passive" does not hold any independent theoretical significance, but mainly serves as a convenient point of reference (cf. Croft, 2001, Chapter 1).

concerns constructicography, although the process of conventionalization as such is outside the scope of investigation. The relevant issues are instead, on the one hand, how well established (in whatever sense) a cxn has to be to warrant representation, and, on the other hand, how to deal with constructional variation.

3.   Constructicon: language as a network of constructions

The notion of 'constructicon' may be traced back to the following quote by Fillmore:

> The grammar of a language can be seen as a repertory of constructions, plus a set of principles which govern the nesting and superimposition of constructions into or upon one another.   (Fillmore, 1988, p. 37)

This "repertory" has subsequently been dubbed a *constructicon* (Jurafsky, 1991, p. 18)[6] and has generally come to be pictured as a network. Although this view of lexicogrammar as a network of constructions is generally assumed in construction-ist literature, the notion as such is vastly understudied. Most work in CxG consists of case studies of individual cxns or small groups of constructions, and the overall structure of the constructicon as such is largely left unexplored.

The second part of the Fillmore quote, how cxns (or rather constructs) are combined, is even less investigated. The standard position is to assume combina-tion by unification (e.g. Fillmore & Kay, 1996; Goldberg, 2006; Sag, 2012), but this simple assumption is clearly insufficient to account for coercion phenomena and other complexities regarding how (by presumption) cxns license constructs which are in turn combined into utterances. Furthermore, as becomes clearly evident once one starts to build a larger constructicon, covering more diverse types of cxns, the network aspect and the combinatory aspect are tightly connected; to be able to account for one of them, one also has to address the other. For example, since constructs licensed by cxns such as ditransitives, passives, clefts, polarity questions and subject-predicate may be combined to form a clefted passive ditransitive inter-rogative sentence (as in *Was it Tiago who was appointed head of department?*), the corresponding cxn descriptions all have to be compatible. This, in turn, depends on their being designed with such combinatory possibilities taken into account.

A somewhat more modest take would be not to assume a single network of cxns, but rather a set of networks (cf. Langacker, 2008, pp. 237ff.). Families of re-lated cxns may thus be treated as smaller networks without direct reliance on a global network. To the extent that actual cxn networks have been presented in the CxG literature so far, they are of this smaller, relatively homogenous kind, although

---

6.   Sometimes, the term is written *construct-i-con* (e.g. Goldberg, 2003; Hilpert, 2014).

they are usually presented under the presumption of a general, superordinate network (a couple of the more ambitious examples are Goldberg, 1995, on argument structure and Sag, 2010, on filler-gap cxns). It should be noted, however, that even under the assumption that there is no global network but only a number of smaller networks, the idea of a constructicon still implies a need to eventually account for how they interact.

Global or not, the set or sets that make up the constructicon are typically conceived of as inheritance networks, with more specific cxns instantiating more abstract ones from which they inherit properties. Exploiting the family metaphor here, the inheriting cxn will be called the *child* in this relation and the superordinate cxn from which it inherits will be called the *parent*. CxG models such as Berkeley Construction Grammar (Fillmore & Kay, 1996) and Sign-Based Construction Grammar (Sag, 2012) strictly assume full inheritance, i.e. that the child inherits all the parent's properties. Thus, the inheritance network amounts to a taxonomic hierarchy. Some versions of cognitive CxG (e.g. Goldberg, 1995, 2013), on the other hand, rather employ default (also called *normal*) inheritance, which means that all properties are inherited *unless specified otherwise*. On this view, the child is not only a more specific instance of the parent construction, but may also deviate from it in one or more respects.[7]

The standard inheritance relation is instantiation: the child is a more specific variant of the parent cxn (on the full inheritance model restricted to a proper subset of instances). In addition, Goldberg (1995) assumes polysemy links and metaphorical links, which are particular cases of default inheritance; and subpart links,[8] also called horizontal links (van der Velde, 2012). The latter are horizontal relations between cxns with certain properties in common, for example presentational and existential cxns being related by their both containing expletive subjects. As an illustration, Hilpert (2014, pp. 62ff.) brings up the many properties shared by the (mono-) `transitive` and the `ditransitive` cxn (in English): both contain an agentive subject and a Patient or Theme as direct object, both instantiate `noun phrase` and `verb phrase` cxns etc.

Subpart links may be characterized in terms of multiple inheritance, i.e. when a cxn (or parts of it) inherits properties from several different sources. This is one reason why subpart links have been presented as a form of inheritance despite the fact that the cxns thus linked do not necessarily constitute a parent-child relation.

---

7.   A related, although secondary, issue is that of complete inheritance vs. redundant representations, i.e. whether inherited properties are represented on all levels where they apply or only on the topmost node (see e.g. Hilpert, 2014, pp. 65ff.).

8.   Note that subpart links are not to be confused with the meronymic relations between a construction/construct and its constituent parts.

Multiple inheritance also applies to inheritance through several "generations", where a cxn not only instantiates its parent cxn but its grandparent etc. as well (for example, the fixed cxn *Don't worry* is an instance of the *Don't X* cxn, which in turn instantiates the `imperative clause` cxn, the `main clause` cxn, etc.; cf. Croft & Cruse, 2004, p. 321).

Such multi-level instantiation has been proposed as a driving force behind the development of the constructicon, especially in usage-based versions of CxG. Cxns are presumably formed by generalization over instances, more specific cxns being acquired first and more general ones developing later on, in a multigranular system continuously being shaped by usage. A low type/token ratio (i.e. many instances per specific type) favors the entrenchment of lower-level generalizations, whereas a high ratio favors the development and strengthening of higher levels (e.g. Hilpert, 2013, 2015). The interrelations between these various generalizations are what fundamentally constitutes the network. How nodes and relations in the network may be established and gradually develop further is being modeled computationally in Fluid CxG (Steels, 2011; van Trijp & Steels, 2012; Steels, 2013) and, from a different perspective, in Embodied CxG (Feldman, Dodge & Bryant, 2010; Bergen & Chang, 2013).

So far, empirical studies exploring these ideas are typically small-scale case studies; the development of larger and more complex networks, not to mention their resulting internal structure, mostly remains uncharted territory, at least from a descriptive point of view (computational modelling having advanced somewhat further). In response to this situation, the constructicon (ccn) resources under development should provide excellent material for the study of more diverse and complex cxn networks. Although nowhere near full coverage, they consist of relatively large and heterogeneous sets of cxn descriptions. As of yet, the network aspect of the resources is underdeveloped, with only smaller sub-networks established and the overall structure largely resembling a set of listemes. Nevertheless, work on developing the internal network structure is indeed in progress, and the chapters in this volume point out the directions where this work is heading.

A central point of consideration is the role of the lexicon. In general, constructionist theories reject the traditional distinction between grammar and lexicon, instead assuming a continuum of cxns. The main motivation for rejecting a modular view of language is the existence – and wealth – of patterns combining lexical and grammatical properties, and the gradual nature of the differences between general and specific structures; most linguistic generalizations are neither fully general ("grammar") nor entirely specific ("lexicon") but located somewhere on a scale between these two points. Note that this theoretical tenet does not preclude a distinction between lexical and phrasal cxns. However, such a distinction is a sorting device within the system, categorizing units according to one differing property

(cf. schematic/substantive, atomic/complex, etc.); it does not construe them as fundamentally different kinds of units or assign them to separate cognitive modules.[9]

Consequently, the lexicon is a subset of the constructicon, namely the subset of lexical cxns, at least in principle. For practical reasons, however, all the ccn resources treated in this volume do distinguish between lexicon and constructicon, albeit in slightly different ways. The main reason for this is that each ccn is developed in relation to one or more lexical resources, importing lexical information from there instead of reinventing the wheel. In particular, most of the ccns are more or less closely related to framenets. Hence, the next section will provide a brief introduction to frame semantics and FrameNet and address their relation to CxG and constructicons (more detailed presentations of FrameNet are given in some of the following chapters, e.g., Lee-Goldman & Petruck, this volume).

4.   Frame semantics and FrameNet

The core idea of frame semantics (e.g. Fillmore, 1982), and consequently FrameNet, is that words are understood in relation to the scenarios – *frames* – in which they occur, and also in relation to other participants in the same frame. A word like *husband* can only be properly understood in relation to concepts such as 'marriage' and 'wife', and vice versa; a verb like *buy* involves a buyer, a seller, some merchandise being transferred from the seller to the buyer, and some means of payment being transferred in the opposite direction, etc. As the organizing scenarios are called *frames*, the participants are called *frame elements*.

Thus, frame semantics is somewhat related to valence, but it is less strictly associated with selecting properties of the head. More importantly, it is also less strictly lexical in that several different words may evoke the same frame. For instance, verbs like *avoid, dodge, escape*, as well as nouns like *evasion* may all be said to evoke the Avoiding frame and to involve the same basic frame elements, even if their respective lexical valence patterns are not identical.

Frames also relate to constructions, in a number of ways. Not only words (lexical cxns) but also many phrasal cxns may be said to evoke frames.[10] Thus, frame semantics is sometimes employed for semantic aspects of CxG analyses. A benefit in this regard is how the multigranular character of frame semantics offers

---

9.   Interestingly, Pulvermüller, Cappelle & Shtyrov (2013) find partial – but only partial – neurolinguistic support for the lexico-grammar continuum. There seems to be something special about words, after all.

10.   For a discussion of relations between cxns and frames, which cxns evoke frames and which do not, etc., see Ohara (this volume); Lyngfelt, Bäckström et al. (this volume).

compatibility with cxns of differing generality. In relation to argument structure, for example, a topic that is extensively discussed in the CxG literature (cf. Goldberg, 1995; Croft, 2003; Boas, 2009a; and others), very general frames correspond to Goldberg-style argument structure cxns, more specific frames to verb-class specific cxns, and verb sense particular valence cxns (also called mini-constructions) to associations between FrameNet's lexical units (pairings of a word and a frame) and certain grammatical configurations.[11]

FrameNet, in turn, is a lexicographic application of frame semantics, in the form of a lexical database of English. At the time of writing, it covers a little over 1,200 frames, presented with corresponding lexical units and sets of annotated sentences (the *FrameNet* website;[12] cf. also, e.g., Fillmore & Baker, 2010; Ruppenhofer et al., 2016). This resource has inspired the development of similar framenets for a large number of languages (see the papers in Boas, 2009b). It is also one of the major sources of inspiration for constructicon resources.

The first constructicon (ccn) project, the one for English, was initiated as an addition, or a complement, to the English FrameNet (Fillmore, 2008; Fillmore, Lee-Goldman & Rhomieux, 2012; Lee-Goldman & Petruck, this volume). The ccn enterprises that followed are all inspired by this initiative and thus, either directly or indirectly, influenced by FrameNet methodology. It should therefore be stressed that the notion of constructicon as such does not depend on FrameNet or frame semantics; the actual connection is essentially a consequence of historic circumstances, in particular of the fact that Charles Fillmore and associates were at the core of both developments.

That said, there are strong connections in actual practice. The English, the Brazilian Portuguese (Torrent et al., this volume), and the Japanese (Ohara, this volume) ccns are all developed as additions to the respective framenets. In principle, although not yet fully developed, framenet is taken to be the superordinate concept, including an FN lexicon and an FN constructicon. The Swedish and Russian ccns, on the other hand, are more or less independent projects. In the case of Russian (Janda et al., this volume) there is no corresponding framenet resource, so any connection is only indirect. For Swedish (Lyngfelt, Bäckström et al., this volume), the ccn and the framenet are interrelated and share some of the same infrastructure but remain essentially independent projects and resources.

---

11. One may also discern more indirect correspondences between constructions and frames. For example, what is the relation between a frame like Request and an imperative clause cxn? It is clearly different from the relation between the frame and typically evoking verbs such as *ask, command, instruct, tell*, etc., which all are used to *refer* to requests whereas an imperative is a way of *performing* one (cf. Lyngfelt, Bäckström et al., this volume, Section 5).

12. <https://framenet.icsi.berkeley.edu/fndrupal/>

The historical relation between the frame-based and cxn-based developments is illustrated in Figure 1 – which may also be seen as an overview of the legacy of the late Charles J. Fillmore.
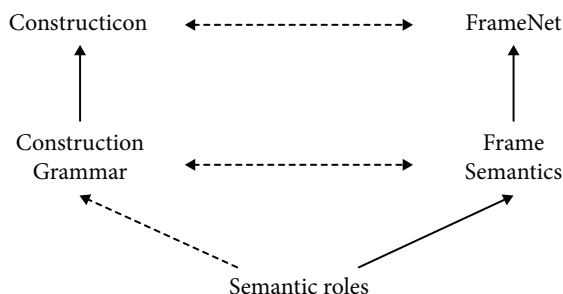
Figure 1. The historical and conceptual connections between construction-based and frame-based developments

As indicated in Figure 1, the connection between cxns and frames can be traced all the way back to Fillmore's (1968) early work on semantic roles, which is a constituting feature of frame semantics and also a central concept in the development of CxG. The two theoretical traditions have then maintained contact. Not only were CxG and frame semantics originally developed at the same place, Berkeley, and to some extent by the same people, frame semantics has often been employed to represent semantic features of CxG analyses (as mentioned above). Similarly, in at least some of the resultant practical applications, ccns and framenets have been connected by links between cxns and frames, where applicable. To what extent and in what ways the two kinds of resources are interrelated in practice varies between the different projects and will be addressed, where relevant, in the chapters to come and more briefly in the overview at the end of this chapter.

5.   Constructicography: construction grammar meets lexicography

As mentioned in the introduction of this introductory chapter, constructicon development – *constructicography* – may be characterized as a combination of construction grammar (CxG) and lexicography. These two traditions overlap in some aspects, such as in their manner of accounting for morphology, but are markedly different in others. Arguably the main challenge is to accommodate the practices of lexicography and syntax, where the most fundamental contrast has only partially to do with dissimilarities between words and larger units. In the following, some degree of stereotyping will be performed in order to highlight the overall differences

between the two traditions. While I am well aware that the fields contain more variation than this rough comparison will make justice to, the issues at stake are quite general concerns across the various approaches.

The way lexicography typically presents language is in the form of listed and/or searchable entries, preferably in a simple and accessible way. The main focus is to account for conventional senses of words, rather than trying to cover all possible uses; "meanings and dictionary senses aren't the same thing at all" (Atkins & Rundell, 2014, p. 311). Although many words arguably have one discrete meaning, others are vague and/or polysemous, and even the discrete ones lend themselves to figurative or other creative uses. Hence, dictionary definitions are idealized generalizations essentially corresponding to prototype descriptions. Precision is of course a desirable goal and a main priority, but peripheral or borderline cases are rarely a concern. (cf. Atkins & Rundell, 2014; Svensén, 2009)

Syntax, on the other hand, is typically concerned with other kinds of idealized generalizations. A central objective is to establish a finite, preferably small, set of general rules to account for an infinite – but still definable – set of grammatical expressions (cf. the famous quote "generating all and only the [grammatical] sentences of a language", Chomsky, 1957: 85). Grammaticality is usually taken to be an in principle binary property, and syntactic categories are treated as discrete entities. Consequently, it is a high priority to delimit categories, account for apparent exceptions, etc., which tends to somewhat direct focus towards untypical cases. It also makes for quite detailed analyses and a high degree of interdependence between definitions.

There is, however, an alternative, *usage-based* approach (e.g. Langacker, 1987; Bybee, 2010), according to which grammar is dynamic – constantly shaped and reshaped by usage. This perspective, which is actually predominant in CxG (see Section 2 above), is more concerned with typicality and prototypes. It is also less reliant on general rules, acknowledging the need for listing particularities of linguistic knowledge, in a multigranular, non-reductionist system where higher-level generalizations are built up from lower-level ones. Thus, a usage-based approach clashes less sharply with lexicographic practices than other models of syntax do. However, along with the interest for usage patterns also follows an interest in variation and an urge to account for both the typical and the untypical, as well as for what motivates the distribution between the two. This leads to detailed analyses far from the lexicographic ideal of concise, accessible definitions.

Furthermore, whatever the theoretical approach, syntax is concerned with relations such as constituency, dependency, and linear order, all of which require other kinds of representation than typically employed in lexicography (although valence descriptions in some dictionaries are clearly a step in this direction, more powerful machinery is required to handle more complex syntactic relations). These

representations make use of an abstract and relatively formalized meta-language, more or less explicitly presupposing discrete distinctions – partly due to the nature of the object of study, but reinforced by tradition. This constitutes yet another fundamental difference from the (prose) prototype descriptions typical of lexicography. While prose definitions may express variability and nuances by modifications such as *typically, especially, marginally, in certain contexts*, etc., grammatical representations tend to be more absolute. There are tools for expressing optionality (usually by parentheses), alternatives (/) and variables, etc., but even these notations are categorical rather than gradable (possibly in contrast to both the reality behind them and the considerations of the analyst).

A constructicon may be roughly referred to as "a dictionary of grammatical constructions". This is because the purposes and usage conditions, and hence the demands on the description format, correlate strongly with those of lexicography. Typical dictionary entries are approximate accounts of ordinary usage, presented in a list or other searchable format. They are also subject to a preference for simple, user-friendly descriptions, not too dependent on technical meta-language. Arguably, then, the same, or at least similar, aims also apply to ccn entries. The challenge is thus to develop a description format to accommodate grammatical, notably syntactic, representations with such conditions. In this regard, some of the crucial issues are:

– accuracy vs. user-friendliness/simplicity
– variability
– granularity
– degree of formalization

We will not go into detail here about how these and other related issues may be resolved. It should be noted, however, that considerations about either of them depend on the purpose of the resource. One of the more fundamental differences is that computational applications require formalized descriptions whereas human users prefer readable ones. Furthermore, any attempt at cross-language application places even higher demands on the meta-language. While grammatical categories are somewhat blunt sorting devices even within a single language, their application across languages is highly problematic (e.g. Croft, 2001, Chapter 1). Hence, attempts to connect ccns for different languages require a meta-language capable of handling dissimilarities between the languages in question. (cf. Lyngfelt, Torrent et al., this volume; Boas & Ziem, this volume)

Finally, ccns may vary greatly regarding the extent and manner they represent relations between cxns, as well as relations between ccns and external resources and applications. As long as a ccn is simply a list of cxn descriptions, the representation of relations between them can be kept to a minimum, but any ambitions to develop

the resource in the direction of a cxn network turn relational aspects into a central concern. Also, the degree of coverage itself clearly correlates with the need to work out the relations between cxns in order to develop a consistent system. Turning to external links, the prime example is, as mentioned above, relations between cxns and frames – or to other lexical resources than a framenet. Regarding other kinds of adaptations, the ccns treated in this volume are mostly designed to be more or less multi-purpose resources, open to different types of application and therefore not strictly tailored to suit any particular one. Nevertheless, they differ somewhat in this regard, at least in terms of foci and priorities; for example, for the ccns of Russian (Janda et al., this volume) and Swedish (Lyngfelt, Bäckström et al., this volume), relevance for language learners is a main concern.

How these and other issues are handled in the various ccn projects will be addressed, where relevant, in each respective chapter.

6.   The chapters in this volume

The present volume consists of nine chapters, including this introduction. There are six language-specific chapters, introducing the English, Swedish, Brazilian Portuguese, Japanese, Russian, and German constructicons, respectively. In addition, there is one chapter addressing the relation between linguistics and language technology, and one devoted to multilingual ccn development.

Chapter 2, *The FrameNet constructicon in action* (Lee-Goldman & Petruck), presents the English ccn, in itself and in relation to the English FrameNet. The chapter offers an introduction to FrameNet, followed by a corresponding introduction to the ccn, the latter focusing on concepts and terminology. The methodology is illustrated by an analysis of the `be_recip`(rocal) cxn (as in *Paul is friends with Chuck*). After sections on annotation and relations between frames and cxns, the chapter is concluded by stating the high relevance of a ccn for language technology.

Chapter 3, *Constructicography at work: Theory meets practice in the Swedish constructicon* (Lyngfelt, Bäckström et al.), is a long chapter about the Swedish ccn. It is first placed in its local context of *Språkbanken* (the Swedish language bank), an infrastructure of linguistic resources with which the ccn is integrated. An exposition of the workflow follows, from selection procedures to the organization of cxn entries into types. After raising issues of idealization and the treatment of constructional variation, the chapter turns to relations between cxns and frames, with a discussion about linking (a subset of) them to each other. The following sections go through the description format and address the ccn from the user's perspective, before wrapping up with a discussion of the ongoing development from a cxn dictionary to a cxn network.

Chapter 4, *Towards continuity between the lexicon and the constructicon in Frame-Net Brasil* (Torrent et al.), presents the Brazilian Portuguese ccn, focusing on how a frame-based lexicon and ccn are integrated in a common database. After introducing the lexicon and ccn of FrameNet Brasil, the chapter details how the integration of the two constituent parts is modeled computationally. This model is then illustrated by sample analyses of `dative with infinitive` and `inceptive aspect` cxns.

Chapter 5, *Relations between frames and constructions: A proposal from the Japanese FrameNet constructicon* (Ohara), also addresses relations between cxns and frames, this time from the viewpoint of the Japanese FrameNet project and with a main focus on conceptual relations and annotation practices. Central topics are why ccns are needed in addition to framenets, fundamental differences between frame annotation and cxn annotation, and the distinction between frame-evoking ("meaning-bearing") and non frame-evoking cxns. The chapter arrives at a five-way classification of cxns: two types of cxns that evoke frames (either semantic or inter-actional frames) and three that do not.

Chapter 6, *A constructicon for Russian: Filling in the gaps* (Janda et al.), approaches ccn development for Russian from a CxG/cognitive linguistics perspective. Unlike the other ccn projects, it has no relation to FrameNet, instead starting out from a general CxG perspective and the lack of coverage of certain types of Russian cxn patterns. After an overview of previous constructionist and related work on Russian, the chapter addresses what types of cxns are the first priority (in the initial stages) of the project, due to lack of coverage and relevance for learners. This is followed by a presentation of the Russian ccn, along with sample cxn analyses, before the chapter is concluded with an outlook at future research and applications for which a Russian ccn would be useful.

Chapter 7, *Constructing a constructicon for German: Empirical, theoretical, and methodological issues* (Boas & Ziem), presents the fairly recently started development of a German ccn. The chapter provides an overview of some characteristic phenomena in the German language, followed by a principle discussion of contrastive issues. Based on this, the authors address, through discussion of a few sample cxns, the possibility of basing German ccn entries on existing English ones, concluding that a language-specific corpus-based approach is preferable. Finally, before rounding up, the chapter introduces the annotation-driven methodology of the German ccn project.

Chapter 8, *Linguistics vs. language technology in constructicon building and use* (Borin, Dannélls & Grūzītis), is concerned with the relation and (often lack of) interaction between linguistics and language technology. After discussing the currently prevalent disassociation between linguistics and language technology, its causes and consequences, the chapter describes the close collaboration between the two branches in the development of the Swedish ccn. It shows how the cxn

analyses both draw upon and simultaneously enrich other resources in an integrated macro-resource, both for general development and in particular subprojects such as using LT tools for identifying potential cxn entries. Then follows descriptions of how information from the ccn is put to use in language technology in various ways, before a concluding wish for continued and deepened collaboration between the disciplines in the future.

Chapter 9, *Aligning constructicons across languages: A trilingual comparison between English, Swedish, and Brazilian Portuguese* (Lyngfelt, Torrent et al.), addresses prospects for connecting ccn resources across languages. After briefly reviewing work in contrastive CxG and multilingual lexicography, the chapter presents a comparison between the entries in the English ccn and corresponding cxns in Brazilian Portuguese and Swedish. In light of this comparison follows a discussion of possibilities and problems for bi- and multilingual ccn development, regarding both LT applications and human users.

References

Atkins, B. T. S., & Rundell, M. (2014). *The Oxford Guide to Practical Lexicography*. Oxford & New York: Oxford University Press.

Bäckström, L., Lyngfelt, B., & Sköldberg, E. (2014). Towards Interlingual Constructicography. On correspondence between constructicon resources for English and Swedish. *Constructions and Frames*, 6:1, 9–32. doi:10.1075/cf.6.1.02bak

Bergen, B., & Chang, N. (2013). Embodied Construction Grammar. In Th. Hoffmann & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 168–190). Oxford: Oxford University Press.

Boas, H. C. (2009a). Verb meanings at the crossroads between higher-level and lower-level constructions. *Lingua*, 120, 22–34. doi:10.1016/j.lingua.2009.03.008

Boas, H. C. (Ed.) (2009b). *Multilingual FrameNets in Computational Lexicography: Methods and Applications*. Berlin & New York: Mouton de Gruyter. doi:10.1515/9783110212976

Boas, H. C. (Ed.) (2010). *Contrastive Studies in Construction Grammar*. Amsterdam: John Benjamins. doi:10.1075/cal.10

Booij, G. (2010). *Construction Morphology*. Oxford: Oxford University Press.

Booij, G. (2013). Morphology in Construction Grammar. In Th. Hoffmann & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 255–273). Oxford: Oxford University Press.

Bybee, J. (2010). *Language, Usage, and Cognition*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511750526

Chomsky, N. (1957). *Syntactic Structures*. The Hague: Mouton.

Chomsky, N. (1995). *The Minimalist Program*. Cambridge, MA: MIT Press.

Chomsky, N. (2000). *New Horizons in the Study of Language and Mind*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511811937

Croft, W. (2001). *Radical Construction Grammar. Syntactic theory in typological perspective*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780198299554.001.0001

Croft, W. (2003). Lexical rules vs. constructions: a false dichotomy. In H. Cuyckens, Th. Berg, R. Dirven & K. -U. Panther (Eds.), *Motivation in Language: Studies in honor of Günther Radden* (pp. 49–68). Amsterdam: John Benjamins. doi:10.1075/cilt.243.07cro

Croft, W., & Cruse, D. A. (2004). *Cognitive Linguistics*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511803864

Culicover, P., & Jackendoff, R. (1999). The view from the periphery: The English comparative correlative. *Linguistic Inquiry*, 30, 543–571. doi:10.1162/002438999554200

Feldman, J., Dodge, E., & Bryant, J. (2010). Embodied Construction Grammar. In B. Heine & H. Narrog (Eds.), *The Oxford Handbook of Linguistic Analysis* (pp. 111–138). Oxford: Oxford University Press.

Fillmore, C. J. (1968). The Case for Case. In E. Bach & R. T. Harms (Eds.), *Universals in Linguistic Theory* (pp. 1–88). London: Holt, Rinehart and Winston.

Fillmore, C. J. (1982). Frame Semantics. In Linguistics Society of Korea (Ed.), *Linguistics in the morning calm* (pp. 111–138). Seoul: Hanshin Publishing Co.

Fillmore, C. J. (1987). Varieties of conditional sentences. *ESCOL*, 3, 79–122.

Fillmore, C. J. (1988). The mechanisms of 'Construction Grammar'. *Berkeley Linguistic Society*, 14, 35–55.

Fillmore, C. J. (2008). Border Conflicts: FrameNet Meets Construction Grammar. In E. Bernal, & J. DeCesaris (Eds.), *Proceedings of the XIII EURALEX International Congress* (pp. 49–68). Barcelona: Universitat Pompeu Fabra.

Fillmore, C. J., & Baker, C. F. (2010). A Frames Approach to Semantic Analysis. In B. Heine & H. Narrog (Eds.), *The Oxford Handbook of Linguistic Analysis* (pp. 313–339). Oxford: Oxford University Press.

Fillmore, C. J., & Kay, P. (1996). Construction Grammar Coursebook. Ms. Department of linguistics, University of California at Berkeley.

Fillmore, C. J., Kay, P. & O'Connor, M. C. (1988). Regularity and idiomaticity in grammatical constructions. The case of *let alone. Language*, 64, 501–538. doi:10.2307/414531

Fillmore, C. J., Lee-Goldman, R., & Rhomieux, R. (2012). The FrameNet Constructicon. In H. C. Boas & I. A. Sag (Eds.), *Sign-Based Construction Grammar* (pp. 309–372). Stanford: CSLI Publications.

FrameNet. <https://framenet.icsi.berkeley.edu/fndrupal/>

Fried, M., & Östman, J. -O. (Eds.) (2004). *Construction Grammar in a cross-language perspective*. Amsterdam: John Benjamins. doi:10.1075/cal.2

Goldberg, A. E. (1995). *Constructions. A Construction Grammar Approach to Argument Structure*. Chicago: The University of Chicago Press.

Goldberg, A. E. (2003). Constructions: A new theoretical approach to language. *Trends in Cognitive Science*, 7, 219–224. doi:10.1016/S1364-6613(03)00080-9

Goldberg, A. E. (2006). *Constructions at Work. The Nature of Generalizations in Language*. Oxford: Oxford University Press.

Goldberg, A. E. (2013). Constructionist approaches. In Th. Hoffmann & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 15–31). Oxford: Oxford University Press.

Hilpert, M. (2013). *Constructional Change in English: Developments in Allomorphy, Word Formation, and Syntax*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139004206

Hilpert, M. (2014). *Construction Grammar and its Application to English*. Edinburgh: Edinburgh University Press.

Hilpert, M. (2015). From *hand-carved* to *computer-based*: Noun-participle compounding and the upward strengthening hypothesis. *Cognitive Linguistics*, 26, 113–147. doi:10.1515/cog-2014-0001

Jurafsky, D. (1991). *An On-line Computational Model of Human Sentence Interpretation: A Theory of the Representation and Use of Linguistic Knowledge*. Doctoral dissertation. University of California, Berkeley.

Kay, P., & Fillmore, C. J. (1999). Grammatical constructions and linguistic generalizations. The *What's X doing Y?* construction. *Language*, 75, 1–34. doi:10.2307/417472

Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What categories reveal about the mind*. Chicago: University of Chicago Press. doi:10.7208/chicago/9780226471013.001.0001

Langacker, R. W. (1987). *Foundations of Cognitive Grammar*, vol. 1. *Theoretical prerequisites*. Stanford: Stanford University Press.

Langacker, R. W. (2008). *Cognitive Grammar. A basic introduction*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780195331967.001.0001

Laviola, A. (2015). Frames e Construções em Contraste: uma análise comparativa português-inglês no tangente à implementação de constructicons. MA Thesis, Federal University of Juiz de Fora.

Lyngfelt, B., & Sköldberg, E. (2013). Lexikon och konstruktikon – ett konstruktionsgrammatiskt perspektiv på lexikografi ['Lexicon and constructicon – a constructionist perspective on lexicography']. *LexicoNordica*, 20, 75–91.

Michaelis, L. A. (2012). Making the case for Construction Grammar. H. C. Boas & I. A. Sag (Eds.) (2012), *Sign-Based Construction Grammar* (pp. 31–67). Stanford: CSLI Publications.

Pulvermüller, F., Cappelle, B., & Shtyrov, Y. (2013). Brain Basis of Meaning, Words, Constructions, and Grammar. In Th. Hoffmann & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 397–416). Oxford: Oxford University Press.

Ruppenhofer, J., Ellsworth, M., Petruck, M. R. L., Johnson, C. R., & Scheffczyk, J. (2016). *FrameNet II: extended theory and practice*. Berkeley: ICSI. Retrieved from https://framenet2.icsi.berkeley.edu/docs/r1.7/book.pdf

Sag, I. A. (2010). English filler-gap constructions. *Language*, 86, 486–545. doi:10.1353/lan.2010.0002

Sag, I. A. (2012). Sign-Based Construction Grammar: An Informal Synopsis. H. C. Boas & I. A. Sag (Eds.) (2012), *Sign-Based Construction Grammar* (pp. 69–202). Stanford: CSLI Publications.

Schmid, H. -J. (2015). A blueprint of the Entrenchment-and-Conventionalization Model. *Yearbook of the German Cognitive Linguistics Association*, 3, 1–27. doi:10.1515/gcla-2015-0002

Steels, L. (2013). Fluid Construction Grammar. In Th. Hoffmann & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 153–167). Oxford: Oxford University Press.

Steels, L. (Ed.) (2011). *Design Patterns in Fluid Construction Grammar*. Amsterdam: John Benjamins. doi:10.1075/cal.11

Svensén, B. (2009). *A Handbook of Lexicography. The Theory and Practice of Dictionary-Making*. Cambridge & New York: Cambridge University Press.

Tomasello, M. (2003). *Constructing a Language: A Usage-Based Theory of Language Acquisition*. Cambridge, MA: Harvard University Press.

van Trijp, R., & Steels, L. (2012). Multilevel alignment maintains language systematicity. *Advances in Complex Systems*, 15(3–4), 1–30.

Van de Velde, F. (2012). Degeneracy: The maintenance of constructional networks. *Leuven Working Papers in Linguistics*, 1, 24–61.

The FrameNet constructicon in action

Russell Lee-Goldman and Miriam R.L. Petruck

The present work provides an overview of the FrameNet Constructicon both as an addition to the FrameNet knowledge base and as proof of concept for theoretical and practical constructicon development. The paper begins with a brief discussion of the fundamental principles of Construction Grammar, which constitutes the theoretical basis of any constructicon development project. Introducing FrameNet Constructicon terminology and illustrating its usage with examples from a variety of constructions, the paper offers an in-depth exploration of the *be*_recip construction (as in *Paul is friends with Chuck*), also demonstrating tools and techniques for analyzing the construction as well as the automatically produced reports that derive from the manual annotation of instances of the construction. The paper concludes by discussing the utility of construction documentation and annotation for language technology.

Keywords: Frame Semantics, FrameNet, Construction Grammar, FrameNet Constructicon, frame, frame element, frame-to-frame relations, grammatical construction, construct, construction-to-construction relations

1. Introduction

A lexicographic approach to linguistic meaning prioritizes the association between semantics and specific words, and FrameNet (framenet.icsi.berkeley.edu) is one such approach, grouping and describing words according to the common scenes or situations that the words describe. However, FrameNet encountered units of language that convey meaning but were not, or not entirely, lexical (Fillmore, Lee-Goldman, & Rhomieux, 2012). Multiword expressions, complex idioms, and even schematic syntactic patterns convey rich meanings. This observation is fundamental to Construction Grammar: a lexicon alone will not capture all of the meaningful units of language.

A Constructicon, i.e. a record or repository of grammatical constructions, is necessary. Ideally, a constructicon would be developed in parallel to a frame-based

lexicon, with connections among and between constructions, lexical items, and semantic frames. This paper describes the Beyond the Core (BTC) project, an attempt to augment the FrameNet lexicon with constructional information. As a pilot study, it aimed primarily to demonstrate the practical and theoretical feasibility of building a constructional repository alongside FrameNet, with common tools and methods. BTC and FrameNet are yet to be fully integrated; in particular, the explicit connections between frames and constructions must be represented. Though this essential element is missing, the FrameNet Constructicon nonetheless illustrates a working process for constructicon development.

We have structured the rest of this paper as follows. Section 2 describes Frame-Net, thus also situating the development of the FrameNet Constructicon in its larger theoretical and technological context. Section 3 introduces the terminology of the FrameNet Constructicon, necessarily also defining the terminology of Construction Grammar. Section 4 discusses *construction-hood*, and explores the use of the `be_recip` construction, as in the sentence *Paul is good friends with Chuck*. Section 5 analyzes the construction; and Section 6 shows annotation for examples of the `be_recip` construction, and several other constructions. Section 7 provides an overview of the parallels between the conceptual apparatus of the FN lexicon and that of the FN constructicon. Lastly, Section 8 summarizes the work presented in this paper and discusses the utility of constructional information for language technology.

2.   FrameNet background

Based on the principles of Frame Semantics (e.g. Fillmore, 1985) and determined to demonstrate its instantiation in a computational environment, FrameNet began as a corpus-based computational lexicography research project, with the goal of providing valence descriptions, or combinatorial possibilities, of each item analyzed for the general vocabulary of contemporary English. As indicated above, the FN Constructicon came into being as an add-on to the existing FN database. In this section, we introduce the basic concepts of Frame Semantics and describe how FrameNet instantiates those concepts.

A FrameNet *frame* is a schematic representation of a situation involving various participants and other conceptual roles, each of which is a *frame element*. A *lexical unit* (LU) is a word sense, expressed by the relation between a lemma and the frame that it evokes. In its lexicographic work, FrameNet focuses on developing frames and analyzing lexical units. Frame development consists of defining frames and frame-specific FEs, determining which LUs belong in a particular frame, as well as refining frame and FE definitions when required by corpus findings. Indeed,

examining and analyzing corpus attestations of groups of semantically related words is central to the process of frame development, with an explicit set of criteria used to decide on groupings of LUs for frame membership (Ruppenhofer et al., 2016, pp. 11–17). While in principle, each LU could be defined in terms of the unique frame it evokes, FrameNet nevertheless sorts LUs into groups that allow consistent analyses in terms of the FEs that define the frame. In practice, FrameNet includes frames of varying degrees of semantic granularity, from general states of affairs (e.g. Subject_stimulus) to very narrowly defined events (e.g. Appeal, as part of a criminal process), even with its preference for "splitting" (as opposed to "lumping"). In lexical frames, groupings of LUs (approximately 14/frame) permit useful generalizations (e.g. about entailments and syntactic realizations) and provide semantically annotated sentences that illustrate paraphrase relations.

To illustrate, consider the Revenge frame, which FN has characterized in terms of an AVENGER performing some PUNISHMENT on an OFFENDER as a response to an INJURY, inflicted on an INJURED_PARTY. Some of the LUs in the Revenge frame are *avenge*.v, *avenger*.n, *get back (at)*.v, *get even*.v, *retaliate*.v, *retaliation*.n, *retribution*.n, *retributory*.a, *revenge*.v, *revenge*.n, *vengeance*.n, *vengeful*.a, and *vindictive*.a, where nouns, verbs, and adjectives are included, as are multi-word expressions. The linguistic realization of each frame element highlights different participants and props of the frame, as shown in the following examples, where the ***target*** (the word being analyzed and with respect to which the Frame Semantics analysis is done) is the verb *avenge*.[1]

(1) [Sven ~Avenger~] **avenged** [his brother ~Injured_party~] [after the incident ~Time~].

(2) [El Cid ~Avenger~] **avenged** [the death of his son ~Injury~] [hastily ~Manner~].

(3) [The monkey ~Avenger~] **avenged** [himself ~Injured_party~] [by growing to the size of a giant and setting fire to the city ~Punishment~].

(4) [Hook ~Avenger~] **avenged** [himself ~Injured_party~] [on Peter Pan ~Offender~].

Notice that AVENGER, PUNISHMENT, OFFENDER, INJURY, and INJURED_PARTY are the **core** frame elements of Revenge, since they uniquely define the frame. As with other events, an act of revenge can be described as having occurred, for example, at a particular TIME as in (#1), or in a particular MANNER (as in #2). TIME and MANNER are two of the *peripheral* frame elements of the frame, describing aspects of events more generally. For each FE that is annotated in an example sentence, FrameNet also records grammatical function (from a modified list of grammatical categories) and phrase type information, thereby collecting triples of information

---

1.   The examples in (1)–(5) are based on sentences in the FN database, reflecting the same phenomena that occur in corpus attestations.

about each FE. Thus, in all of the above sentences AVENGER is recorded as an External NP.[2] The INJURED_PARTY (#1), (#3), (#4) is realized as an Object NP, as is INJURY (#2), while PUNISHMENT is realized as a PP-*ing* phrase, and OFFENDER (#4) is realized as a PP. The peripheral frame element TIME (#1) is instantiated as a PP and MANNER is instantiated as an AVP.

When a conceptually necessary and salient (i.e. core) frame element is not represented in the surface syntax of a sentence, FrameNet records it as a null in-stantiation, of which there are three types: constructional (CNI); definite (DNI); and indefinite (INI). Constructionally omitted constituents are licensed by a gram-matical construction in which the target occurs. Examples of CNI are the omitted agent in a passive sentence and the omitted subject in an imperative, as in *Her honor was **avenged** by murdering her assailant* and ***Get even** with that bum*, where the AVENGER is not mentioned explicitly, although clearly understood as a participant in the event. The other types of null instantiation are lexically specific. In (#1)–(#3), above, no lexical or phrasal material for the OFFENDER occurs in the sentences; FrameNet records that information because it provides lexicographically relevant information about omissibility conditions. In these examples, OFFENDER is omit-ted under DNI, since the referent is understood from the linguistic or discourse context. INI is the other lexically specific null instantiation, and it is illustrated with the missing objects of verbs such as *eat, bake*, and *sew*, which are usually transitive, but can be used intransitively. With such verbs the nature of the missing element can be understood without referring back to a previously mentioned entity in the discourse. In the Revenge frame, all of the verbs allow the FE PUNISHMENT to be omitted under INI; thus, for sentences (#1), (#2), and (#4) the FrameNet database would record PUNISHMENT as INI.

FrameNet also distinguishes a third type of FE, namely *extrathematic*. A frame element with extrathematic status places the current frame against the backdrop of a larger situation, as seen in the following example with the extrathematic FE Iteration.[3]

(5)   [The looters $_{Avenger}$] **revenged** [themselves $_{Injured\_party}$] [again and again $_{Iteration}$] during the demonstration.

FrameNet lexicographers annotate many example sentences for a given LU to en-sure coverage of all patterns in which it occurs. Automatic processes summarize

---

2.   FrameNet uses *external* for subjects of target verbs, and any constituent controlling the subject of a controlling verb.

3.   Ruppenhofer et al. (2016) gives a detailed description of FrameNet's FE types, and current annotation practice.

the findings, and present them in displays that show explicit information about the mapping of semantic roles to syntactic structure. One such (web-based) display is given in Figure 1, the valence patterns table for the LU *avenge*.v, which also provides clickable links to the annotated sentences. For example, clicking on the first occurrence of "2" (next to the word *TOTAL*) in the table results in the display of the two annotated sentences shown in Figure 2.

Number annotated	Patterns				
2 TOTAL	Avenger	Injured_Party	Injury	Offender	Punishment
(1)	NP	NP	PP[for]	INI	PPing[by]
	Ext	Obj	Dep	–	Dep
(1)	NP	NP	PP[of]	DNI	PPing[by]
	Ext	Obj	Dep	–	Dep
(1) TOTAL	Avenger	Injured_Party	Instrument	Offender	Punishment
(1)	NP	NP	PP[in]	INI	INI
	Ext	Obj	Dep	–	–
(10) TOTAL	Avenger	Injured_Party	Offender	Punishment	
(2)	CNI	NP	DNI	DNI	
	–	Ext	–	–	
(1)	CNI	NP	PP[on]	INI	
	–	Ext	Dep	–	
(5)	NP	NP	DNI	INI	
	Ext	Obj	–	–	
(1)	NP	NP	DNI	PPing[by]	
	Ext	Obj	–	Dep	
(1)	NP	NP	PP[on]	PPing[by]	
	Ext	Obj	Dep	Dep	
1 TOTAL	Avenger	Injured_Party	Time		
(1)	NP	NP	AVP		
	Ext	Obj	Dep		

Figure 1. Some valence patterns for *avenge*.v

[X] [Avenger A Herefordshire hobgoblin] would *AVENGE*[Target] [Injured_Party himself] [Injury for any insult][Punishment by stealing all the family's keys and refusing to return them until his favourite cake had been baked and left on the hob for him to eat] .[Offender INI]

[X] Others say [Avenger she] *AVENGES*[Target] [Injured_Party herself] [Injury of the insult offered by her erstwhile lover] [Punishment by luring fishermen and other sailors to their doom]. [Offender DNI]

Figure 2. Two annotated sentences for *avenge*.v

The FrameNet database holds nearly 1,200 frames, approximately 13,000 lexical units, and almost 200,000 manually annotated example sentences.[4]

In addition to frames, FEs, lexical units, and annotations, FrameNet also records frame-to-frame relations in the database, the most important of which are *Inheritance* and *Subframes*. Frame inheritance is a relationship in which a child frame is a more specific elaboration of its parent frame. Thus, all of the frame elements, other frame relations and (semantic) characteristics of the parent have equally or more specific correspondents in the child frame. For example, the `Revenge` frame inherits from the `Rewards_and_punishment` frame, some of whose LUs are *discipline*.v, *reward*.n and *punitive*.a, and where the FE Evaluee corresponds to the more specific FE Offender in the `Revenge` frame. Subframes is a relationship that characterizes the different sequential parts of a complex event in terms of the sequences of states of affairs and transitions between them, each of which can itself be separately described as a frame. For instance, the complex `Traversing` frame has two subframes, `Departing` and `Arriving`, the former bearing the *Precedes* relation to the latter.[5]

FrameNet appeals to the concept of *prototype* in the process of defining frames: a LU that best exemplifies the situation described in the frame is chosen for analysis and serves as the vehicle for characterizing the frame more generally. In addition, FrameNet implements the concept of *perspective* by categorizing words into separate frames based on differences in perspective, and recording such information with the frame-to-frame relation *Perspective_on*. To illustrate, the often-cited `Commercial_transaction` frame constitutes a family of frames, whose participants and props are Buyer, Seller, Money, and Goods. The family includes one frame that characterizes the transfer of goods and another for the transfer of money. Each of the lexical frames associated with the transfer of goods, `Commerce_buy` and `Commerce_sell`, takes a different perspective, the former of the Buyer and the latter of the Seller, as in Figure 3. Likewise, each of the lexical frames associated with the transfer of money, `Commerce_pay` and `Commerce_collect`, takes a different perspective, the former of the Buyer and the latter of the Seller.

Ellsworth, Ruppenhofer and Ziem (2014) argues that although the FrameNet database was developed originally as a frame-based *lexicon*, it constitutes a repository of constructions too. FrameNet provides a wealth of information about the mapping of form to meaning through the theory of Frame Semantics, primarily focusing on *lexical constructions*. Thus, the database includes much information

---

4.   These numbers are current as of February 2015.

5.   Petruck and de Melo (2012) offers a compelling use-case, and demonstrates the need for the relation Precedes in natural language processing applications.
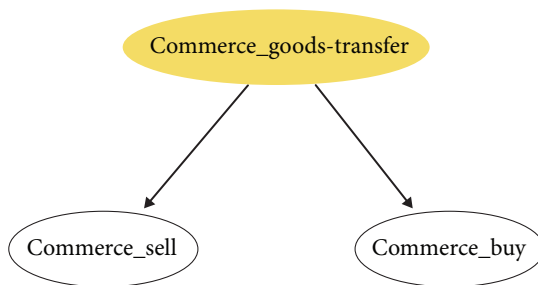
Figure 3. Frame-to-Frame relation perspective_on for `commerce_sell` and `commerce_buy`

about constructions, including the fundamental (to Construction Grammar) understanding that frames constitute the semantic side of constructions. Somewhat similarly, every valence description in FN includes form-side information in its collection of phrase types paired with frame elements. (See Section 7 for further discussion of the parallels between the conceptual apparatus of the FrameNet lexicon and that of the FrameNet Constructicon.)

3. Constructicon terminology

This section introduces and defines the terminology of Construction Grammar (CxG) and that of the FrameNet Constructicon (FN-CXN), recognizing that to a certain extent terminology development for the constructicon relied on some of FN's existing terminological practices and conventions.

The fundamental unit of analysis in CxG is the *grammatical construction*, defined as a pairing of a form and a meaning that specifies a particular external syntax, along with its semantic-pragmatic interpretation for a combination of syntactic, morphological, and/or lexical elements. Linguists working in the CxG framework have identified and analyzed numerous grammatical constructions, both schematic, for instance, the `subject_predicate` construction (e.g. *Sam runs*), or the `double_object` construction, (e.g. *Jerry sold Chuck a car*), and lexically specific, for instance the *way* construction (e.g. *Jo pushed his way into the bar*), or the LTN[6] construction (e.g. *last Thursday*), etc. A *construct* is a linguistic form that instantiates one or more constructions, as for example in the sentence *Eric doesn't like reading novels, let alone romance novels*, which instantiates the `let_alone` construction (Fillmore, Kay & O'Connor, 1988). In the simplest case, a construct instantiates just

---

6. Fillmore (2002) dubbed this construction the *LTN_construction*, where *LTN* stands for *last this next*, the words that may modify one of the possible day names of day part names.

one construction; for example the time phrase *Monday morning* is a construct that instantiates the day_name_plus_day_part_name construction, devoted to identifying parts of days. The terms *constructional form* and *constructional meaning* together capture the two sides, or specifications of a construction, namely the form side and the meaning side, respectively. A *constructional form* is a lexemic and/or morphosyntactic specification of a construction, for example the lexeme *want*, or [VP [V $_{Transitive}$ + NP $_{Direct\ Object}$ + NP $_{Indirect\ Object}$]], a schematic representation of the double_object construction; and a *constructional meaning* is a semantic and/or pragmatic specification of a construction, for example, Promising, Causation, Conjunction, and deference.

A *Construct Element* (CE) is a constituent part of a construction. CEs themselves are constructions, although these constructions may be very general. For example, the only form-side constraint on the SUBJECT CE in the Subject-Predicate construction is that the subject must be an NP.

A *Construction-evoking Element* (CEE) is lexical material that is central to, or that cues the existence of, a particular construction, as for example *way* in the *way* construction (e.g. *Greg elbowed his way through the crowd*). Obviously, not all constructions will have a CEE; in particular, purely grammatical constructions, such the shared_completion construction (as in *Deborah wants and I have a red cashmere sweater*) and the Imperative construction (*Shut up!*), do not have CEEs.

In addition to the above items, the FrameNet Constructicon also includes a number of relations, which we define and exemplify here (where appropriate).

Constructional Inheritance characterizes the situation when one construction, a child construction, inherits another construction, the parent, if all of the formal, semantic, and pragmatic constraints of the parent are also true of the child, with the potential of additional constraints in the child. Constructional inheritance entails CE inheritance, for instance the mileage (e.g. *30 miles per gallon*) construction inherits all of the characteristics of the (most general) rate construction (e.g. *Andy read **three essays per hour***), and the yes-no_question construction (e.g. *Did Abby attend the conference?*) inherits from subject_aux_inversion construction, a schematic construction that licenses subjects and auxiliaries.

Construct Element Inheritance involves one CE, a child CE, inheriting another CE, i.e. the parent CE, if all of the formal, semantic, and pragmatic constraints of the parent also hold true of the child, with the potential of additional constraints on the child. CE-to-CE inheritance is always part of construction-to-construction inheritance. For example, the CEs in the comparison_equality construction (e.g. *That's **as good a reason as any***) inherits all of the CEs from the general comparison construction, whose CEs include COMPARISON_MARKER and a noncomparative BASE_EXPRESSION. Note that the child construction may have additional construction-specific restrictions.

Construct Element-to-Construction link: A CE in a construction is specified as necessarily instantiating another construction. For example, in `subjectless_tag_question` sentences (e.g. *Fooled you, didn't I?*), the tag must instantiate the `predicate_ellipsis` construction (*…didn't I?*).

Construction-to-Frame link: A construction's meaning is specified to be the meaning of a particular frame, or that of one which has inherited frames that do so. For instance, the meaning of the `rate` construction (e.g. *He types **75 words per minute***) is the meaning of a `Rate` frame[7]. Notionally, having a construction-to-frame link is like placing a construction in a frame, and analogous to placing LUs in frames. (See the discussion of `be_recip` in Section 5.)

4.   Choosing a construction and exploring the construction's use

The process of choosing a construction to analyze proceeds in similar ways to that of choosing a semantic frame. To begin with, a number of constructional patterns have been described in the literature: *let alone* (Fillmore, Kay & O'Connor, 1988), argument structure constructions (Goldberg, 1995), applicatives (Michaelis & Ruppenhofer, 2001), `nominal extraposition` (Michaelis & Lambrecht, 1996), resultative constructions (Boas, 2003), and *big mess* (Kay & Sag, 2012), among many others, for numerous languages (Boas, 2010). Or, the analyst may have noticed interesting phenomena while analyzing structurally or semantically related constructions. For example, the analysis of `time_when` constructions (e.g. *during the previous week*) might motivate working on the `time_unit_after_time_unit` construction (e.g. *week after week*), which in turn could prompt studying the `time_unit_by_time_unit` construction (e.g. *week by week*), and so on. Finally, the Construction Grammarian may wish to describe constructions that will cover a significant portion of a corpus of interest. For instance, a biomedical corpus could suggest focusing on series of noun compounds (e.g. *hip pain diagnosis procedure*). Before describing the exploratory and analytical steps that follow the selection of a construction to analyze, we discuss how the analyst determines construction-hood.

Consider the sentence in (6). Begin with the simple observation that it does not fit the expected pattern of number agreement between pre- and post-copular noun phrases. We use a corpus search tool that facilitates viewing items of interest

---

7.   As of this writing, FrameNet has defined `Rate_description` and `Rate_quantification` frames, neither of which capture the more general `Rate` frame that the `rate` construction requires for its meaning.
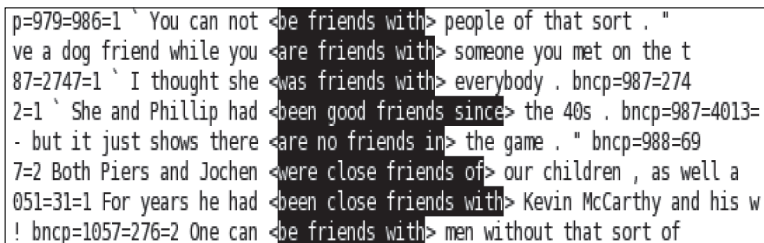
in a "keyword in context" manner (such as xkwic)[8], which FrameNet has used since its earliest days, or the web interface for the Corpus of Contemporary American English (http://corpus.byu.edu/coca)). This search provides minimal confirmation of our suspicions about the existence of the construction, and by scanning the results, reveals any common surrounding patterns.

(6)   Mitchell is friends with Kaley.

An initial exploratory search for the construction is simply to request the lemma *be* preceding the word form *friends*. Examining the results leads to a few observations:

–   Occasionally, the word *friends* has a modifier: *close*, *best*, *childhood*, etc.
–   Some additional patterns may or may not be related, including:
    –   [plural subject] + *be* + *friends* (*the two women were close friends*)
    –   [plural subject]+*be* + *friends* + *of* (*many were middle-aged friends of Sue*)

```
p=979=986=1 ` You can not <be friends with> people of that sort . "
ve a dog friend while you <are friends with> someone you met on the t
87=2747=1 ` I thought she <was friends with> everybody . bncp=987=274
2=1 ` She and Phillip had <been good friends since> the 40s . bncp=987=4013=
- but it just shows there <are no friends in> the game . " bncp=988=69
7=2 Both Piers and Jochen <were close friends of> our children , as well a
051=31=1 For years he had <been close friends with> Kevin McCarthy and his w
! bncp=1057=276=2 One can <be friends with> men without that sort of
```

Figure 4. Keyword-in-context view for *be …friends* + preposition

Each of these observations leads to follow-up questions, which may be formulated as increasingly refined corpus queries. For instance, (1) What are the possible modifiers of *friends*? (2) Do the modifiers change depending on whether the subject is singular or plural, or whether the word *friends* is followed by *of* or *with*? (3) Is the subject's plurality a crucial part of the construction, or is whether the members of the relationship are split across multiple noun phrases more important? (4) Does this behavior interact with possible modification and the choice *of* as compared to *with*? (5) What words other than *friend(s)*, fit this pattern?

These questions are familiar to any constructional analysis, namely, to figure out the distributional or co-occurrence restrictions found among all the parts of the (potential) construction. This activity is in service of asking one of the crucial

---

8.   XKWIC is a graphical interface to corpus search tools that the Institute for Natural Language Processing of Stuttgart University developed (http://www.ims.uni-stuttgart.de/forschung/projekte/CorpusWorkbench.html).

questions for determining construction-hood: is this pattern of language use, including both form and meaning, predictable on the basis of other (perhaps more general) principles of the language, or does it have features that demand specific analysis? If specific analysis is required, then this pattern is a construction that should be recorded in the constructicon.[9]

Through the process of answering these questions and determining that, in fact, there is a bundle of linguistic features that are idiosyncratic, or otherwise unpredictable, we build an argument in favor of this pattern being a construction. In doing so we also define the boundaries of the construction, the syntactic and semantic (among others) constraints on its constituent parts, and its overall meaning. The remainder of this section is devoted to examining such features and building an argument for construction-hood. While constructing a complete argument is beyond the scope of this section, the analysis presented is sufficient to create a definition and annotate the construction.

Consider the first two questions listed above: what are the possible modifiers of *friends* and how, if at all, does this impact the other putative parts of the construction? Initially, we may limit the search to sentences with a singular subject, given the near certainty that such structures are of interest. Looking at the most frequent adjectives in the template *is [adj.] friends*, yields the following:

(7)   good, close, best, longtime, old

We then contrast this finding with the adjectives found with *are [adj.] friends*, which given its canonical syntax and agreement, seems less idiosyncratic. Here we find the same adjectives as in (7), plus others: *trusted*, *peaceful*, *artistic*, *male*. From these patterns we start to form a characterization of the possible adjectives in our construction: they pick out attributes of the friendship (its quality and duration), rather than, say, some characteristics of the individuals in the friendship.

At this point we have established a potential collocational pattern, namely a restricted set of adjectives associated with a singular subject. Going through one more step, we now compare the adjectives with any following prepositions. The most common prepositions following *friends* are *of*, *in*, and *with*. They appear with the following sets of adjectives:[10]

---

9.   While some who adopt the constructional approach to language description propose frequency as a criterion of construction-hood (e.g. Goldberg, 2006; Hilpert, 2014), we do not consider frequency definitional.

10.   Only looking at linear order ignores the fact that some of these prepositions are associated with distant verbs, but not with *friends*. However, the idea is to determine whether any general patterns exist for focusing on with more targeted searches.

(8)   a.   of: close, good, old, other, American, longtime, dear, great, personal, best
      b.   in: best, old, close, good, new, closest, visiting, other, powerful, dear
      c.   with: good, close, best, fast, still, old, great, personal, lifelong, better

While the sets show significant overlap, we again see that *with* seems less likely to appear with certain individual-describing adjectives like *American* and *powerful*.

Introspection also serves as a guide in the early stages of formulating the arguments for the existence of a construction. Based on the above observations, some illustrative sentences can be tested for grammaticality, or for expected semantics:

(9)   a.   Sue is good friends with Bob.
       b.   *Sue is powerful friends with Bob.
       c.   *Sue is good friends of Bob.
       d.   Sue is a friend of Bob.
       e.   *Sue is a friend with Bob.

(10)   a.   Sue is the most powerful friend of the president.
       b.   *Sue is (the most) powerful friends with the president.

(11)   a.   {Sue and Bob/They} are good friends.
       b.   {Sue and Bob/They} are good friends {of /with} the president.
       c.   {Sue and Bob/They} are powerful friends {of/*with} the president.
       d.   {Sue and Bob/They} are powerful friends.

Examples (9a)–(c) and (10b) illustrate, as a hard grammatical constraint, the corpus patterns observed above: with a singular subject and plural *friends*, only certain adjectival modification is possible, and *friend* takes a PP-*with*, rather than PP-*of*. With a typical copular clause (*X is a friend*), we only see *friend of*, not *friend with*. Example (10a) confirms that with *friend of*, there are no limits on adjectival modification. This constellation of observations forms the core of the features we want to associate with the construction.

The sentences in (11) test these preliminary ideas against plural subjects. While restrictions seem to disappear with the adjective and the prepositions *of/with*, they remain in subtle ways. When the members of the friendship are expressed with two noun phrases (one subject and one oblique), as before, we see that adjectives like *powerful* cannot appear with the preposition *with*. Thus, the true generalization about the construction does not involve a singular subject specifically. Instead, it involves a choice between grouping all the friends in one noun phrase or splitting them across the predicate. In other words, the putative `friends_with` construction participates in an alternation common across English words that express reciprocal relations, as shown in (12), below.

(12) a.   Kim danced with Pat/Kim and Pat danced.
      b.   Kim met (with) Pat/Kim and Pat met.
      c.   Kim is married to Pat/Kim and Pat are married.

As with these alternations, sentences such as those in (11) are ambiguous: Sue and Bob may be friends with one another, or with some unnamed other(s). When it is the former, we may consider the sentence to be an instance of the *friends with* construction; when it is the latter, perhaps it is a more conventional combination of syntactic and semantic elements. Aside from the elegance of treating (9a) and (11a) as constructional instances of a well-established valence alternation pattern, there is semantic evidence that this is on the right track. Because *friends with* permits limited adjectival modification, we expect (11d) to not be an example of this construction. And indeed, (11d) is readily understood as meaning that Sue and Bob are both powerful and both friends with some unnamed other person – but not that they are both powerful and friends of one another.

At this point we have determined the following elements of the construction:

– A plural noun, *friends*, which appears as a post-copular noun
– A limited set of modifiers that describe the friendship relationship
– A valence alternation, with
    –   a plural subject, indicating the mutual friends, or
    –   a subject and a PP-*with*, indicating the two parties

Needless to say, each point could be expanded upon and further bolstered with additional corpus evidence. The two elements that certainly need further refinement concern the noun and possible modifiers: are other nouns possible, and what is the best way to characterize the set of licit modifiers? Again, a combination of intuition and corpus searches are necessary to yield good results. The BTC study (some of which Fillmore, Lee-Goldman & Rhomieux, 2012 documents) found that many reciprocal-relationship nouns were possible in the construction: *buddy*, *roommate*, *sibling*, *partner*, *colleague*, *arch-rival*, and so on. Among modifiers, aside from those already mentioned, we also found nouns indicating the origin of the relationship: *high-school buddies*, *work friends*.

5.   Defining the `be_recip` construction

Using the analysis worked out in the previous section, this section lays out the definition and specific parts of the construction. Figure 5 displays the definition of the construction.

Definition	A plural nominal is used as a reciprocal predicate. The head of the nominal is a term of reciprocal personal relationship (e.g., friends). The nominal may include modifiers that describe the relationship itself (e.g., close friends, college roommates), though not the individuals (e.g., wealthy). This predicate participates in the reciprocal alternation: the parties to the relationship may be expressed as a single (plural) subject (*They are good friends*) or as a subject and a with-marked PP (*She is good friends with her mother*). Note that with is used regardless of the preposition the head noun would normally appear with (*I met a close friend of/*with the President*).
Construction Elements	Head_noun Modifier Individuals Individual_1 Individual_2
Construction-evoking Element	None
Construct's properties	Nominal predicate
Evoked Frame	Reciprocality. The Head_noun CE specifies the type of the reciprocal relationship.

Figure 5. The `be_recip` construction

Each part of the construction about which something special must be said (and which is realized in sentences containing the construction) is designated as a CE. All the parts discussed in the previous section, i.e. the main sentential arguments, the ʜᴇᴀᴅ ɴᴏᴜɴ, and the ᴍᴏᴅɪꜰɪᴇʀs, exhibit idiosyncratic behavior in this construction; these parts are set up as CEs. No CEE is posited. Although the set of possible ʜᴇᴀᴅ ɴᴏᴜɴs is limited, this limitation can be stated in semantic terms; and, in principle, the class is open, unlike, for instance, *let alone* or the *way* construction (*elbow your way*), which have fixed lexical elements. The structure that the construction licenses directly, i.e. a plural nominal, such as *close friends*, is specified as a nominal predicate. As such, it appears in copular constructions (*be close friends*), with *become* and *as* (*become close friends with her*, *see myself as close friends with him*), as a pre-posed subject modifier (*close friends with the president, she …*), and so on.

Constructions may be syntactic arrangements, such as the `shared_comple-tion` construction, or form-meaning pairings. The latter type always evokes frames, much as though they were LUs. The *way* construction in (13) is a typical example of a frame-evoking construction. The main CEs are: (i) a ᴠᴇʀʙ like *punch* or *whistle* and (ii) a ᴘᴏssᴇssᴇᴅ NP headed by *way*. Neither of these elements evokes a

frame of motion, but the construction as a whole does (specifically, FrameNet's `Self_motion` frame). Thus, the arguments of the construction (the subject, the path-expressing PPs, etc.) are properly analyzed as FEs of `Self_motion`, just as if the motion-indicating predicate was a simple verb like *run*.

(13)   The kids *whistled their way* down the street.

Likewise, *be_recip* evokes the `Reciprocality` frame. FrameNet has defined this frame as characterizing situations with "[p]rotagonists in relations with each other that may be viewed symmetrically." The frame also specifies an "equivalence of Protagonist_1 + Protagonist_2 [with] Protagonist." In other words, the parties in the relation may be conceived of as a complex whole or as two related parties. No LUs in English evoke this frame directly, although phrasal constructions do evoke the frame. Frames that inherit `Reciprocality` include words like *meet* or *collaborate*, which participate in the reciprocality valence alternation. However, because *be_recip* is separate from the specific reciprocal relationships (like friendship or colleague-hood), it does evoke `Reciprocality` directly.

The CEs of *be_recip* are, therefore, realizations of the `Reciprocality` frame's FEs. Namely, ɪɴᴅɪᴠɪᴅᴜᴀʟ_1 is Pʀᴏᴛᴀɢᴏɴɪsᴛ_1, ɪɴᴅɪᴠɪᴅᴜᴀʟ_2 is Pʀᴏᴛᴀɢᴏɴɪsᴛ_2, and ɪɴᴅɪᴠɪᴅᴜᴀʟs are Pʀᴏᴛᴀɢᴏɴɪsᴛs. The ʜᴇᴀᴅ_ɴᴏᴜɴ CE indicates the type of reciprocal relationship (which we may consider an implicit FE of `Reciprocality`) and adds its semantic information to the CEs (e.g., for *friend*, the individuals are in a friendship relationship).

With the construction, its form- and meaning-side properties, and its subparts determined, the next step is annotation of instances of the construction from natural language corpora. The following section describes the resulting annotations, along with possibilities for viewing the information.

6.   Annotation

Annotation with respect to a construction proceeds similarly to that for a LU in a frame. The annotation must indicate the construct, the CEE (if any), the CEs (if any), and any external segments that play a key role in the construction in a sentence. The last category includes mainly support words like *have* and *do* as well as the copula. The annotation may also include, on separate layers, indications of grammatical function and phrase type. These categories are familiar from FrameNet annotation, and serve similar roles, i.e. to highlight variations in the realization of the construction and how it fits with the remainder of the sentence.

The sentence labeled (16) in Figure 6 illustrates the following elements.

16. The same year, [Individual_1 Clinton] was also to become^{Supp} {Be_recip[Head_noun friends}]
    [Individual_2 with a trio of women who had advanced ideas about sex]. TOP
17. This year's event on Sunday July 5 will start and finish at Southlands Centre, rather
    than Albert Park and [Individual_1 Middlesbrough and District Harriers] are^{Cop}
    {Be_recip[Head_noun co-organisers}] [Individual_2 with the council]. TOP
18. '[Individual_1 Sally] used to be^{Cop} {Be_cecip[Modifier good] [Head_noun mates}] [Individual_2 with Zaria],
    did n't she ?

Figure 6. Annotation of *be_recip*

- Individual_1: *Clinton*
- Support word: *become*
- Construct span (indicated with curly braces): *friends*
- Head_noun: *friends*
- Individual_2: *with a trio of women who had advanced ideas about sex.*

In other sentences, the construct span and HEAD_NOUN would not cover the exact same span. The sentence (labeled 18) has *mates* as HEAD_NOUN, and *good mates* as the construct span.

The `adjective_as_nominal.people` construction licenses expressions such as *the very young*, as in *The greatest danger to the very young is to be taken from home*. Constructs of this construction contain a definite determiner and an adjective phrase, and as a whole denote generic groups (usually of people) with the attribute characterized by the adjective phrase. The construction has only two CEs (the DETERMINER and the ADJECTIVE PHRASE) and a CEE (the DETERMINER). It licenses no CEs outside the construct span. Figure 7 shows the annotation.

23. {Adjective_as_nominal.people[Definite_determiner <The>] [Adjective_phrase innocent}] suffer and
    {Adjective_as_nominal.people[Definite_determiner <the>] [Adjective_phrase conscientious}] are trodden
    down into conformity . [allLayers] TOP
24. {Adjective_as_nominal.people[Definite_determiner <The>] [Adjective_phrase less well-off}] are doing things
    like swapping chocolate biscuits for cheaper plain ones;
    {Adjective_as_nominal.people[Definite_determiner <the>] [Adjective_phrase better-heeled}] are staying out of
    restaurants and eating gourmet meals at home instead . [allLayers] TOP
25. {Adjective_as_nominal.people[Definite_determiner <The>] [Adjective_phrase rich}] live as fearful princes :
    {Adjective_as_nominal.people[Define_determiner <the>] [Adjective_phrase poor}] live as angry beggars .
    [allLayers] TOP

Figure 7. Annotation of `adjective_as_nominal.people`

Not shown in this view are additional layers of annotation that indicate the formal (morphosyntactic) properties of the construction and its parts. The text that receives the ADJECTIVE_PHRASE CE label also receives a phrase type label indicating that it is an adjective phrase.[11] The construct span as a whole is also labeled with a phrase type (NP), indicating the formal features that dictate its placement within the larger sentence.

Constructions that cover larger spans are annotated as well: for instance, subjectless tag sentences, described in constructional terms by Kay (2002), consist of a HOST sentence, and a TAG. Kay enumerates several features of the construction that constrain and contribute to its overall interpretation, notably the presence or absence of modal auxiliaries and negation in the HOST and the TAG. These are annotated on secondary layers, represented in Figure 8 as italic subscripts. (Note the labels *Host_negation* and *Tag_subject*). The construction's name, `tagged_sentence.subjectless`, contrasts with that of the `tagged_sentence.canonical` (*It isn't raining, is it?*).

4. {_{Tagged_sentence.subjectless}[_{Host}Told you I was half-dead] , [_{Tag}<did _{Tag subject}she>]} ? TOP

5. {_{Tagged_sentence.subjectless}[_{Host}Give you a flipping chance to even enjoy it] , [_{Tag}<do _{Tag subject}they>]} ?! TOP

6. {_{Tagged_sentence.subjectless}[_{Host}Do _{Host negation}n't want sheep] , [_{Tag}<do _{Tag subject}we>]} ? TOP

Figure 8. Annotation of tagged_sentence.subjectless

As with LUs, annotation of constructions (or, what FrameNet calls *constructicographic annotation*) illustrates the range of a construction's realizations as straightforwardly as possible. Constructions have a number of parameters that are ideally illustrated in the annotation, including the possible CEEs, the presence or absence of optional elements, and, if the construction is frame-evoking, external elements that realize the frame's FEs. Additionally, annotation aids in the process of developing and refining construction definitions. In the course of annotation discovering that the data does not perfectly match the stated properties of a construction is not uncommon, thus necessitating updates to the construction's definition.

---

11. Naming the CE *Adjective_phrase* does not constitute a syntactic analysis; doing so is simply a name. Therefore, we also annotate CE spans (when appropriate) with one of a set of labels specifically used to indicate syntactic phrase types.

7.   FrameNet Lexicon-Constructicon analogues

Given the shared intellectual history of the sister theories Frame Semantics and Construction Grammar, which hearkens back (at least) to the often cited visionary paper 'The case for case' (Fillmore, 1968), as well as the more recent explicit call for an "Articulation of Lexicon and Constructicon" (Fillmore, 2006, p. 35), drawing analogies between the FrameNet lexicon and the FrameNet Constructicon seems both natural and necessary (Petruck, 2014). Figure 9, shows lexicon-constructicon analogues.

FrameNet Lexicon	FrameNet Constructicon
Frame	Construction
Frame Evoking Element (FEE) (Lexical Unit, LU)	Construction Evoking Element (CEE)
Frame Element (FE)	Construction Element (CE)
lexicographic annotation	constructicographic annotation

Figure 9. Lexicon-Constructicon analogues[*]

[*] Figure 9 also appeared in Petruck and Ziem (2014). The observant reader will note the absence of *construct*. A construct is the specific instantiation of a construction (i.e., in a particular sentence), thus, the closest analogue to a construct in FrameNet is a particular realization of a valence pattern of a LU, again in a particular sentence. As neither of these units have any particular theoretical status, we excluded from the comparison.

Comparing frames and constructions on the level of complex entities is quite natural. The FEE is an indication of the specific semantic frame's relevance to interpretation, just as a CEE indicates that a construction may license the current linguistic structure. Likewise, both frames and constructions have their own distinctive elements (FEs and CEs, respectively), with annotation from natural language corpora illustrating these elements.

At the same time, a construction may be compared to a lexical unit, as both are connected in analogous ways to semantic frames. To illustrate, the Friendship frame has a schematic representation of a friendship between two parties. The specific LUs of `Friendship`, including *buddy*, *pal*, *friend*, and *mate*, elaborate on the frame's definition, adding more specific information about the two parties, their relationship, and the linguistic (e.g., syntactic) means of expressing those parties. A construction like *be_recip* also evokes a frame, the more general `Reciprocality` frame. It too specifies a means of expressing the FEs of that frame, and combines with various relationship-denoting words (i.e. the HEAD_NOUN CE described in Section 5) to elaborate upon the basic meaning of `Reciprocality`.

These analogues arise naturally when considering the types of information that LUs, constructions, and frames encode, along with the connections between them. This natural development only emphasizes the need to develop frames and constructions in parallel and the benefits of establishing thorough connections between the lexicon and constructicon.

8.   Summary and conclusion

The ideal constructicon will be consistent with construction-grammatical theory and will embody claims that corpus data support. The BTC pilot project demonstrated the feasibility of building such a constructicon alongside FrameNet. Fillmore, Lee-Goldman and Rhomieux (2012) described the connections to grammatical theory, and provided detailed analyses of a number of constructions in English (e.g. `rate, way`, and `degree_modification`, etc.). The present paper has shown the basics of the creation of that constructicon, starting from the basic components of a construction, to determining construction-hood, and finally creating a definition and annotating examples from a corpus.

A constructicon should also provide utility to language technology applications. Current approaches to segmentation and parsing (both syntactic and semantic) are limited by the types of structures used at present in language representation, such as parse trees and simple predicate-argument structure. A constructicon recognizes the breadth of linguistic structures, which go beyond words and multi-word expressions to partially filled templates with various degrees of schematicity. More importantly, by connecting constructions and frames from the beginning, it places these structures within a semantic network, providing a link between complex linguistic form and rich representations of meaning.

Acknowledgments

References

Boas, H. C. (2003). *A Constructional Approach to Resultatives*. Stanford: CSLI. Amsterdam/ Philadelphia: John Benjamins.

Boas, H. C. (Ed.). (2010). *Contrastive Studies in Construction Grammar*. Amsterdam/Philadelphia: John Benjamins. doi:10.1075/cal.10

Boas, H. C. (2014) *What happened to Frame Semantics?* Plenary Talk at Eighth International Conference on Construction Grammar, Osnabrueck.

Ellsworth, M., Ruppenhofer, J., & Ziem, A. (2014). *FrameNet as a repository of constructions: A case study of non-local argument realization*. Paper presented at the Eighth International Conference on Construction Grammar, Osnabrueck.

Fillmore, C. J. (1968). The Case for Case. In E. Bach & R. Harms (Eds.), *Universals in Linguistic Theory* (pp. 1–88). London: Holt, Rinehart, and Winston.

Fillmore, C. J. (1985). Frames and the semantics of understanding. *Quaderni di Semantica*, 6, 222–254.

Fillmore, C. J. (1988). The Mechanisms of Construction Grammar. In Proceedings of the Fourteenth Annual Meeting of the BerkeleyLinguistics Society (pp. 35–55). Berkeley.

Fillmore, C. J. (2002). Mini-grammars of some time-when expressions in English. In J. Bybee & M. Noonan (Eds.), *Complex Sentences in Grammar and Discourse* (pp. 31–59). Amsterdam/ Philadelphia: John Benjamins. doi:10.1075/z.110.04fil

Fillmore, C. J. (2006). The Articulation of lexicon and constructicon. In S. Fujii, T. Morita, & C. Sakuta (Eds.), [Abstracts to] Proceedings of the Fourth International Conference on Construction Grammar (pp. 53–63).

Fillmore, C. J., & Baker, C. F. (2010). A Frames Approach to Semantic Analysis. In B. Heine & H. Narrog (Eds.), *The Oxford Handbook of Linguistic Analysis* (pp. 791–816). Oxford: Oxford University Press.

Fillmore, C. J., Kay, P., & O'Connor, M. C. (1988). Regularity and idiomaticity in grammatical constructions: The Case of Let Alone. *Language* 64(3), 501–538. doi:10.2307/414531

Fillmore, C. J., Lee-Goldman, R. R., & Rhomieux, R. (2012). The FrameNet Constructicon. In I. A. Sag & H. C. Boas (Eds.), *Sign-Based Construction Grammar* (pp. 283–322). Stanford: CSLI Publications.

Fillmore, C. J., Petruck, M. R. L., Ruppenhofer, J., & Wright, A. C. (2003). FrameNet in Action: The Case of Attaching. *International Journal of Lexicography*, 16(3), 297–332. doi:10.1093/ijl/16.3.297

Fried, M. (2015). Construction Grammar. In T. Kiss, & A. Alexiadou (Eds.), *Syntax – Theory and Analysis: An International Handbook* (2nd ed.), (pp. 974–1003). Berlin: Walter de Gruyter.

Goldberg, A. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.

Goldberg, A. (2006). *Constructions at Work: The Nature of Generalization in Language*. Oxford: Oxford University Press.

Hilpert, M. (2014). *Construction Grammar and its application to English*. Edinburgh: Edinburgh University Press.

Kay, P. (2002). English Subjectless Tag Sentences. *Language*, 78(3), 453–481. doi:10.1353/lan.2002.0161

Kay, P., & Fillmore, C. J. (1999). Grammatical constructions and linguistic generalizations: the What's X doing Y? Construction, *Language* 75(1), 1–33. doi:10.2307/417472

Kay, P., & Sag, I. A. (2012). Cleaning up the Big Mess: Discontinuous Dependencies and Complex Determiners. In I. A. Sag & H. C. Boas (Eds.), *Sign-Based Construction Grammar* (pp. 211–238). Stanford: CSLI Publications.

Michaelis, L. A., & Lambrecht, K. (1996). Toward a Construction-Based Model of Language Function: The Case of Nominal Extraposition. *Language*, 72(2), 215–247. doi:10.2307/416650

Michaelis, L. A., & Ruppenhofer, J. (2001). *Beyond Alternations: A Constructional Account of the Applicative Pattern in German*. Stanford: CSLI Publications.

Petruck, M. R. L. (2014). *The Sisterhood of Frame Semantics and Construction Grammar*. Paper presented at Eighth International Conference on Construction Grammar, Osnabrueck.

Petruck, M. R. L., & de Melo, G. (2012). Precedes: A Semantic relation in FrameNet. *Proceedings of the Workshop on Language Resources for Public Security Applications* (pp. 45–49), Eighth Conference on International Language Resources and Evaluation, Istanbul.

Petruck, M. R. L., & Ziem, A. (2014). *Semantic Relations for Frames and Constructions: Perspective_on*. Paper presented at the Eighth International Conference on Construction Grammar, Osnabrueck.

Ruppenhofer, J., Ellsworth, M., Petruck, M. R. L., Johnson, C. R., Baker, C., & Scheffczyk, J. (2016). *FrameNet II: Extended Theory and Practice*. Retrieved from https://framenet.icsi.berkeley.edu/fndrupal/the_book.

Constructicography at work

Theory meets practice in the Swedish constructicon

Benjamin Lyngfelt, Linnéa Bäckström, Lars Borin,
Anna Ehrlemark and Rudolf Rydstedt

This chapter addresses central topics in constructicography from the viewpoint of the Swedish constructicon project (SweCcn), focusing on practical constructicon development. The full process of construction description is described and discussed, from selection via corpus analysis to finished constructicon entry and beyond, towards structuring the set of entries into a network. Particular attention is given to the description format and the treatment of constructional variation. A main theme in the chapter is the interdependence and alignment of SweCcn and related resources, on the one hand in the local context, notably the infrastructure of Språkbanken (the Swedish language bank), and on the other hand with respect to corresponding resources for other languages. Of key concern is the relation to FrameNet, both the Swedish and other framenets, and a major section is devoted to conditions for linking constructions and frames.

Keywords: constructicography, constructicon, construction, construction grammar, FrameNet, language technology, lexicography, Swedish

1. Introduction

The Swedish Constructicon (SweCcn)[1] is a freely available online repository of Swedish construction descriptions (e.g. Lyngfelt et al., 2012; Sköldberg et al., 2013). While primarily designed as a linguistic tool, it is also intended as a resource for language technology and pedagogical applications. A particular concern is relevance for learners of Swedish as a second or foreign language. At the time of writing, the database consists of about 400 entries, accounting for a large variety of constructional patterns. The current set is somewhat biased towards partially schematic constructions in the borderland between grammar and lexicon, but our ambition is to account for constructions across almost the whole grammar-lexicon

---

1. <https://spraakbanken.gu.se/konstruktikon>

continuum, excluding only lexical constructions. (See Lyngfelt, this volume, for an introduction to constructions and related concepts.)

SweCcn is a practical instantiation of Construction Grammar (CxG), while also borrowing much of its methodology from lexicography – that of traditional dictionary making (e.g. Atkins & Rundell, 2008; Svensén, 2009) as well as that of FrameNet (e.g. Ruppenhofer et al., 2016). The design is to a large extent inspired by the English FrameNet Constructicon in Berkeley (Fillmore 2008; Fillmore, Lee-Goldman & Rhomieux, 2012; Lee-Goldman & Petruck, this volume), which in turn is based on the format in FrameNet. Due to practical concerns, constructicon analyses are simplified in comparison to more elaborate CxG formalisms (e.g. Fillmore & Kay, 1996; Sag, 2012); a constructicon entry in SweCcn may be characterized as a cross between a CxG account and a dictionary entry.

Thus, as CxG collapses the traditional distinction between grammar and lexicon, practical constructicon development – constructicography – combines grammar description and lexicography. This combination is not straight-forward in all respects, since the two traditions have developed according to partially different goals. In the present chapter, we provide a comprehensive presentation of SweCcn, accounting for the choices made in accommodating these goals.

Section 2 presents the local context of SweCcn, with particular focus on the infrastructure of *Språkbanken* ('the Swedish language bank') and the Swedish FrameNet++ project. In Section 3, we illustrate the practical constructicon development. Section 4 addresses the treatment of constructional variation, and Section 5 deals with the alignment of SweCcn and FrameNet. Section 6 accounts for the SweCcn description format, and Section 7 is concerned with the user's perspective. In the concluding Section 8, we present an outlook, focusing on future constructicon development into a structured network of constructions.

2.   Constructicon site: the local context

While firmly couched in the CxG tradition and heavily influenced by the English FrameNet Constructicon development, SweCcn is also a product of its local environment. Within the Department of Swedish, in addition to the general occupation with linguistics and language education, there are three specialized research and development units, all three of which play a role in the development of SweCcn:

– *Språkbanken* ('the Swedish language bank')
– *Lexikaliska institutet* ('Centre for lexicology and lexicography')
– *Institutet för svenska som andraspråk* ('The institute of Swedish as a second language')

Construction Grammar got introduced in Gothenburg in the early 2000s and gradually attracted more researchers and students. With inspiration from Fillmore (2008), the idea arose to build a Swedish constructicon. Around the same time, the language technology unit of Språkbanken had (1) started building a Swedish framenet, (2) developed corpus tools which are very useful for the study of constructions, and (3) was developing a resource infrastructure which a constructicon could both contribute to and benefit from. Hence, it was only natural to work together. In addition, Lexikaliska institutet, with decades of experience of dictionary development, provided valuable expertise on lexicography. Furthermore, we had long entertained the idea to apply a constructionist approach to second language education, and figured that a constructicon should be a useful resource in this regard. Therefore, the project also involves specialists on second language research.[2]

Thus, SweCcn is a collaboration between grammarians, computational linguists, lexicographers, and second language researchers. This means that SweCcn both draws upon and is adapted to the language technology resources of Språkbanken, is methodologically influenced by the dictionary production at Lexikaliska institutet, and strives to make the constructicon useful for second language education. In the following, we will focus on the resource infrastructure of Språkbanken, with which the technical possibilities and conditions of SweCcn are tightly connected.

2.1   Språkbanken

Språkbanken[3] is a research and development unit at the Department of Swedish, University of Gothenburg. It grew out of an initiative aiming at making Swedish lexicology and lexicography firmly corpus-based (Allén, 1970) and was established with Swedish government funding as a national center in 1975. Over time, the activities developed in two different directions, language technology on the one hand and dictionary production on the other, eventually branching off as separate units. The lexicography unit got formally established as *Lexikaliska institutet* ('Centre for lexicology and lexicography') in 2003,[4] whereas the language technology unit retained the name Språkbanken. As mentioned above, both units are involved in SweCcn.

---

2.   The L2 aspects of SweCcn are not a main focus in this presentation, but see e.g. Loenheim et al. (2016).

3.   < https://spraakbanken.gu.se/eng>

4.   For a historical and theoretical overview of the lexicography tradition in Gothenburg, see Rydstedt (2012, Chapter 2).

The main focus of Språkbanken's present-day activities is the development and refinement of language resources and language technology tools, and their application to research in language technology, in linguistics, and in several other disciplines, notably in the Humanities and Social Sciences, as well as in medical research.

2.2   SweFN++ – a richly structured lexical macroresource for Swedish

Lexical resources have formed a central part of Språkbanken's research activities throughout most of its history. Because of the way funding for these activities has generally been awarded – for specific projects over a limited period, typically three years – Språkbanken has grown "organically" for several decades, and each new project has had its own objectives and generally no necessary connection to other projects. After almost half a century of work on Swedish linguistic resources and Swedish lexicography, this had resulted in Språkbanken having accumulated a number of digital linguistic resources of various kinds – including both data and processing resources – with various degrees of coverage, and in various formats. The resources were rarely mutually compatible, and consequently there was no effective way to utilize the valuable linguistic information painstakingly compiled in these resources.

For this reason, Språkbanken initiated a concerted effort in 2009 on developing an open-content – i.e., freely available and modifiable – integrated lexical macroresource for Swedish (called Swedish FrameNet++, or SweFN++) to be used as a basic infrastructural component in Swedish language technology (LT) research and in the development of LT applications. The SweFN++ project had four main objectives:

1. to build a Swedish framenet (SweFN), with wide lexical coverage, on the same principles as the English Berkeley FrameNet (BFN) and to be developed in collaboration with the BFN team at ICSI Berkeley;
2. to integrate a number of existing free lexical resources (both in-house and external), by harmonizing and merging them, thereby reusing their valuable manually defined linguistic information;
3. to develop a methodology and workflow which makes maximal use of LT and other tools in order to minimize the human effort needed to build SweFN++; and
4. to use the SweFN++ resource, especially the new SweFN component, in concrete LT applications.

This endeavor has produced four components employed in the development of SweCcn: the Swedish FrameNet (2.2.1), the lexical macroresource (2.2.2), the lexical

infrastructure Karp (2.2.3), and the corpus infrastructure Korp (2.2.4). While SweCcn is not part of the SweFN++ project *per se*, it is incorporated in the lexical infrastructure as an integral part of the macroresource.

2.2.1   *Results of the SweFN++ project: SweFN*

At the time of writing (December 2017) Swedish FrameNet has over 39,000 LUs contained in close to 1,200 frames, and is thereby the world's largest framenet in terms of number of LUs.[5] As a feature unique to SweFN, it contains analyses of compound patterns in terms of frame elements being instantiated within compounds. SweFN also contains around 50 frames which do not yet exist in other framenets. Several of these frames describe nominal concepts, others are more fine-grained elaborations of frames in BFN, and a few have been created due to linguistic or cultural differences (Friberg Heppin & Toporowska Gronostaj, 2014).

SweCcn and SweFN employ the same basic editorial system, with corresponding similarities regarding description format and user interfaces. To the extent that constructicon entries in SweCcn correspond to frames in SweFN, we establish links between them (see Section 5 below). Nevertheless, they remain essentially independent resources, intended for somewhat different purposes, and are therefore less closely integrated than most other constructicons and framenets treated in this volume (see Section 8.1).

2.2.2   *Results of the SweFN++ project: The lexical macroresource*

Resource integration has turned out to be a many-faceted problem. The available lexical resources were heterogeneous as to their content and coding, having been developed for different purposes by different groups with different backgrounds and assumptions, some by linguists, some by language technology researchers – possibly with little linguistic background or none at all – and yet others in Wikipedia-like collective efforts. Thus one of the main challenges for the SweFN++ project has been to ensure content interoperability not only among the lexical resources but also between the available tools for text processing and lexical resources to be used by various pieces of software, and to formulate strategies for dealing with the uneven distribution of some types of information in the resource (e.g., syntactic valence information at present being available for about one fourth of the entries).

The resulting lexical macroresource is organized as a hub-and-spokes architecture, with one designated resource taking the central position. This is SALDO, a full-scale semantic and morphological lexicon of modern Swedish (Borin, Forsberg,

---

5.   The number of frames is on a par with BFN, while SweFN has far fewer annotated corpus examples than BFN.

& Lönngren, 2013). The information model of SALDO has been carefully designed for this purpose. Its two main item types are word senses[6] and lemgrams,[7] both identified using system-wide persistent identifiers (PIDs). Resources are interlinked either – ideally – on the content side, using SALDO's word sense identifiers, or – for practical reasons – on the form side, using the lemgram identifiers.[8] Thus, lexical units in SweCcn are represented as SALDO units, whereby they are linked to the rest of the macroresource and, by extension, to instances in corpora (see 2.2.4 below).

It is easy to achieve on the order of 80% correct sense linkages between resources automatically, simply because of the Zipfian distribution of word senses over lemmas in any lexical resource (Borin, 2010; Borin, Forsberg, & Lyngfelt, 2013). Interlinking of the most polysemous lemmas, which are also the most frequent ones in text, turns out to be a much slower and more laborious process. Work is still ongoing on utilizing the structure of the resources themselves, e.g., determining which SALDO sense should be chosen for a polysemous lemma in a Bring thesaurus class (Borin, Allwood, & de Melo, 2014) based on the semantic distances (as determined by the SALDO topology) of the alternatives to other, monosemous lemmas in the class (Borin, Nieto Piña, & Johansson, 2015).

The SweFN++ macroresource now contains, wholly or in part, the following component resources:[9]

– SALDO (Borin, Forsberg & Lönngren, 2013)
– Swedish FrameNet (Borin, Dannélls et al., 2010; see also above)
– Swesaurus (Borin & Forsberg, 2014)
– Core WordNet (Pedersen et al., 2013)
– IDS/LWT lists (Borin, 2012; Borin, Comrie, & Saxena, 2013)
– PAROLE
– SIMPLE
– Dalin's dictionary (19th c.) (Borin & Forsberg, 2011)

---

6.   Word senses are chosen as the basic content unit in SALDO, rather than, e.g., concepts as in WordNet synsets (Fellbaum, 1998), the latter being quite difficult to delimit in a satisfactory way (Margolis & Laurence, 1999; Murphy, 2002).

7.   A lemgram is a citation form together with a part of speech and an inflection table, i.e., a lemma together with its grammatical behavior.

8.   If needed, approximate sense linking can be achieved using SKOS (simple knowledge organization system) relations in order to handle non-isomorphisms between resources (e.g. words or word senses in the historical lexicons which have no counterpart in the modern language).

9.   See <https://spraakbanken.gu.se/research/swefn/publications> for references to further relevant publications.

- Old Swedish dictionaries (Borin & Forsberg, 2009, 2011)
- Swedberg's dictionary (17th c.)
- Gothenburg Lexical Database
- the Lexin dictionaries
- Bring's thesaurus (Borin, Allwood, & de Melo, 2014; Borin, Nieto Piña, & Johansson, 2015)
- The Swedish Constructicon

For many of these, integration work is ongoing. Linking historical dictionaries to modern resources raises many intricate methodological problems (Andersson & Ahlberg, 2013; Ahlberg, Forsberg, & Hulden, 2014). The lexical resources are downloadable from Språkbanken's web site as XML files structured using the ISO standard Lexical Markup Framework (ISO, 2008; Francopoulo, 2013), and some of the central resources have also been published as Linked Open Data (Borin, Dannélls et al., 2014).

2.2.3   *Results of the SweFN++ project: Karp*
Minimizing the human effort needed to build SweFN++, and also SweCcn, requires advanced technical support and an efficient methodological approach. This is the domain of Karp, Språkbanken's open lexical infrastructure (Borin, Forsberg et al., 2012), developed in the SweFN++ project: to provide an adequate support to integrate, create and curate modern and historical lexical resources.

Karp combines 32 lexical resources, including SweCcn (although it is not strictly speaking a lexical resource), which are interlinked with SALDO identifiers. It offers several kinds of search and editing functionality (Borin, Forsberg et al., 2013). Lexical information may be accessed from Karp either through a web interface intended for human users or through REST-based web services for computer program access. Karp is developed in parallel with Språkbanken's corpus infrastructure Korp (see below).

Karp offers custom-designed lexical editing functionality, used in the development and enhancement of SweFN (and SweCcn). It integrates BFN and usage examples supplied by Korp. There are facilities to semi-automatically extract frames and frame information from BFN and select lexical units from SALDO, automatically extract sentences from Korp and manually select and annotate them for semantic structure. There is also support for adding useful information on the annotation of compounds, the topic domain or other information the user wishes to emphasize.

The same basic editorial system is also used by SweCcn, adapted to handle constructional information. The information categories employed are accounted for in Section 6.

2.2.4   *Spinoff from the SweFN++ project: Korp*

Korp is the open corpus infrastructure of Språkbanken (Borin, Forsberg, & Roxendal, 2012). It provides online access to about 13.5 billion words of Swedish text corpora, out of which roughly one billion are made up of historical texts, and the remainder is dominated by social media (about 8.5 billion words) but contains corpora of a large number of genres. Korp provides a rich source of genuine linguistic examples in support of the work with the lexical resources.

Korp is characterized by a modular design, where a corpus search engine residing on one or several servers in the backend communicates with other software through well defined interfaces. As in the case of Karp, there is a web-based graphical user interface frontend for use by humans, providing sophisticated search facilities. Other computer programs – such as the Karp web interface – call the software interface directly.

The development of the corpus infrastructure started at the same time as the SweFN++ project, and an important aim informing the work on Korp has been a strong bidirectional connection to Karp. In essence, this aim involves up-to-date lexical annotations of the corpus texts, together with corpus search facilities enhanced by lexical information, for suggesting lexical entries as query terms when the user types a word in the search box, and for generating lists of semantically related words which can be used to expand the query. Conversely, this bidirectional connection facilitates the use of corpus examples and corpus statistics in the lexical infrastructure.

Korp is open-source software,[10] and has developed into a real success story. There are now Korp installations in several places in the world, providing sophisticated corpus search facilities in a number of languages. One explicit reason mentioned for choosing Korp is exactly the strong integration of the corpus search machinery with lexical resources and the ease with which a user can move back and forth between corpus search and lexicon consultation.

In SweCcn, Korp is used extensively, both as an empirical base for the construction analyses and as a source for example sentences in the constructicon entries. Some of the corpora are also employed in efforts to identify potential constructions by automatic means (see Section 3.2.1).

Figure 1 presents a graphical overview of the relationship between SweCcn and the other components of Språkbanken's lexical macrostructure, as well as its connections to corresponding resources for other languages.
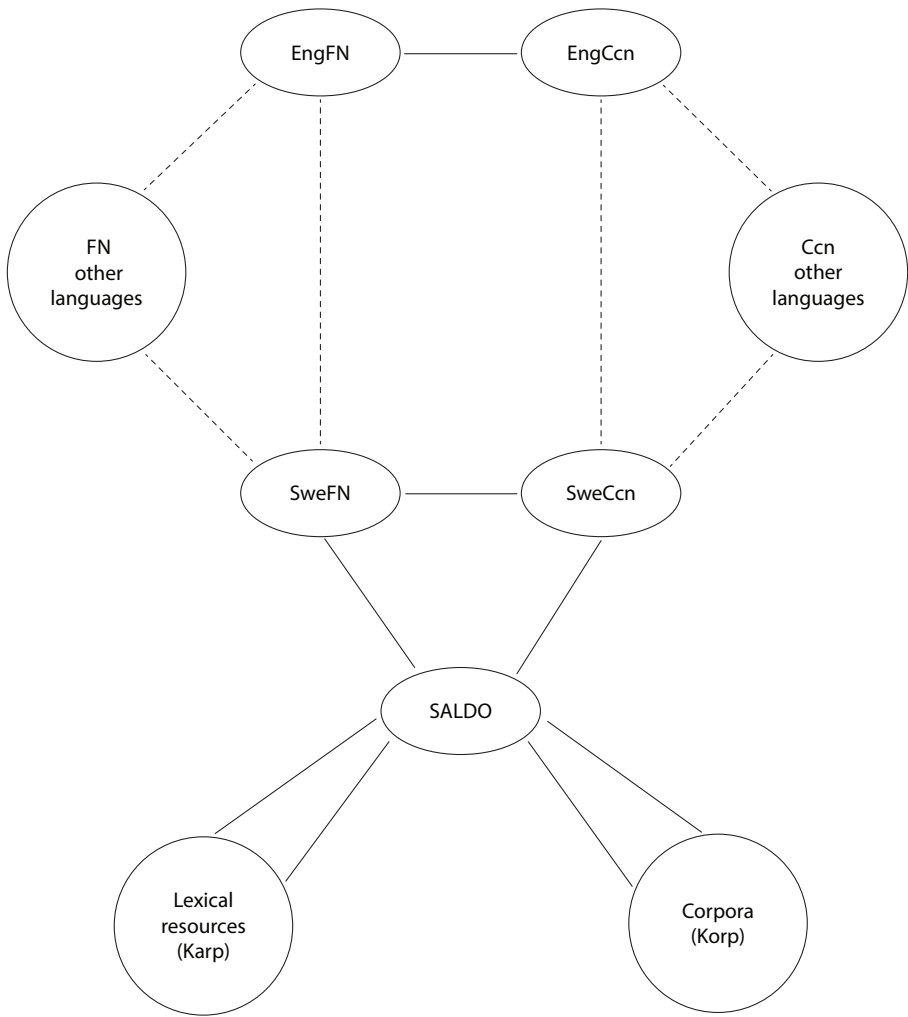
---

10. See <https://spraakbanken.gu.se/eng/korp>

Figure 1. SweCcn in relation to other resources

3.   Building the constructicon

SweCcn is intended to be a resource of broad applicability, catering to the needs
of language pedagogy as well as those of language technology. Also, to enable
some degree of size and coverage, the procedures and formalisms involved cannot
be too time-consuming. In short, a constructicon entry should be descriptively
adequate, concise, user-friendly, and formalized. Hence, there are a number of

methodological issues to consider, and the analyses and their formal description might be conceived of as an amalgamation of construction grammar, lexicography, and language technology.

In this section we outline the process from identification to presentation of a construction. Section 3.2 exhibits an overview of different methods used to identify Swedish constructions. The analysis process is described in Section 3.3 and finally some remarks on the overall structure of the database are given in Section 3.4. However, before this exploration we give a short description of the constructicon entries in the SweCcn, in Section 3.1.

3.1   Some brief notes on the constructicon entries

Every constructicon entry includes up to fifteen fields of information (for a detailed account, see Section 6 below). In this section we present the most central fields, namely *definition*, *structure sketch* and *examples*. In combination, these three represent the basic general description of the construction in terms of its grammatical structure, semantics, pragmatics and distribution. An illustration of the entry `i_adjektivaste_laget` (roughly, 'in adjectivest measure'),[11] in English translation, is given in Figure 2 (for a glossed translation of the example sentence, see (1) below).[12]

Cxn ID	`i_adjektivaste_laget`
Definition	A [phenomenon]$_{Theme}$ has a [property]$_{Property}$ of excessive proportions in relation to an (implicit) standard.
Structure sketch	[i² Adj$_{suv}$ laget]
Examples	[Jag]$_{Experiencer}$ ska erkänna att det här är [en kladdkaka]$_{Theme}$ [[i]$_{P}$ [sötaste]$_{Property}$ [laget]$_{N}$]$_{i\_adjaste\_laget}$
	[I]$_{Experiencer}$ must admit that this is [a sticky chocolate cake] $_{Theme}$ [[in]$_{P}$ [the sweetest]$_{Property}$ [measure]$_{N}$]$_{i\_adjaste\_laget}$

Figure 2. The SweCcn entry `i_adjektivaste_laget`

---

11.   Throughout, names of constructions in SweCcn are indicated in a sans serif font (`Consolas`).

12.   In the glossed examples, the following notation is used for grammatical markers: COMP comparative, DEF definite, EXPL expletive, GEN genitive, INDF indefinite, NEUT neuter, PASS passive, PL plural, PRS present tense, PST past tense, REFL reflexive, SG singular, SUV superlative, VB (derived) verb. The construct instantiating the construction in question is highlighted in boldface, where relevant.

(1) *Jag ska erkänna att det här är   en kladdkaka   i*
I   shall admit that it here be.PRS a sticky.chocolate.cake in
sötaste   laget
sweet.SUV.DEF measure.DEF

'I must admit that this is a sticky chocolate cake a bit too sweet'

The `i_adjektivaste_laget` construction in Figure 2 is a partially schematic construction (see Section 3.2). It consists of three elements of which two are lexically specific, *i* 'in' and *laget* 'the measure', and one is schematic, a superlative adjective. The definition is written in prose, on the model of dictionary descriptions.

The structure sketch is a simple, linear representation of the grammatical structure of the construction. The elements are defined in terms of part of speech, phrase type, grammatical function and/or lexical item, possibly with further specifications. Superscripts specify word sense (according to SALDO; see 2.2.2) while subscripts indicate inflection, subtype or ordinal indexing. For example, Adj_{suv} means that the construction element may be instantiated by any adjective, but only in the superlative. To indicate specific word forms we employ italics, as in *laget* (the definite singular form of *lag*) in Figure 2. The surrounding brackets delimit the construction proper; external construction elements, if shown in the structure sketch, are placed outside the brackets

To every entry we add at least three annotated example sentences, which include authentic instances of the construction in question (for reasons of space, only one example is included in Figure 2). Normally the examples are collected through the corpus tool Korp (see 2.2.4). The main purpose of the example sentences is to illustrate how the construction is used with respect to syntactic structure, information structure, variability and distribution.

All the construction elements, including external ones, are described in detail in separate fields of information (not included in Figure 2). The unique tag name given to each element is then used for the annotation in both the example sentences and the definition (if appropriate). Consequently, the annotations connect the grammatical, lexical and semantic description given throughout the entry. In the example sentences, we also annotate the construction as a whole.

3.2 Selection

Constructions are form-meaning pairings "at varying levels of complexity and abstraction" (Goldberg, 2013, p. 17), from the very general to the very specific. Much of the earliest work in CxG focused on semi-general constructions, "peripheral" patterns according to prevalent views in linguistics at the time (see Hoffmann & Trousdale, 2013, pp. 2–3) but rather towards the middle of an assumed

grammar-lexicon continuum. To some degree, a similar focus is characteristic of SweCcn, and a considerable portion of the current constructicon entries account for partially schematic patterns that combine lexical and grammatical properties.

There are several good reasons for this. One is empirical: since major language descriptions are typically either grammars or lexica, patterns that do not conform to such a modular distinction tend to be overlooked. Possibly for the same reason they also pose problems for (non-native) language learning (e.g. Wray, 2008) as well as language technology applications (e.g. Sag et al., 2002). Hence, a major purpose of SweCcn is to provide better coverage of such constructions. Furthermore, there is the methodological benefit that constructions with both grammatical and lexical properties make a good starting point for developing an adequate description format, capable of dealing with a wide range of properties. "[T]he machinery needed for describing the so-called minor or peripheral constructions […] will have to be powerful enough to be generalized to more familiar structures" (Fillmore, Kay & O'Connor, 1988, p. 534).

Note, however, that this is an area of high priority, not a delimitation of the scope of SweCcn. The database includes many constructions of various kinds, and our long-term aim is to provide good coverage of general as well as specific constructions and everything in between. The only part of the grammar-lexicon continuum that is considered outside the scope of SweCcn are lexical constructions, mainly because these are generally accounted for in lexical resources – in particular in SALDO and SweFN, with which the constructicon is interconnected.

In order to avoid too much bias towards familiar and linguistically striking constructions, we have employed several different methods to identify constructional patterns, using both primary and secondary sources. On the one hand we investigate corpora of authentic texts (3.2.1). On the other hand we extract constructional information from language descriptions such as dictionaries and grammars (3.2.2).

3.2.1   *Collecting constructions out of authentic language use*
There are several ways to extract construction candidates out of authentic language use, both manual and automatic.

Manual extraction has primarily been applied to advanced L2 student essays. These texts are advantageous for a couple of reasons. One of our long-term goals with the SweCcn project is to make the resource a useful tool for L2 education. Studying L2 learners' texts may reveal what constructions they yet fail to master. Advanced learners, although capable of producing quite complex and varied texts, are usually still lacking with respect to idiomatic accuracy, which develops late in the L2 acquisition process (e.g. Abrahamsson & Hyltenstam, 2009). Hence, these texts consist of highly developed language use but also of marked constructional

errors. To a native language speaker these unintended mistakes stand out and are easy to discover. By studying L2 texts we have thus been able to detect constructions that are problematic for L2 learners.

Many constructicon entries also derive from less systematic elicitation. Every now and then one notices the use of interesting constructions in the newspaper, on the television, or among colleagues and friends in the everyday life.

As manual methods tend to be rather time consuming, we have also developed automatic tools to identify constructions. The rest of this section presents an experiment with automatic methods to detect potential constructions in digital corpora. Utilizing the aforementioned infrastructure of *Språkbanken*, primarily the Korp and Karp tools, an experiment was set up to discover relevant construction candidates through the use of hybrid n-grams (cf. Bäckström et al., 2013; Forsberg et al., 2014). A hybrid n-gram is a linear pattern of units, comprised of specific word forms, lexical units, and grammatical categories, according to relative frequency. The method has the advantage of discovering patterns that are highly conventional and regular, which otherwise might be overlooked. In the first round of the experiment we focused on partially schematic patterns (Bäckström et al., 2013); in the second round we also included fully schematic ones (Forsberg et al., 2014).

The experiments were run on manually annotated corpora, consisting of balanced text, to extract token-phrase patterns. This operation brings excessive amounts of strings, and further algorithms were applied in several steps to prune the results. For instance, we had to distinguish patterns that are proper phrases and reject strings that are too fragmental. We also used different algorithms to establish a relevant frequency ranking (cf. Forsberg et al., 2014).

The final step in the process was to let three members of the project manually go through the top ranked results. In the end we had obtained approximately 200 relevant patterns, out of 1,200 candidates, deemed suitable for further development into constructicon entries. Thus, this method for automatic discovery turned out to be fairly successful, and it is now applicable to any corpus within Korp. Furthermore, the method does not only provide immediate construction candidates, it also brings to attention indirectly related patterns. To every candidate string were supplied examples generated from the corpus. These example sentences both illustrate the construction candidate and help to point out constructional patterns more or less related to the actual string.

An example of how a 3-gram string from the process indirectly resulted in two constructicon entries is described and illustrated in Figures 3 and 4. Although the original automatically generated string is not a construction in its own right, i.e. not in itself a conventionalized linguistic pattern, the supplied examples and some considering led to two constructicon entries. The original string is given in (2).

(2)   PP *att   göra*
      PP *to   do*

The string in (2) consists of one grammatical element, a PP, and two lexical items, *att* 'to' and *göra* 'do'. During the manual analysis of the construction candidates the staff found this pattern relevant. Although the candidate itself does not seem to be meaningful, the accompanying corpus examples (such as the one in (3a)) revealed it to be a cut off part of the pattern *ha (X) med Y att göra*, 'have (X) to do with Y'. By expanding an original somewhat inaccurate string, we are able to produce complete constructions.

   The following corpus analysis (see Section 3.4) of *ha_med*_Y_*att_göra*, also brought to light a related but somewhat different construction: *vad_har*_X_P_Y_*att_göra*. The two constructions are described in Figure 3 and Figure 4, respectively. Formally the constructions are quite similar and semantically they both concern the relation between two items. The pragmatics, however, differ considerably.

Cxn ID	ha_med_Y_att_göra
Definition	[A phenomenon]$_{Theme}$ [is about or is related to]$_{SoA}$ [a topic]$_{Corresponding}$. The construction is often used negated.
Structure sketch	NP$_1$ [ha$^1_{fin}$ (Pn) med^1 NP$_2$ att^1 göra^1] / NP$_1$ [ha$^1_{fin}$ (Pn) att^1 göra^1 med^1 NP$_2$]
Examples	Ett märkligt beslut, [jakt]$_{Theme}$ [[har]$_{State}$ [ingenting]$_{Quantifier}$ [med]$_P$ [religion]$_{Corresponding}$ [att]$_{inf-m}$ [göra]$_{SoA}$]$_{X\_har\_med\_Y\_att\_göra}$
a.	An odd decision, [hunting]$_{Theme}$ [[has]$_{State}$ [nothing]$_{Quantifier}$ [to]$_{inf-m}$ [do]$_{SoA}$ [with]$_P$ [religion]$_{Corresponding}$]$_{X\_har\_med\_Y\_att\_göra}$
b.	Vad som helst kan hända när [man]$_{Theme}$ [[har]$_{State}$ [att]$_{inf-m}$ [göra]$_{SoA}$ [med]$_P$ [teknik]$_{Corresponding}$]$_{X\_har\_med\_Y\_att\_göra}$ Anything can happen when [one]$_{Theme}$ [[has]$_{State}$ [to]$_{inf-m}$ [do]$_{SoA}$ [with]$_P$ [technology]$_{Corresponding}$]$_{X\_har\_med\_Y\_att\_göra}$

Figure 3. The SweCcn entry *ha_med*_Y_*att_göra*

(3)   a.   *Ett märkligt beslut,   jakt   **har**   **ingenting med religion***
           a   odd   decision,   hunting   have.PRS   nothing   with religion
           ***att göra**.*
           to   do
           'An odd decision, hunting has nothing to do with religion'
      b.   *Vad   som   helst kan hända   när   man **har**   **att göra***
           what   which ever can happen when   one have.PRS   to   do
           ***med teknik**.*
           with technology
           'Anything can happen when one has to do with technology'

The construction in Figure 3 expresses a relation between two elements. The THEME item is an external element whereas the CORRESPONDING item is an internal part of the construction.[13] The construction as a whole is a VP with the predicate *ha* 'have', and the preposition from the original string is invariably *med* 'with'. The grammatical elements are specified in the structure sketch, and the function of the construction is described in the definition. The order between the *med*-PP and the infinitive phrase *att göra* 'to do' is flexible, which is shown in the structure sketch and by the example sentences in (3).

   The construction *vad_har_X_P_Y_att_göra* in Figure 4 is formally a *wh*-question. Its function, however, is different from that of an ordinary question. Instead of asking for information, this construction is used to imply that the THEME element should not be associated with the CORRESPONDING element.

Cxn ID	vad_har_X_P_Y_att_göra
Definition	While expressed as a question, this construction rather implies that a [phenomenon]$_{Theme}$ should not be associated with a [topic]$_{Corresponding}$.
Structure sketch	[vad^1 ha$^1_{fin}$ NP$_1$ P NP$_2$ att^1 göra$^{1?}$]
Examples a.	[[Vad]$_{Adv}$ [har]$_{State}$ [jämlikhet]$_{Theme}$ [med]$_P$ [det]$_{Corresponding}$ [att]$_{inf-m}$ [göra]$_{SoA}$?]$_{vad\_har\_X\_P\_Y\_att\_göra}$ [[What]$_{Adv}$ [has]$_{State}$ [equality]$_{Theme}$ [to]$_{inf-m}$ [do]$_{SoA}$ [with]$_P$ [that]$_{Corresponding}$?]$_{vad\_har\_X\_P\_Y\_att\_göra}$
b.	[[Vad]$_{Adv}$ [har]$_{State}$ [Sverige]$_{Theme}$ [i]$_P$ [EM]$_{Corresponding}$ [att]$_{inf-m}$ [göra]$_{SoA}$?]$_{vad\_har\_X\_P\_Y\_att\_göra}$ [[What]$_{Adv}$ does [Sweden]$_{Theme}$ [have]$_{State}$ [in]$_P$ [the European championship]$_{Corresponding}$ [to]$_{inf-m}$ [do]$_{SoA}$?]$_{vad\_har\_X\_P\_Y\_att\_göra}$

Figure 4. The SweCcn entry *vad_har_X_P_Y_att_göra*

(4) a. *Vad har   jämlikhet med det   att göra?*
         what have.PRS equality with that to do?
         'What does equality have to do with that?'
     b. *Vad har   Sverige i EM att göra?*
         what have.PRS Sweden in EC to do?
         #'What does Sweden have to do in the European Championship?' (cf. *What's Sweden doing in the European Championship?*)

An important formal difference between the constructions is that *vad_har_X-_P_Y_att_göra* allows for other prepositions than *med* 'with', as illustrated in (4b), whereas the preposition in *ha_med_Y_att_göra* is lexically fixed. In this regard,

---

13. References to individual construction elements in plain text are marked by SMALL CAPS.

*vad_har*_X_P_Y_*att_göra* also differs from its closest English equivalent construction, as shown by the contrast in grammaticality between the English translations of (4a) and (4b).

The *vad_har*_X_P_Y_*att_göra* construction is strikingly similar, both in form and function, to the more well-known *What's X doing Y?* construction (as in *What's this fly doing in my soup?*, cf. Kay & Fillmore, 1999). Both have the form but not the function of a *wh*-question, and both contain lexically specific elements, in both cases including *vad* 'what' and *göra* 'do'. Incidentally, there is also a SweCcn entry for the Swedish construction *vad_gör_* X_Y? '*What do*-PRS X Y?', which is closely equivalent to *What's X doing Y?*.

3.2.2   *Collecting constructions from secondary sources*

There is an ample amount of constructions to be collected from secondary sources, such as grammars, dictionaries, and earlier construction studies. Throughout the years, linguists have presented a large amount of case descriptions of particular constructions, from different theoretical perspectives, providing thorough construction analyses that are often easily adapted to the SweCcn format. Notably, many SweCcn entries derive from students' essays, often written in close association to the project.

Dictionaries are systematically organized, and information concerning specific constructions and patterns is presented according to established templates, although only with respect to particular lexical items and therefore with limited capability of capturing the productivity of the patterns (cf. Lyngfelt & Sköldberg, 2013). These patterns are often excellent points of departure for identifying more general constructions. With access to digital lexicographic tools we can efficiently extract the relevant information.

Similar opportunities are provided by grammars, in particular the national reference grammar (Teleman, Hellberg & Andersson, 1999). Although different in format from the dictionary entries, the sections are sufficiently uniform to facilitate systematic elicitation. An example, *för_X_skull* 'for X's sake' is illustrated in Figure 5 (and Example 5). It was excerpt from a paragraph describing prepositional phrases with a genitive modifier in the complement (cf. Teleman et al., 1999, vol. III, p. 648).

Cxn ID	för_X_skull
Definition	An action is performed, or a state of affairs maintained, for the sake of a beneficiary or other concern of priority.
Structure sketch	VP [för[1] NP$_{gen}$/Pn$_{poss}$ *skull*]
Examples	Jag [gjorde det]$_{Event}$ [[för]$_P$ [din]$_{Beneficiary}$ [skull]$_N$]$_{för\_X\_skull}$
	I [did it]$_{Event}$ [[for]$_P$ [your]$_{Beneficary}$ [sake]$_N$]$_{för\_X\_skull}$

Figure 5. The SweCcn entry *för_X_skull*

(5) *Jag gjorde det **för din**   skull.*
     I   do.PST it   for   your   sake
     'I did it for your sake'

The construction in Figure 5 is one of several instances of the more abstract pattern P
NP$_{gen}$ N, where P can be any preposition, and NP$_{gen}$ can be any genitive noun phrase,
but the noun N is conventionally associated with the preposition. All of the P NP$_{gen}$
N constructions are highly lexicalized and most of them include a conventionalized
association between the preposition and the noun. Apart from `för_X_skull` there
are also e.g. `å_X_vägnar`, 'on behalf of X', `i_X_ställe` 'in place of X', and `på_X_be-`
`kostnad` 'on X's expense', all of which were admitted as entries in SweCcn.[14]

3.3   Construction analyses

To get an accurate picture of how the constructions are used, the constructicon en-
tries in SweCcn are corpus-based. Our main resource for the analyses is the corpus
tool Korp, which is a very valuable tool for investigating constructional patterns
in corpora, due to its rich annotation and its large number of corpora of various
genres (see Section 2.2.4).

     When a construction candidate has been identified we search for constructs in
suitable corpora. Those most commonly used contain newspaper texts (currently
594 M tokens) and blogs (currently 616 M tokens), respectively. On average we need
to study a few hundred out of some thousands of hits to achieve a sufficient basis
for the entry description, although these numbers vary a lot due to e.g. frequency,
genre and distribution. Occasionally the patterns are difficult to search for in cor-
pora, particularly when it comes to infrequent constructions. In those cases we
also employ general internet searches. Also, the role of our own language intuition
is not to be disregarded.

     A construction candidate might be reformulated or even rejected on account
of the corpus analysis. The initial pattern may have been too specific or too general
concerning particular properties. Regarding candidates from authentic language
use (except for the automatically generated ones) the initial patterns tend to be
too specific, and the variability and productivity revealed by the corpus analysis
then leads to a more general constructicon entry. Also note that, as the database is

---

14. In the case of hierarchical associations between more abstract and more specific cxns, inher-
itance relations are recorded. It should be noted that no theoretical distinction is made between
cxns at different levels of abstraction; SweCcn is in principle a multigranular resource, where all
cxn entries are considered units of the same kind.

under constant development, even a completed constructicon entry is still subject to revision in light of further data.

In this manner, the corpus analyses provide the basis for our construction descriptions, especially as regards the prose definition, the structure sketch, and the example sentences.

3.4   Organizing the constructicon entries

Building a constructicon is not only a matter of creating constructicon entries; the set of entries has to be structured somehow. Aside from the theoretical notion of a constructicon as a network of constructions, there is a practical need for both developers and users to be able to navigate the database. To this end, we are on the one hand establishing an inheritance hierarchy among the constructions, a task that is far from completed (see Section 8.2). On the other hand, the constructions are sorted into categories and types, both of which serve to identify groups of constructions (see also Section 7.1).

Every construction is assigned a basic *category*, according to the phrase type of the construction as a whole. For instance, *i_adjektivaste_laget* (Figure 2) is of category PP (prepositional phrase), *ha_med_Y_att_göra* (Figure 3) is a VP (verb phrase), and *vad_har_X_P_Y_att_göra* (Figure 4) is an S (sentence/clause). Constructions that are not specified for phrase type, such as coordination and gapping constructions, are assigned the category XP.

The constructions are also assigned to one or more *types*. This categorization is more versatile in that types can be based on any salient property shared by a group of constructions, such as

– functional properties (e.g. contrast, comparison, resultative)
– overall structure (compound, coordination)
– particular construction elements (reflexive, verb particle, implicit elements)
– 'learner focus', marking constructions that are considered particularly relevant for second language learners.

Even 'construction' is a type, consisting of all construction entries (distinguishing them from other items in the database, in particular semantic roles). Although the type attribute is not restricted to particular kinds of properties, at least not yet, they are generally based on simplex properties. Hence, we do not assume complex types such as *concessive subordinate clause*. Instead, the constructions in question are assigned two different types: *concessive* on the one hand and *subordinate clause* on the other.

The type attribute is a relatively late addition to the setup, and therefore it has not yet been consistently applied throughout the database. The current set of types

in SweCcn will surely be expanded and revised. It is also conceivable that the general role of this attribute may change somewhat as other aspects of the construction network are further developed (see Section 8.2). Nevertheless, types have proved very useful in a number of respects and will surely remain an important feature of SweCcn.

As a bonus, the type attribute shows promise as a tool for visualizing patterns of co-occurrence by combining types. For example, there are currently nine constructions of the type *implicit elements* (constructions involving null instantiation), and four of those also belong to the type *coordination*. This correlation indicates that coordination facilitates omission of various kinds. While this particular connection is no surprise, there is an obvious potential for also capturing less familiar patterns as the network grows.

4.   Idealization and variation

Descriptive linguistic resources require an amount of idealization, and there is always a tension between the aim for descriptive adequacy and practical concerns based on the purpose of the resource. A central problem in this regard is to what extent and in what way one accounts for variation. In this section, we will first discuss a few basic perspectives on descriptive adequacy, after which we account for the treatment of variation in SweCcn.[15]

4.1   Descriptive adequacy: defining grammaticality or characterizing usage?

Linguistic works aimed at non-linguists can hardly avoid the tension between descriptivism and prescriptivism. Although descriptivism is the favored approach among linguists, even strictly descriptive works are open to, and often receive, a prescriptive interpretation. Hence, a resource like a constructicon will be taken to make claims about correct usage of the constructions covered, regardless of whether this is intentional; even the mere inclusion of a certain construction may be interpreted as a claim about its appropriateness (cf. e.g. Teleman et al., 1999, vol. I, pp. 17–18, 29–30; Svensén, 2009, p. 24). Therefore, although SweCcn is primarily a descriptive resource, we have a responsibility to take such considerations into account.

---

15. Since SweCcn is primarily a monolingual resource, *variation* in this context refers to language internal variation. Regarding cross-linguistic variation between constructions, see e.g. Croft, 2001; the papers in Boas, 2010; Lyngfelt, Torrent et al., this volume).

Within the linguistic discourse, on the other hand, prescriptivism is rather frowned upon, and a basic descriptive approach is usually taken for granted. In fact, descriptive adequacy is taken to be a prerequisite for explanatory adequacy (in the sense of e.g. Chomsky, 1957). However, there are clearly more than one notion of descriptive adequacy, and the various approaches may be grouped around two very basic, and partially conflicting goals: to define what is grammatical and to characterize what is actually used (cf. Lyngfelt, Magnusson Petzell, & Wide, 2017).

To define what is grammatical is a fundamental goal for most varieties of generative grammar, in the sense of a system intended to delimit and account for "all and only the grammatical sentences" of a language, as Chomsky phrased it in 1957 (pp. 20–21), but the goal as such is way older than that. Arguably, it is implied by the very notion of grammaticality, at least if this is taken to be a sharp distinction (cf., however, Fanselow et al., 2006; Lau, Clark, & Lappin, 2014; and others). On this view, a description should ideally amount to a definition, making correct predictions about grammaticality. Such an approach requires some focus on borderline cases and peripheral structures that are rarely if ever used. Whether an expression is common or uncommon is in principle irrelevant; what matters is whether it is grammatical.

Characterizing what is actually used, on the other hand, is a core purpose of corpus linguistics and the constituting idea behind usage-based approaches to language (e.g. Langacker, 1987; Bybee, 2010). Such approaches instead tend to focus on the common, the typical, on the one hand, and linguistic variation, on the other, often addressed in terms of frequency.

Both generative and usage-based approaches are theoretically grounded in assumptions about the mind of the individual language user, regarding how linguistic patterns are acquired, stored etc. However, descriptive claims tend to be made with respect to the grammar of a whole speech community, for example (standard) Swedish. In many generative approaches this step is essentially justified by the assumption that the intuitions of individual speakers are similar enough for reliable generalizations to be made. This may in turn be verified by, e.g., grammaticality surveys or, indirectly, in corpora. According to usage-based approaches, these intuitions are presumably shaped by frequency of use in the language community. Hence, corpus data are considered a valid basis for linguistic generalizations in a more direct manner.

These are basic issues, typically resolved according to tradition (that is, the tradition in which the researcher is schooled) and in practice often disregarded. Nevertheless, *how* they are resolved has fundamental consequences for constructicon development, and in this case there is no single tradition to adhere to.

Constructicography combines grammar description with lexicography, and these traditions hold quite different positions on these matters.[16]

Grammar traditions generally tend to favor the definitional approach, assuming distinct categories, necessary and sufficient conditions, distributional tests, etc. A certain amount of idealization is deemed acceptable, but the ideal is that all legitimate instances be included (and everything else excluded). By and large, this ideal seems to hold for both generative and usage-based approaches, although it essentially reflects principles characteristic of the former. Even the latter tend to employ descriptive tools such as constraint-based representations, which expressions either satisfy or not, and typically aim for analyses which preclude potential counter examples. The main alternative would be to account for linguistic phenomena in terms of prototypes rather than distinct categories, a practice sometimes frowned upon for alleged lack of precision. Even admitting that exact definitions may be unattainable, as high a degree of precision as possible is generally the goal. Simplicity may be considered a virtue, but not at the expense of precision. Accordingly, grammatical descriptions tend to get quite detailed.

In lexicography the tradition is quite different. With the exception of language for special purposes, where technical terminology requires exact definitions, it is striking that lexicographers, although explicitly producing definitions, generally do *not* aspire to cover all and only the legitimate uses of a word.[17] Instead, dictionaries account for conventionalized uses of a word (although there may be different opinions regarding how well established a certain usage has to be before it qualifies), in practice if not in principle.[18] This practice seems to be increasingly recognized in principle as well, as reflected in the following quote from Atkins & Rundell (2008, p. 280; cf. also Hovmark, 2012):

---

16. Yet another basic issue is the relation between spoken and written language, which is becoming increasingly complex due to the development and spreading of more informal writing in electronic media. On this matter, however, there is no crucial conflict between the grammar and lexicography traditions. Accounting for the standard variety of a language basically means a focus on structures employed in both speech and writing – albeit with a slight bias towards written language, which is both more standardized and more easily available in corpora etc. In SweCcn, we adhere to this standard practice.

17. Note that this characterization concerns *lexicography*, in particular dictionary-making – as opposed to *lexicology*, the theoretical study of the lexicon. In lexicology, as well as in formal semantics, there is a wealth of work with a more generative perspective, for example Pustejovsky (1995).

18. This position seems to be more or less taken for granted, since it is hardly subject to discussion in standard works like Atkins & Rundell (2008) and Svensén (2009).

– It [a prototype approach] reflects the way people create meaning when they communicate, and thus it goes with the grain of the language, and accommodates creativity and fuzziness.
– It makes the lexicographer's task more manageable, because it allows us to focus on the prototype and its common exploitations, rather than requiring us to predict and account for every possible instantiation of a meaning.

In lexicography, a main concern is also to produce dictionaries that are user-friendly. Therefore, although both precision and simplicity are high priorities in lexicography as well as in grammar description, they tend to be valued differently in cases where they are in conflict.

When building a constructicon, one of the main challenges is accounting for grammatical phenomena under conditions typical of lexicography. The detailed analyses typical of construction grammar would be both too time-consuming and too complex for any project attempting to account for a large number of constructions, especially if there is an ambition to do so in a fairly user-friendly way (we will address the user's perspective in Section 7 below). Instead, we strive for brief, simple construction descriptions, basically constructional counterparts of dictionary entries.

However, we cannot simply adopt a lexicographer's approach, since the kinds of data to account for and especially the descriptive tools available are quite different. Although the meaning/function of a construction may be characterized by a prose definition much like the meaning/function of a word, other tools are required to represent word order, constituency, and other grammatical properties. Such tools seem to be less well suited to approximate characterizations, especially regarding word order and constituency. While grammatical categories do not necessarily have to be treated as uniform and discrete – in fact, there is evidence suggesting they should not (e.g. Croft, 2001, Chapter 1) – a phenomenon like word order does not easily lend itself to other than a linear representation, with a discrete order of elements. How variability of word order and other constructional properties are treated in SweCcn will be addressed in the following section.

4.2   Accounting for constructional variation

Aiming for brief and simple yet adequate construction descriptions, an ever-present issue is how to deal with constructional variation. In this case, idealization is not always a matter of favoring simplicity over precision. While disregarding constructional variants is misleading in the sense that it obscures the range of variability, including all possible variants of a construction may be misleading as well, at least if they are all presented on equal terms. Typically, there is one main pattern and

a number of more or less marked variants. If they all are presented as equally available, the description gives an incorrect impression of the usability of the less common variants. Unless one chooses to include frequency information, which is both time consuming to obtain and adds to the complexity of the description, the choice between including or excluding constructional variants often boils down to which alternative is the least misleading in the case in question. An always available option is also to treat the variants as constructions of their own, either as subtypes of or instead of the more general pattern.

As a general rule of thumb, we try to include central/common variants and usually exclude marginal/uncommon ones. We are also more inclined to account for variability associated with the construction itself than variation due to interaction with other constructions. Furthermore, different parts of the construction description are more hospitable to variability than others. For example, consider the construction *snarare/hellre_än_*samordning, which corresponds to the English construction *rather than* coordination (cf. Fillmore, Lee-Goldman, & Rhomieux, 2012). It consists of a preferred element, a dispreferred element, and the coordinating expression *hellre/snarare än*,[19] which can be combined in a number of different relative orders:

(6) a. *Amerikaner dricker* **hellre öl än vin**.
        Americans drink.PRS rather beer than wine
        'Americans would rather have beer than wine'
   b. *Amerikaner dricker* **öl hellre än vin**.
        Americans drink.PRS beer rather than wine
        'Americans would have beer rather than wine'
   c. **Hellre än vin** *dricker amerikaner öl*.
        Rather than wine drink.PRS Americans beer
        'Rather than wine, Americans would have beer'

Of the variants in (6), both (a) and (b) are relatively common, (a) a bit more so, while (c) is more rare. In addition, the word order in (c) is arguably due to topicalization, which may affect word order in a number of different constructions and should not have to be accounted for in every one of these (we will return to this issue below). Hence, the SweCcn entry for this construction should account for (a) and (b) but not necessarily (c).

---

19. The lexical items *hellre* and *snarare*, which both correspond to English *rather*, differ slightly in meaning: *hellre* indicates subjective preferences, whereas *snarare* marks behavioral tendencies (cf. Bäckström, Lyngfelt, & Sköldberg, 2014, p. 14, where, unfortunately, the semantic characterizations have been reversed due to a typo).

However, inclusion in a SweCcn entry is not a strictly binary feature, since an entry consists of several components. In the prose definition, both (a) and (b) are mentioned, while the structure sketch only displays (a), which is the unmarked variant. Among the annotated examples, even an instance of (c) is included.[20] On the other hand, the structure sketch accounts for yet another variant, namely the case where the dispreferred element is left out: *Amerikaner dricker hellre öl* ('Americans would rather have beer'). This possibility is marked by parentheses, as in (7):

(7)   [*snarare*/*hellre* XP$_1$ (*än* XP$_2$)]

Another illustrative example is the Swedish comparative correlative `ju_desto` construction, correponding to *the* X-er *the* Y-er in English, as illustrated in (8).

(8)   a.   *Ju högre   de   kom, desto   glesare blev   skogen.*
           JU higher they   came DESTO sparser became forest.DEF
           'The higher they went, the sparser the trees became'
      b.   *Ju klumpigare desto   bättre.*
           JU clumsier   DESTO better.
           'The clumsier, the better'
      c.   *Ju mer   soja och nötkött, desto   mindre regnskog*
           JU more soy and beef   DESTO less   rainforest
           'The more soy and beef, the less rainforest'
      d.   *Ju mer självsäker han verkar, desto   mindre självsäker är han.*
           JU more confident he   seems DESTO less   confident is he
           'The more confident he seems, the less confident he is'

In Swedish, as in English, the comparative form may be rendered by a suffix (*-are* '-er') as well as by a determiner (*mer, mindre* 'more, less'), depending on the adjectival expression. Although this variation follows from more general constructions, it is a central feature of the `ju_desto` construction and therefore has to be accounted for. Further variation regards the form of the compared expressions, which may be anything from simple adjectives to full clauses, and the possibility to add a third conjunct (*ju* X, *ju* Y, *desto* Z). There is also an amount of lexical variation, in that *ju … desto* is sometimes replaced by either *ju … ju* or *desto … desto*.

In the SweCcn entry for `ju_desto`, the structure sketch (9) accounts for the variable form of the conjuncts, including the different forms of adjectival inflection. The possibility of a third conjunct, which is uncommon but perfectly acceptable, is

---

20. Displaying more variants in the examples section than in the definition and structure sketch corresponds to a practice employed in some dictionaries, in which the valence descriptions cover the more established patterns whereas more variants are included among the usage examples.

included in the prose definition. The lexical variation, however, is not even included among the annotated examples but merely mentioned in the field for comments, since the lexical variants are not only less established but also deemed less acceptable in standard Swedish.

(9)   [*ju* Adj$_{komp}$ (XP) *desto* Adj$_{komp}$ (XP)] / [*ju mer*/*mindre* XP *desto mer*/*mindre* XP]

In a recent account of the `ju_desto` construction, Dooley (2014) presents a two-level analysis, first distinguishing the prototypical properties of the construction (basically in agreement with the SweCcn entry), and then addressing the range of further variation. Generally speaking, the ambition of SweCcn is to cover the prototypes.

As shown by the above examples, some kinds of variation are easier to incorporate than others. Had the difference between (6a) and (6b) lent itself to a more straight-forward and easily readable representation, the SweCcn entry of `snarare/hellre än_samordning` would have included both in the structure sketch. On the other hand, the lexical variation regarding `ju_desto`, which would have been easy to incorporate, was judged too marginal to warrant representation. Choosing whether to include additional variants always involves weighing the value of providing more detailed information against the cost of obscuring the general pattern as well as making the description more complex and less accessible. In this regard, representing paradigmatic variation within a single slot, such as the alternation between *snarare* and *hellre* in (7) and between *mer* and *mindre* in (9), clearly costs less, and so does marking optionality. Accounting for variation regarding the arrangement of the construction elements, by contrast, costs considerably more.

To the extent that variants can be perceived as individual patterns, they may also be treated as individual constructions, possibly as subtypes of a more general construction from which they inherit their mutual properties. In a multigrain system such as SweCcn, the conceptual difference between distinct constructions, on the one hand, and variants of the same construction, on the other, is often a matter of different levels of abstraction. Hence, constructional variation that is difficult to incorporate within a single general entry may instead be accounted for through distinct subordinate constructions.

In addition to questions regarding which constructional variants to represent, as well as how and where to do it, there is also the question of whether to indicate additional variability as such. Both of the variants represented in the structure sketch of `ju_desto` (9) have fixed word order; whereas the structure sketch for `snarare/hellre_än_samordning` (7) is subject to word order variation. Should this difference be indicated somehow? One could either treat fixed word order as default and mark points of possible variation – or treat variation as default and mark fixed sequences. In SweCcn we do neither, which means that the structure sketches are underspecified for variability. There is no formal indication whether

the structure is a prototype open to alternations or a fixed pattern.[21] Nonetheless, the structures displayed are claimed to be the standard forms of the constructions; the underspecification normally only concerns the occurrence of marginal variants (since more established variants would typically be included) and the possibility for variation due to other constructions. If deemed important enough, however, such information may be included in the definition of the construction. This practice mirrors that of lexicography; in dictionaries there is usually no indication whether usage variants beyond those represented are possible or not, but the definitions may contain information about typicality and usability.

In summary, constructional variation is treated according to the following general guidelines:

–   We include central, standard and common variants, but tend to exclude more marginal ones.
–   We account for variation associated with the construction as such rather than variation following from interaction with other constructions.
–   We are more restrictive in the definition and structure sketch of construction entries, which should display the standard, but are more generous among the annotated examples; some additional variation may also be recognized in the comments field of an entry.
–   For practical reasons, simple paradigmatic variation and optionality get represented to a higher degree than more complex variation.
–   Variant patterns may be represented as variants of the same construction, as distinct (subordinate) constructions of their own, or both.
–   Variability as such is not explicitly indicated.

5.   Constructions and frames

There is a strong historical and conceptual correlation between construction grammar and frame semantics: both originated in Berkeley, developed by Charles Fillmore and associates, both are characterized by a holistic view of linguistic units, and the semantic component of CxG analyses has often been represented in terms of frame semantics. This correlation is even stronger in the subsequent development of descriptive resources based on these theories: framenets and constructicons. The first constructicon database, the English FrameNet Constructicon, was created

---

21. Some indication of variability follows from inheritance relations, however, especially from the category assignment. Unless specified otherwise, properties of general constructions such as NP, PP, etc. presumably apply by default to more specific instances of these categories.

as an appendix to the English FrameNet (Fillmore, Lee-Goldman & Rhomieux, 2012) to accurately depict constructions that cannot be accounted for by lexically associated frames alone. Following this initiative, framenets and constructicons for other languages are designed to complement each other, typically even construed as integrated parts of the same database (e.g., Ohara, this volume; Torrent et al., this volume). Thus both kinds of resources are intentionally designed for cross referencing, and the interlinking between them is facilitated by the similar description formats.

While this interrelation is somewhat weaker in the case of SweCcn and SweFN, which are in principle independent resources, they are nonetheless intended to be compatible both with each other and with their counterparts for other languages. One important undertaking in this regard is to establish links, where applicable, between constructicon entries in SweCcn and their corresponding frames in SweFN (and, by extension, framenets in general) (cf. Ehrlemark, 2014). There are many potential advantages in doing so, besides the virtue of general compatibility, for example the possibility to compare constructions from different languages via the frames they evoke (Bäckström, Lyngfelt, & Sköldberg, 2014; cf. Boas, 2009).

At the time of writing, about half of the entries in SweCcn are linked to frames in SweFN. The continuing work with comparing and linking the two resources does not aim to link all constructions to frames, but rather to distinguish frame-bearing cxns from non frame-bearing cxns. Future work also concerns systematic distinctions between different kinds of construction-to-frame relations. After a quick introduction to FrameNet and some general remarks about linking (Section 5.1), we illustrate some frame-bearing cxns in Section 5.2 and turn to non frame-bearing cxns in Section 5.3. It should be noted that the distinction between the two sets is by no means absolute but follows from the kinds of frames assumed and the kinds of relations recognized.

Frame-bearing and non frame-bearing cxns are also discussed by Ohara (this volume), based on Japanese data and with a slightly different perspective than our approach. She proposes a five-way classification of cxns: those invoking semantic frames (ordinary FN frames), those invoking interactional frames (a novel category proposed by Ohara), and three types of non frame-bearing cxns.

5.1   Linking constructions and frames

A framenet is a lexical-semantic resource in which lexical units are defined by the cognitive frames they presumably evoke in the mind of the language user. The frames can be described as schematic scenes populated by frame-unique elements (FEs) like participants, objects or states of affairs, and represent the background

knowledge needed to make sense of language. In an attempt to mimic a cognitive model of how meanings are related to each other, frames are explicitly connected via frame-to-frame relations such as inheritance, causation, precedence or perspective.

To illustrate, the frame `Kidnapping` is described as a situation where a PERPETRATOR carries off and holds a VICTIM against his or her will by force.[22] Lexical units like *shanghai* (v), *snatcher* (n) and *abducted* (a) all evoke the `Kidnapping` frame and example sentences containing these words are annotated with frame elements, to exemplify different distributional patterns of the LUs in the frame, as in Figure 6.

Frame	Kidnapping
FEs	PERPETRATOR, VICTIM
Example	[I]$_{Perpetrator}$ was planning to [abduct]$_{LU}$ [Kirsty]$_{Victim}$

Figure 6. `Kidnapping` frame[23]

The `Kidnapping` frame is related to other event frames in the network, like `Committing_crime`, that defines the role of the perpetrator as a villain intentionally breaking the law, and `Getting`, that describes the event of coming to possess something one did not have before. A framenet matrix thus represents meaning as a network of related cognitive scenes and keeps track of the hierarchical relations between frames.

The target of annotation in framenet-style text analysis is always the frame-evoking lexical unit (including some multi-word expressions), usually verbs, nouns, adjectives or prepositions. But also more complex constructions can correspond to frames, in which case the correspondences are represented by links between SweCcn and SweFN. The most straightforward relation would be that a frame-bearing construction evokes a frame in the same manner as a lexical unit, with the only difference that the frame-evoking unit is itself complex (Fillmore, Lee-Goldman, & Rhomieux, 2012). Prototypical frame-bearing constructions have a clearly defined referential meaning that corresponds to the frame description of the target frame. In the ideal case there is also full correspondence between the construction elements and the (core) frame elements.

---

22. Analogous to constructions, frame names are in a sans serif font (`Consolas`); frame elements are indicated by SMALL CAPS with an initial capital letter.

23. All frame definitions in this section are taken from the Berkeley English FrameNet, see <https://framenet.icsi.berkeley.edu/fndrupal/home>. The same definitions also apply to SweFN, which generally adopts the frames and frame definitions from BFN, except for a few frames unique to SweFN (Friberg Heppin & Toporowska Gronostaj, 2014).

However, some constructions evoke frames, some do not, and some correspond to frames in other ways. In the terminology of Fillmore, Lee-Goldman, & Rhomieux (2012), frame-evoking constructions are called *frame-bearing*. We adopt this term here, but expand the notion somewhat to also include other correspondences between constructions and frames than the prototypical evoking relation. Sometimes the frame and the construction apply to the same situation but from different perspectives or portray different profilings. In other cases the frame corresponds to a domain restriction on the construction rather than the constructional meaning as such. At present we do not distinguish systematically between such relations, but simply classify all constructions linked to frames as *frame-bearing*. In due course, we expect to define a more fine-grained set of relations.

There is, however, one kind of link that is handled separately, namely the association between cxn elements and semantic roles. Roles in SweCcn are features assigned to construction elements; they are defined globally and thus constitute individual entities in the database.[24] Since semantic roles are not isolated properties but rather relations between an entity and a scene (cf. Rydstedt, 2012), they essentially correspond to frame elements. Therefore the roles in SweCcn are defined as such, by links to frame elements in very general frames; see Section 6.2.2.

The prototypical case of a non frame-bearing construction is a grammatical pattern with precise formal features but no referential meaning, like constructions for passive voice or gapping. Other constructions have a pragmatic meaning or function that is not satisfactory captured by FrameNet frames, such as specifications of modality or information structure. Such constructions certainly mean something, but their meaning lies beyond the (current) scope of FrameNet. Alas, there are certainly cases where one has to ask – does the meaning of this construction correspond to a frame that does not exist in FrameNet *yet*?

5.2   Frame-bearing constructions

In this section we will present some examples of frame-bearing constructions roughly organized in tentative groups, not by any means definite or mutually exclusive.

---

24. Semantic roles are the only entries in SweCcn which are not constructions, since they are only defined in terms of meaning and therefore do not constitute form-meaning pairs. The association between form and meaning takes place through the construction elements that instantiate the roles.

5.2.1   *Argument structure constructions*

Many frame-bearing constructions are argument structure constructions with verbal heads, with a potential of altering the verb's inherent valence. These constructions refer to types of events, which, arguably, is closely equivalent to evoking a frame. Depending on the degree of schematicity, the frame evoked will depict a more or less general scenario.

A typical example is the `reflexiv_resultativ` 'reflexive_resultative' construction, as in *äta sig mätt* 'eat oneself full' (Figure 7). This construction roughly means 'achieve something by V-ing' and occurs with both transitive and intransitive verbs, potentially altering the verb's inherent valence restrictions. The syntactic structure of the construction is [V Pn$_{refl}$ AP] and the construction elements have the semantic roles of Actor/Undergoer, Activity and Result.[25] Typical instantiations include *dricka sig full* 'drink oneself drunk' and *springa sig varm* 'run oneself warm', while a superficially similar example like *känna sig trött* 'feel tired, lit. feel oneself tired' does not fit the description since it does not mean 'get tired by feeling'. (cf. Lyngfelt, 2007, pp. 109–110)

Construction	reflexiv_resultativ
Structure sketch	[V Pn$_{refl}$ AP]
CEs	NP, PN.REFL, ACTIVITY, RESULT
Example	[Hon]$_{NP}$ [skrek]$_{Activity}$ [sig]$_{Pn.refl}$ [hes]$_{Result}$
	[She]$_{NP}$ [screamed]$_{Activity}$ [herself]$_{Pn.refl}$ [hoarse]$_{Result}$

Figure 7. The `reflexiv_resultativ` construction

The `reflexiv_resultativ` construction corresponds to the `Causation` frame, defined as a situation where a Cause or an Actor causes an Effect. The entity Affected by the causation matches the reflexive. The frame contains lexical units like *force (v), make (v), result (n)* and *responsible (a).* In framenet terms, it is the lexical unit that evokes the `Causation` frame, as in the first example in Figure 8.

In the case of `reflexiv_resultativ`, however, it is the construction as a whole that evokes the causative meaning, with no specific lexical material standing in for the frame-evoking unit. To control whether the construction fits the frame we take the same sentence as above in Figure 7 and test if it is possible to annotate it with the frame elements of the target frame in Figure 8. Since the construction elements

---

25. The CEs are labeled by their most salient distinctive characteristic, which in the case of the reflexive and the external NP argument is their category (they both have the role of Actor/Undergoer, and the reflexive element is a highly prominent feature of the cxn). The other two elements are annotated as Activity and Result, that is, by their semantic roles (their categories are V and AP, respectively).

overlap neatly with the frame elements, we can safely conclude that the construction
is frame-bearing and establish a link between the two respective entries.

Frame	Causation
FEs	ACTOR, AFFECTED, CAUSE, EFFECT
Example: LU	[Screaming]$_{Cause}$ [made]$_{LU}$ [her]$_{Affected}$ [hoarse]$_{Effect}$
Example: Cxn	[She]$_{Actor}$ [screamed]$_{Cause}$ [herself]$_{Affected}$ [hoarse]$_{Effect}$

Figure 8. Causation frame

A similar case is the `indirekt_kausativ_bort` 'indirect_causative_*away*' con-
struction, as in *äta bort sin huvudvärk* 'eat one's headache away', which can be
instantiated by almost any verb-object relation where the ACTIVITY indirectly leads
to the removal of a THEME, as in Example (10) and Figure 9. The verb-particle *bort*
'away' combines with a verb to achieve the causative removing relation between the
activity and the theme. (cf. Sjögreen, 2015)[26]

(10) *Hon spelade   bort sina   sparpengar.*
      she   gamble.PST away her.REFL savings
      'She gambled her savings away.'

Construction	Indirekt_kausativ_bort
Structure sketch	[V *bort* NP]
CEs	ACTOR, ACTIVITY, THEME, RESULT
Example	[Hon]$_{Actor}$ [spelade]$_{Activity}$ [bort]$_{Result}$ [sina sparpengar]$_{Theme}$
	[She]$_{Actor}$ [gambled]$_{Activity}$ [away]$_{Result}$ [her savings]$_{Theme}$

Figure 9. The `indirekt_kausativ_bort` construction

Our analysis concludes that the construction evokes the Removing frame, in which
an AGENT causes a THEME to move away from a SOURCE. But unlike the Removing
frame, the construction has no correspondent to the frame element SOURCE. This
is a case of null instantiation (NI) – indicating that the source is implicit. Figure 10
illustrates how the Removing frame may be evoked by either an LU such as *vräka*
'evict' or by the `indirekt_kausativ_bort` construction.

---

26. This fairly general Swedish *bort* 'away' construction subsumes instances corresponding to
the well-known English time *away* construction (cf. *Dancing the night away*, Jackendoff, 1997),
although this particular function is not as strongly conventionalized as a construction in its own
right in Swedish.

Frame	Removing
FEs	AGENT/CAUSE, THEME, SOURCE, GOAL
Example: LU	[Hyresvärden]$_{Agent}$ vill [vräka]$_{LU}$ [nazisterna]$_{Theme}$ [NI]$_{Source}$
	[The landlord]$_{Agent}$ wants to [evict]$_{LU}$ [the nazis]$_{Theme}$ [NI]$_{Source}$
Example: Cxn	[Hon]$_{Agent}$ [spelade]$_{Manner}$ [bort]$_{Goal}$ [sina sparpengar]$_{Theme}$ [NI]$_{Source}$
	[She]$_{Agent}$ [gambled]$_{Manner}$ [away]$_{Goal}$ [her savings]$_{Theme}$ [NI]$_{Source}$

Figure 10. Removing frame

A more complex case concerns the group of reflexive verb particle constructions of the form [V *upp* REFL], i.e. a verb combined with the particle *upp* 'up' and a reflexive. This general pattern is semi-productive, but its specific subtypes display a wide range of mostly unrelated meanings, as in *stressa upp sig* 'get stressed', *lösa upp sig* 'dissolve', *mopsa upp sig* 'get cocky', *piffa upp sig* 'dress up' and *jobba upp sig* 'work one's way up'. Since the various instances form distinct groups, we have created five particular construction entries. Some of these have matching frames in FrameNet, such as the Progress frame for verba_*upp_sig*.höja_rang 'verb_*up_oneself*.higher_rank' or the Dispersal frame for verba_*upp_sig*.upplösa 'verb_*up_oneself*.dissolve'. For others, however, there is no good match available. Short of adding new frames, which accurately depict the semantic constraints of the constructions, our options are either to resort to inferior matches or to refrain from linking these particular constructions to frames.

Currently, there is no general verba_upp_sig 'verb up oneself' construction in SweCcn. Such a construction would have a very general semantic characterization, with a range of which only small portions are actually instantiated.[27] Disregarding the directional component indicated by *upp* 'up', which has several different metaphoric interpretations, the most salient semantic property seems to be 'change of state'. Assuming such a construction, it could be linked to the frame Change_of_state_scenario. In that case, even those subconstructions lacking individual matches would be indirectly linked to FrameNet by inheritance.

5.2.2   *Formulas*

Many constructions define the syntactic and morphological regular patterns to express unit relations for time, place, measures, proportions and dates. They are formulas with slot fillers that link them with certain domains. The formal constraints of the construction do not always overlap with the semantic constraints, and therefore a single construction may be linked with several different frames.

---

27. Given its partial productivity, [V *upp* REFL] is the kind of structure sometimes referred to as a *pattern of coining* rather than a (fully productive) construction (e.g. Kay, 2013).

Such is the case with the pair of rate constructions `proportion_i_om` and `proportion_per`, distinguishing two different syntactic patterns for expressing proportions in Swedish. In both cases, the construction combines two entities, a numerator and a denominator, joined by a preposition, but they differ regarding choice of preposition, definiteness on the denominator noun phrase, and domain of use. In terms of form, the prepositions *i* 'in' and *om* 'about' occur with a definite denominator, whereas *per* 'per' occurs with an indefinite. In terms of meaning, the construction `proportion_i_om` is restricted to temporal relations, and therefore corresponds to frames such as `Frequency` (11a) and `Speed_description` (11b). The construction `proportion_per` is a more general construction that in addition to `Frequency` (12a) and `Speed_description` (12b) also expresses ratio relations regarding `Relational_quantity` (12c), and `Price_per_unit` (12d).

(11) a. *två gånger om dagen*
   two time.PL about day.DEF
   'twice a/per day'

   b. *120 km i timmen*
   120 km. in hour.DEF
   '120 km. an/per hour'

(12) a. *fem lektioner per vecka*
   five lecture.PL per week.INDF
   'five lectures a/per week'

   b. *20 meter per sekund*
   20 meter.Ø per second.INDF
   '20 meters per second'

   c. *tjugo datorer per klassrum*
   twenty computer.PL per classroom.INDF
   '20 computers per classroom'

   d. *10 kronor per kilo*
   10 krona.PL per kilo.INDF
   '10 Swedish kronor a/per kilo'

Accordingly, `proportion_i_om` is linked to two frames and `proportion_per` to four. Note that these links do not represent ordinary evoke relations, but rather indicate domains of applicability. The basic meaning of both constructions is the proportional numerator-denominator relation as such, while speed, frequency etc. are different domains to which this relation is applied.

By comparison, the corresponding English rate constructions are organized differently in the English FrameNet Constructicon, with four different constructions split by domain/frame (*cost-time, frequency, mileage* and *speed*) and disregarding formal differences (*a/an* vs. *per*).

5.2.3   Grading constructions

Grading constructions are constructions that are used for things like comparing, enhancing and evaluating attributes. They are common in use and usually easy to link to corresponding frames in the cluster of frames under Gradable_attributes.

For example, the two comparison constructions jämförelse.likhet 'comparison.equality' and jämförelse.olikhet 'comparison.inequality' both evaluate the relation between two entities with respect to some attribute, and both can be linked to the frame Evaluative_comparison. However, while the Evaluative_comparison frame is structured around LUs like *compare (v), rival (v)* and *measure up (v)*, the constructions instead profile the comparative value of the attribute. The inequality construction is formed by a comparative adjective or adverb plus *än* 'than' (13), while the equality construction is expressed by the pattern *lika* AP *som* 'as AP as' (14). In Figure 11 it is shown how the FrameNet annotation fails to capture the comparative elements of the comparison constructions.

(13) *Hon är   smartare   än   de   flesta.*
she   be.PRS   smart.COMP   than   the.PL   many.SUV.DEF
'She is smarter than most people'

(14) *Hon spelar   gitarr **lika dåligt som du**.*
she   play.PRS   guitar   like   badly   as   you
'She plays the guitar as badly as you do'

Frame	Evaluative_comparison
FEs	PROFILED_ITEM, STANDARD_ITEM, ATTRIBUTE, PROFILED_ATTRIBUTE, STANDARD_ATTRIBUTE
Example: LU	[Ingen av svenskarna]$_{Profiled\_item}$ kan [mäta sig]$_{LU}$ med [norrmannen]$_{Standard\_item}$. [NI]$_{Attribute}$
	[None of the Swedes]$_{Profiled\_item}$ can [rival]$_{LU}$ [the Norwegian]$_{Standard\_item}$. [NI]$_{Attribute}$
Example: Comparison Inequality	[Hon]$_{Profiled\_item}$ är [smartare]$_{Attribute}$ än [de flesta]$_{Standard\_item}$.
	[She]$_{Profiled\_item}$ is [smarter]$_{Attribute}$ than [most people]$_{Standard\_item}$.
Example: Comparison Equality	[Hon]$_{Profiled\_item}$ [spelar gitarr]$_{Attribute}$ lika dåligt som [du]$_{Standard\_item}$.
	[She]$_{Profiled\_item}$ [plays the guitar]$_{Attribute}$ as badly as [you]$_{Standard\_item}$ do.

Figure 11. Evaluative_comparison frame

The `Evaluative_comparison` frame best corresponds to a general comparison construction. In the Swedish Constructicon we have organized the comparison constructions in the same way as in the English FrameNet Constructicon, with a parent construction `comparison` and two children that distinguish the different patterns for equality and inequality comparisons. In the Japanese FrameNet the researchers have chosen to complete this distinction by adding a `Comparison_in-equality` frame to the network (Ohara, 2014). This is yet another example of how the two resources influence each other and how difficult it is to arrive at a level of generality that satisfies the goals and aims of both constructicons and framenets.

5.2.4   *Figurative constructions/idioms*

Figurative constructions and productive idioms have a transferred meaning that cannot be inferred from the individual components alone. Linking such constructions to frames is a way of pointing out the target domain of the metaphor, although the correlation does not usually extend to analogous associations between the construction elements and frame elements. For example, the Swedish expression *inte den vassaste kniven i lådan* 'not the sharpest knife in the drawer' roughly means that somebody is stupid, clueless, or incompetent. The construction *kniven_i_lådan* 'the_knife_in_the_drawer' is thus a humorous understatement, since 'not the sharpest' is interpreted as 'pretty blunt'. On the surface it takes the form of a comparison, just as 'the smartest girl in the class', but it stands out not only in its conventionalized metaphoric meaning but also in requiring a negation and in restricting the adjective to synonyms for smart, bright or competent. As it is a partially productive idiom, knives and drawers can be replaced by other objects in other closed confinements, as in the example in Figure 12.

Construction	kniven_i_lådan
Structure sketch	Neg [Det$_{def}$ Adj$_{sup}$ N$_{def}$ i^2 NP$_{def}$]
CEs	PROPERTY, ENTITY, LOCATION, NEG, DET, P
Example	[Han]$_{Entity}$ är [inte]$_{Neg}$ [den]$_{Det}$ [ljusaste]$_{property}$ [kritan]$_{Entity}$ [i]$_P$ [asken]$_{Location}$
	[He]$_{Entity}$ is [not]$_{Neg}$ [the]$_{Det}$ [brightest]$_{property}$ [crayon]$_{Entity}$ [in]$_P$ [the box]$_{Location}$

Figure 12. The *kniven_i_lådan* construction

Since this construction has nothing to do with sharpness, knifes or drawers, but is rather a judgment about somebody's mental abilities, we choose to link it to the frame `Mental_property`. The `Mental_property` frame accurately describes how mental properties may be attributed to a person by a (usually implicit) judge. As you can see

in Figure 13, this gives a good hint that the whole expression should be interpreted analogously to a more literal description about somebody's (lack of) intelligence.

Frame	Mental_property
FEs	PROTAGONIST, BEHAVIOUR, PRACTICE
Example: LU	[Han]$_{Protagonist}$ är en [klipsk]$_{LU}$ ung man.
	[He]$_{Protagonist}$ is a [smart]$_{LU}$ young man.
Example: Cxn	[Han]$_{Protagonist}$ är inte [den ljusaste kritan i asken]$_{Property}$
	[He]$_{Protagonist}$ is not [the brightest crayon in the box]$_{Property}$

Figure 13. Mental_property frame

5.3   Some non frame-bearing constructions

Not all constructions have frame-like meanings, but what exactly is 'frame-like'? Distinguishing non frame-bearing constructions is a way of evaluating FrameNet, and in the process we are faced with questions about the limits of framenet-style meaning representation. In this section we present a few examples of constructions that have not been linked to FrameNet and that may be considered non frame-bearing. Typically, these constructions have less referential meaning, in the sense of referring to actual scenes or entities, but are primarily characterized by grammatical or pragmatic functions.

Prime examples include regular syntactic patterns to organize constituents into phrases and clauses, or words in inflectional categories and paradigms. This group includes general grammatical constructions for predication, complementation, modification and phrase structure (few of which are included in SweCcn at this point) as well as syntactic configurations such as passives (15a), interrogatives (15b), and information packaging constructions such as clefts (15c).

(15)   a.   *Boken   skrevs   av hans   fru.*
             book.DEF write.PST.PASS by his   wife
             'The book was written by his wife'

       b.   *Skrev   hans fru   boken?*
             write.PST his   wife book.DEF
             'Did his wife write the book?'

       c.   *Det   var hans fru   som   skrev   boken.*
             EXPL was his   wife which write.PST book.DEF
             'It was his wife who wrote the book'

Certainly, passives refer to the same situations as their active counterparts, only from a different perspective, and the same holds for polarity questions and their corresponding declaratives as well as for cleft sentences and simple clauses. Therefore their differences in meaning do not seem to be matter of what frames they evoke – at least not unless we assume frames of a very different kind than those in FrameNet, possibly straying away from the basic ideas of frame semantics.[28]

Another case in point are reflexive constructions, which indicate various relations centered around the prototypical function of marking coreference between two co-arguments. Consider for example the deobjective reflexive construction in (16), where the reflexive complement of a communicative verb indicates a metonymic relation between the subject/communicator and the object/message, thereby obscuring the actual content of the object argument (Lyngfelt, 2007, p. 105). The construction resembles English expressions like *express yourself*, but is much more productive.

(16) *När Gardner pratar är   han noga   med att **precisera sig**.*
    when Gardner talk.PRS be.PRS he careful with to precise.VB REFL
    'When Gardner speaks he is careful to specify what he means'

While the verbs in this particular construction clearly evoke the Communication frame, they do so regardless of the reflexive and the relation has little to do with the reflexive construction as such. What the deobjective reflexive construction contributes in terms of meaning are the metonymic relation and the deprofiling of the message, neither of which would be captured by a link to Communication. Other reflexive constructions, which are less restricted in terms of domain, would be even harder to provide with an adequate characterization in terms of frames.

Other constructions license diversions from regular argument structure patterns, like `adjektiv_som_nominal` 'adjective as nominal' in (17) and `nominal_som_adjektiv` 'nominal as adjective' in (18). Such constructions are often characterized in terms of coercion, type-shifting, or null instantiation (of an implicit head), neither of which are typically frame-like phenomena. The same applies to

---

28. Petruck & Ziem (2014) suggest that this kind of perspectivization may be accounted for by some version of the *Perspective_on* relation used in FrameNet to link, for example, the frames Commerce_sell and Commerce_buy via the more general frame Commerce_goods-transfer (cf. Ruppenhofer et al., 2016, pp. 82–83). Provided that some technical issues are dealt with, for instance the absence of clausal constructions that are neutral with respect to voice, clause type etc., this frame-to-frame relation might be adapted to handle relations between constructions. It is less clear, however, how it would translate into a relation between constructions and frames. In any case, linking general clausal constructions to frames seems to require the creation of some very abstracts frames, such as Predication.

derivational constructions such as `exocentric_adj.sammansättning` 'exocentric adjectival compound' (19), formed by an adjective, a noun, and a participial suffix.

(17) *Kvinnor   är   överrepresenterade bland   **de   fattiga**.*
      woman.PL   be.PRS   overrepresented.PL   among   the.PL   poor.PL
      'Women are overrepresented among the poor'

(18) *Ironi är   så   **90-tal**.*
      irony be.PRS so   90-number
      'Irony is so 90's'

(19) *Riv   dem   och   lägg över i   en   **tjockbottnad**   gryta.*
      grate them and   lay   over in a   thick-bottomed   pot
      'Grate them and put in a thick-bottomed pot'

There are also many constructions used to avoid redundant repetition, omitting otherwise obligatory constituents whose content is provided by the context. These include `ellips.komplement` 'shared completion' (20) and `ellips.fragment` 'stripping' (21).[29] Since such constructions mostly concern the form of a sentence, it is hard to see any kind of frame evoked by them.

(20) *Det   är   förbjudet   att både **köpa och sälja droger**   i   Sverige.*
      EXPL be.PRS   forbidden   to   both   buy   and   sell   drug.PL   in   Sweden
      'Both the buying and selling of drugs are forbidden in Sweden'

(21) *Han brukade äta kött   men **inte nu   längre**.*
      he   use.PST eat meat but   not now   longer
      'He used to eat meat, but not anymore'

As shown by these examples, regular grammatical patterns are typically hard to account for in terms of frames. A perhaps better candidate, however, is the integrated appositive construction illustrated in (22). Arguably, its main function is that of specification, and it may therefore be claimed to evoke a `Specification` frame (although no such frame currently exists in FrameNet).

(22) *Hon bor   i   Stockholm, **Sveriges   huvudstad**.*
      she   live.PRS in Stockholm Sweden's   capital
      'She lives in Stockholm, the capital of Sweden'

---

29. While the SweCcn names of these constructions literally translate into *ellipsis.complement* and *ellipsis.fragment*, respectively, the labels *shared completion* and *stripping* are the names of the corresponding English constructions in the Berkeley Constructicon (cf. Fillmore, Lee-Goldman, & Rhomieux, 2012, p. 327).

If we were to create such a frame, it is arguably also evoked by, e.g., relative clauses (*Stockholm, which is the capital of Sweden*) and copular constructions (*Stockholm is the capital of Sweden*) and would presumably inherit from an abstract `Predication` frame. While this is of course conceivable, these constructions do not relate to the frames in the same way as an LU does. Whereas a verb like *specify* refers to a specification scene, appositives and relatives constitute specifications in themselves. Although this may still be considered a construction-to-frame relation, it is markedly different from evoking.

Turning to constructions with more of a pragmatic function, we encounter a similar lack of correspondence between constructions and frames, often due to a difference in perspecive. Take for example the exclamatives in (23), formed by an initial adverb or pronoun and an XP.

(23) a. *Vilken dag!*
          which day.DEF
          'What a day!'
      b. *Vad tiden går!*
          what time.DEF go.PRS
          'How time flies!'
      c. *En sådan gullig kattunge!*
          a such cute kitten
          'What a cute kitten!'

We might call them degree markers, but they do not conform to either `Degree` or `Beyond_compare` since both of those frames are based on a gradable attribute or variable, which is usually absent in these constructions. 'What a day!' can mean that the day was terrible, wonderful, boring, or whatever. We could probably get a better match by creating an `Expressivity` frame, but would still be left with the same kind of difference in perspective as regarding the appositive in (22).

Another illustrative case is `instruerande_passiv` 'instructing_passive' (24), which stands out from other passives by its deontic modality.[30] Although it has the form of a statement, its intended reading is that of a request.

(24) *Severas kyld*
      serve.PASS chilled
      'To be served chilled'

There is a `Request` frame in FrameNet, evoked by lexical units like *demand, beg, ask,* and *instruct*, but its relation to imperatives, let alone these passives, is less

---

30. The omitted subject in (24) is a genre-specific property of so-called labelese (cf. Ruppenhofer & Michaelis, 2010).

straight-forward. One way to describe the connection is that the construction as a whole represents a single frame element: the MESSAGE. Another is that the construction does not *refer* to a request, as LUs do, but rather *perform* it – again a difference in perspective.

Some constructions are distinguished by the way they behave in interaction, such as the reactive constructions in (25)–(26). First, we have an interesting type of coordination cxn (25).

(25) **Kul och kul**, *så roligt var   det inte*
        fun and fun so fun   be.PST it   not
        'Actually it wasn't that fun'

The construction illustrated in (25) is called `reaktiv_X_och_X` 'reactive x and x'. It picks up an expression X from the previous utterance and repeats it twice, normally followed by a clarification, to indicate that the expression X is not quite adequate in the present context (Lindström & Linell, 2007). Any part of the utterance can be singled out for renegotiation this way. A simple sentence like *I like my car* may be reconsidered in several ways: *Like and like, actually I love it* or *Car and car, it's more like a wreck* or *Mine and mine, it's actually my mother's*.

Something similar is going on in the `reaktiv_dubbel_aux` 'reactive double auxiliary' construction (26), except that it is restricted to auxiliaries and does not involve coordination (Linell & Norén, 2009). An expression from the preceding sentence – in this case the auxiliary – is repeated twice, followed by a corrective specification; for instance, (26) could be a response to *Can you do the dishes?*.

(26) **Kan kan** *jag väl,   men jag vill   inte*
        can can I   surely but I   want not
        'Sure I can, but I don't want to'

Both constructions, which have similar functions, are probably best characterized in terms of their role in interaction, but that cannot be captured by any existing communicative frames. The closest match would probably be `Concessive`, corresponding to the relation between the repeated part and the clarifying continuation. However, even disregarding the difference in perspective, this only pertains to the internal structure of the constructions and fails to capture the key interactional aspects.

The constructions in (25)–(26) both involve reduplication, a feature commonly used across languages for enhancing significance, duration, etc. or to partition a process. A few Swedish examples are the constructions `upprepad_komparativ` 'repeated comparative' (27a), `X_efter_X` 'x_*after*_x' (27b) and `X_för_X` 'x_*by*_x' (27c).

(27)   a.   *Han simmade **närmare   och närmare   ön.***
           he   swim.PST close.COMP and close.COMP island.DEF
           'He swam closer and closer to the island'
       b.   *Samma dröm   återkommer **natt   efter natt.***
           same   dream   return.PRS   night after night
           'The same dream returns night after night'
       c.   *Jag blir   bättre   **dag för dag.***
           I   become.PRS well.COMP day for day
           'I get better day by day'

These reduplication patterns express (gradual) augmentation (27a), iteration (27b), and stepwise development (27c), respectively. Neither corresponds to existing frames, although one can perceive loose associations to frames like `Activity_ongoing`, `Degree` and `Change_position_on_a_scale`. Assuming new frames, for instance `Iteration`, would provide a better fit but would still leave us with the difference in perspective noted for so many other constructions.

To conclude, many of the constructions currently considered non frame-bearing could be linked to frames if we assume more frames, some of which would be quite different from the existing ones,[31] and/or apply a more generous treatment of construction-to-frame relations. However, the first strategy would weaken the connection to other framenets and thereby weaken the reasons for establishing the links in the first place. By the second strategy, if we link too generously, the links would be less meaningful and possibly less useful. In any case, the further we get from ordinary evoking, the greater the need for a more fine-grained classification of construction-to-frame relations.

The decisive factor must be what the links are intended for. If they are to improve the interconnections between constructicon and framenet they must be meaningful. If we mean to relate constructions across languages via frames they must be even more so. Therefore, we have no ambition to link all constructions to frames. Some constructions correspond to frames, others do not, and only the former should be linked to FrameNet. The distinction between the two sets follows from what relations the links represent, and therefore it can only be as precise as the definitions of those relations.

---

31. For the sake of argument, we temporarily disregard the fact that we have no direct influence over SweFN.

6.   Description format

This section presents the description format of SweCcn. An entry consists of up to 15 parts, not counting a handful of fields utilized for project internal housekeeping. Section 6.1 introduces core parts giving a general description of the construction, including a prose definition, a simple sketch of the syntactic structure, and annotated examples. Section 6.2 discusses parts elaborating the general description, including grammatical category, construction types, and descriptions of the construction elements. Section 6.3 describes parts with additional information, including comments, references, and relations to external resources such as (the Swedish) FrameNet and the Berkeley English Constructicon.

6.1   General description

This section presents those parts of a SweCcn entry which give a general description of how the construction is identified (the name), how it is used (a canonical illustration and authentic examples), what kind of meaning it conveys (the definition), and how it is structured from a syntactic point of view (the structure sketch).

6.1.1   *Name and illustration*

The Name and the Illustration work in conjunction to make it easy to identify specific constructions, in particular in listings containing results from searches, such as:

–   `reflexiv_resultativ` 'reflexive resultative' – *supa sig full* 'to drink oneself drunk'
–   `så_grad.resultat` 'so degree.result' – *så stor att den inte gick att lyfta* 'so big it was not possible to lift'
–   `verba_som_en_X_verbar` 'to verb like an X verbs' – *Det smakade som en stia luktar* 'It tasted like a pigsty smells'

Names of units in SweCcn should give a strong pointer to the construction described in the entry, but a name does not have to capture the analyzed structure in its entirety. Constructions have different salient features, making it hard to adhere to a single and simple naming principle. We are in fact using two basic naming principles, as well as mixtures between them:

–   Linguistic descriptions, as in `ellips.fragment` 'ellipsis fragment', and `reflexiv_resultativ` 'reflexive resultative'.
–   Stylized type examples, as in *kniven_i_lådan* 'the knife in the drawer' and *dö_av_skam* 'die of shame'.

– Hybrid names, as in X_*för*_X 'X by X' and verba_*som_en*_X_verbar 'to verb like an X verbs'.

The illustration is a short, idealized example intended to make the construction easier to identify.

6.1.2   *Definition*

The constructions are defined in free text, with definitions in ordinary dictionaries as a general model. A difference from common dictionaries is that the definition is not restricted to purely semantic and functional information but may also describe important formal aspects. Parts of the definition corresponding to semantically important construction elements are annotated. These features are evident in the definition of the reflexiv_resultativ cxn:

> [Någon eller något]$_{NP}$ utför eller undergår [en aktion]$_{Activity}$ som leder (eller antas leda) till att aktören/det utsatta, uttryckt med [reflexiv]$_{Pn.refl}$, uppnår [ett tillstånd]$_{Result}$
>
> '[Someone or something]$_{NP}$ performs or undergoes [an action]$_{Activity}$ whereby the actor/undergoer, expressed by a [reflexive pronoun]$_{Pn.refl}$, reaches [a state]$_{Result}$'

The annotation of definitions resembles the one of examples (cf. 6.1.4), but it does not separate external construction elements from internal ones. It is also worth noting that the annotated definition is somewhat redundant since the same information is often conveyed both by the text and the annotation. This is intentional. The annotation is necessary to make a precise connection between the definition and the various construction elements, but it is also important that the definition is as informative as possible for human readers in applications where the annotation is removed in order to enhance readability.

6.1.3   *Structure sketch*

The structure sketch is basically a list of the formal elements in the construction, assuming as flat a structure as possible. Brackets are used to separate internal construction elements (the construction proper) from external ones (valence bound elements). This can be illustrated by reflexiv_resultativ. As can be inferred from the definition in Section 6.1.2, the construction is built around four formal elements: A noun phrase expressing the actor or undergoer, a verb expressing the activity, a reflexive pronoun referring to the actor or undergoer, and an adjective phrase expressing the resultant state. The construction is analyzed as a verb phrase including the verb, the pronoun and the adjective phrase, with the noun phrase as an external element. This motivates to the following sketch:

(28)   NP [V Pn$_{refl}$ AP]

The actual sketch in the SweCcn omits the NP, since external construction elements may be left out unless they are a salient part of the construction as a whole. This is the case in `kniven_i_lådan` 'the knife in the drawer':

(29)   *Han är   inte **den**   vassaste   kniven   i   lådan.*
       he   be.PRS not the.SG sharp.SUV knife.DEF in drawer.DEF
       'He's not the sharpest knife in the drawer'

This construction is analyzed as an NP, in the translated example *the sharpest knife in the drawer*. However, in this construction, which is a conventionalized metaphorical understatement, the negation is an integral part, although it is not included in the same constituent (see Section 5.2.4). Therefore this external construction element warrants representation in the structure sketch.

Most elements in structure sketches are described either by simple category labels, denoting part of speech, phrase type, or grammatical function, or by lexical units or specific word forms. Special cases are notations for ellipsis and such. The complete list currently contains 27 categories. The category may be specified further when appropriate, as in Pn$_{refl}$ (reflexive pronoun) in the sketch of `reflexiv_re-sultativ` above. Other common variations are examplified by N$_{indef}$ (indefinite noun), VP$_{inf}$ (infinitival verb phrase) and N$_{sg}$ (singular noun). Currently, there are 30 specifications defined in our internal manual.

Elements in structure sketches are not limited to grammatical categories. The most common alternative is lexical units (e.g. på³, äta¹). These are defined in SALDO, a lexical resource with both semantic and morphosyntactic information (see 2.2.2). The system also accommodates literal strings (e.g. *det*). All units may take specifications (e.g. ha$^{1}_{fin}$ for a finite form of *ha* 'have' in its first SALDO sense, and *det*$_{expl}$ for the expletive use of the literal string *det*).

Other devices include specification of aspects on ordering and indices to note coreference, formal identity and categorical identity. The latter is used in constructions such as `samordning` 'conjunction':

(30)   [XP$_1$ (Konj) XP$_1$]

The sketch in (30) reads: Two XPs (i.e., two elements of an arbitrary category) with an optional conjunction in between. The XPs must be of the same category.

As discussed in Section 4.2, the format allows for alternatives and groupings, with a bar separating alternatives and brackets delimiting groups. The delimiting of the construction proper with brackets is a special case. Ordinary parentheses indicate optional parts. Alternatives, groupings and optionality are used sparingly.

For example, the progressive construction *hålla_på*.prog (31), consisting of the complex auxiliary *hålla på*, a linking element, and a VP, comes in two variants. Either it contains an infinitival marker followed by an infinitival VP – or the linking element is the conjunction *och* 'and' and the following VP is in the same tense as *hålla på* (Blensenius, 2015). The structure sketch to represent these alternatives is shown in (31a), and the two variants are illustrated by the examples in (31b)–(31c).[32]

(31) a. [hålla_på[2] att[1] VP$_{inf}$] | [hålla_på[2] och[1] VP]
      b. *Passagerarna höll på att gå ombord.* 'The passengers were boarding' (lit. The passengers held on to board)
      c. *Jag håller på och lär mig.* 'I'm learning' (lit. I hold on and learn myself)

An example of optionality is the unikhet 'uniqueness' construction (32), in which a unique individual or group with a certain property is singled out. As shown in the structure sketch (32a), it consists of a definite article, the word *enda* 'only', an optional nominal head, and a relative clause (which expresses the distinctive property).[33]

(32) a. [Al$_{def}$ enda[1] (N) S$_{rel}$]
      b. *den enda som vet sanningen* 'the only (*one) who knows the truth'
      c. *de enda varelser som jag riktigt kommer överens med* 'the only creatures that I quite get along with'

Strictly speaking, this is not a true case of optionality, since the head is consistently omitted in those cases when it corresponds to English *one* (32b) but otherwise obligatory (32c). Nevertheless, this kind of variation is presented as optionality in

---

32. There is also a prospective *hålla_på*.punkt construction, indicating that something *almost* happened, as in *Jag höll på att ramla av stolen* 'I almost fell off my chair' (lit. I held on to fall off the chair). This construction only occurs with an infinitival marker and infinitival VP. Hence, the *hålla på* data could also be accounted for by an alternative pair of constructions distinguished by form instead of function: a progressive *hålla_på*.och construction (a kind of pseudo coordination) and an ambiguous *hålla_på*.att construction. It should be noted, however, that to most speakers of Swedish, the infinitival marker *att* and the conjunction *och* are homophones, which makes a formal distinction less reliable.

33. The unikhet construction also illustrates an interesting complication regarding inheritance relations between constructions. The variant without a nominal head is arguably a special case of the adjektiv_som_nominal 'adjective as nominal' construction (see Example (17) in Section 5.3). However, we cannot simply treat that construction as a mother to unikhet, since the inherited properties only apply to one of the two structural variants. Inheritance relations are discussed in Section 8.2 (see also Section 6.3.1).

the simplified description format employed in SweCcn. It is of course possible to represent the distinguishing constraints in a more detailed analysis, for instance in the form of specific subconstructions. A more straight-forward example of optionality is given in (6)–(7), Section 4.2.

6.1.4   Authentic examples

The entry includes a series of annotated examples. These are, with few exceptions, gathered from Korp (cf. 2.2.4). The annotation identifies (i) the part of the example corresponding to the construction proper, and (ii) the parts of the example corresponding to individual construction elements, as in the following example from the entry for `reflexiv_resultativ` 'reflexive resultative':

(33)   Har du$_{NP}$ [ölat$_{Activity}$ dig$_{Pn.refl}$ odödlig$_{Result}$]$_{reflexiv\_resultativ}$?
       'Have you aled yourself immortal?'

External construction elements are valence bound components, and not a part of the construction proper, motivating that the brackets delimiting the construction includes the internal elements ACTIVITY, PN.REFL, and RESULT, but not the external NP element.

6.2   Elaborating the description

This section describes how the SweCcn entry clarifies the grammatical category of the construction itself (category), the construction elements and their attributes (construction elements), how the construction is classified in various dimensions (type), lexically specific elements (keywords), and words commonly appearing in the construction (common words).

6.2.1   Category

The Category field contains the grammatical category of the construction as a whole, and should be consistent with the syntactic analysis in the Structure Sketch. It always consists of basic category types (e.g. NP, V, Adv, XP). Since `reflexiv_resultativ` concerns the argument structure of a verb it belongs to the category VP, for instance. In some cases a dual label is motivated; `så_grad.resultat`, for example, may be instantiated by either an adjective phrase or an adverb phrase (AP or AdvP). Hence, we have to allow for alternatives. If a construction is unspecified for category it is labeled XP.

The list of grammatical categories is slowly evolving as the need for a more expressive formalism makes itself felt. This is a deliberate strategy, since a new construction may at any time present a new distinction that has to be taken into

account. The fundamental problem is that constructions may appear at any level of abstraction, which means that traditional grammatical categories, like NP and S, may not always be the right level of abstraction. Currently, we are using 27 grammatical categories.

6.2.2   *Construction elements*

Construction elements (CEs) are divided into internal and external elements. The internal CEs are part of the construction proper, whereas the external ones are valence bound elements. CEs are described by the attributes *name, category* (*cat*), *lexical unit* (*LU*), *role, grammatical function*, and *other*. The *other* attribute is used to capture important information that does not fit in the well defined attributes. The name of the CE is always based on its most salient attribute.

Thus, CEs in SweCcn are analyzed as feature bundles in that every CE is described by the combination of the values of the attributes. While the combination of feature values is specific to the CE in question, features such as categories and roles are globally defined in the database. A notable feature is that the sets of categories and roles both are defined to facilitate analyses at different levels of abstraction. The category noun phrase (NP) includes noun (N), pronoun (Pn), and more. In the same way, the role Actor includes Agent, Cause, and so forth.

As an illustration, `reflexiv_resultativ` 'reflexive resultative' is analyzed as having three internal CEs and one external CE:

Internal construction elements:

– ACTIVITY: cat = V; role = Activity
– PN.REFL: cat = Pn_{refl}; role = Actor/Undergoer
– RESULT: cat = AP; role = Result

External construction elements:

– NP: cat = NP; role = Actor/Undergoer

ACTIVITY corresponds to the verb, PN.REFL to the reflexive pronoun, RESULT to the adjective phrase denoting the result, and NP to the external noun phrase. The values of the different attributes are discussed below.

All CEs must have a name and one attribute supplying the name, but otherwise attributes are optional. They should only be defined if they add something relevant to the description.

The *name* attribute is based on the most salient attribute, but it may be modified to avoid ambiguity or to add clarity. The most common variation is to add a suffix, as in NP1 and NP2.

The *category* attribute (cat) gives a morphosyntactic description consisting of a grammatical category and specifications when appropriate. The basic categories are the same as for the category of the construction as a whole, plus a small number of additions to deal with ellipsis and such. The category attribute may also include specifications in the same way that the structure sketch does. There are currently 30 different specifications. Some examples:

- V_{fin} (finite verb)
- NP_{def} (definite noun phrase)
- $Adj_{suv,def}$ (superlative definite adjective)

The cat attribute allows for alternatives. It is thus possible to define categories such as:

- AP/AdvP (adjective phrase or adverb phrase)
- NP/VP_{inf} (noun phrase or infinitival verb phrase)
- $AP_{comp}/AdvP_{comp}$ (adjective phrase or adverb phrase in the comparative)

The *lexical unit* attribute (LU) refers to units in SALDO (cf. 2.2.2 and 6.1.3). As a result, any specification of the LU attribute in SweCcn is automatically integrated with a vast collection of other linguistic resources. Units in SALDO have a semantic specification, indicated in superscript (as in *äta[1]* 'eat'), as well as information about part of speech and inflectional class.

The *role* attribute: Semantic roles come in many flavors. Fillmore's personal development started with a small set of roles intended to solve the linking problem for early generative grammar and ended with a multitude of frames, each with its own local frame elements (Fillmore, 2003). Goldberg (1995) and others make a distinction between highly abstract *argument roles*, useful for the discussion of general grammatical principles, and *participant roles* defined in the context of single, or minor groups of, lexemes. When analyzing what becomes the subject in ergative and accusative languages, it has been fruitful to introduce metaroles, such as Actor and Undergoer, which are families of argument roles (Van Valin & LaPolla, 1997; cf. also the *proto-roles* of Dowty, 1991).

In developing SweCcn, it is necessary to use semantic roles at all these levels. Constructions, after all, span from the most abstract to the most concrete entities in the language. FrameNet is useful in this regard, since it contains both very abstract frames and rather concrete ones ordered in a network. However, in contrast to FrameNet, we employ a set of semantic roles with the same definition wherever they might appear. They are mostly at the same level of abstraction as traditional argument roles or slightly more concrete, but may in principle be as concrete as required. These are the roles assigned to the role attribute. Whenever

possible, they are defined by a reference to a frame element in a FrameNet frame.[34] Examples:

– Agent: Agent @ Act_intentionally
– Goal: Goal @ Motion.

The *grammatical function* attribute indicates functions such as subject, adverbial, etc., and is a relatively late addition to the description format. Since our main approach to syntactic structure is based on constituency and category rather than dependency relations, this attribute is not regularly used unless grammatical function is an essential defining property of the CE in question. Typical cases are adverbials and determiners, which may often vary in terms of phrase type/part of speech, and clausal word order constructions, which usually depend on grammatical function. In many other cases, however, the grammatical function attribute is largely redundant and/or depends on interaction with other constructions. For example, verb phrase internal NPs are not regularly marked as objects, partly because the relevant information follows from their category, role, and verb phrase internal status; partly because the grammatical function depends on whether the construct (the particular instance of the construction) occurs in an active or passive sentence, etc. Accordingly, CEs such as nominal arguments are always specified for grammatical function in clausal word order constructions but, as of yet, usually not in argument structure constructions.

The *other* attribute is used for information that does not fit into the cat, LU, role or grammatical function attribute. Anything in the other attribute may be an indication of weaknesses in our format, so we are using it to find out where improvements are in order. We use it systematically when the CE has to be a certain literal string, and to note when the CE is the former or latter part of a compound (*förled* or *efterled*). Other uses include some semantic restrictions.

6.2.3   *Type*

As mentioned in Section 3.4, the basic purpose of the types is to bring order to the constructicon. This is a very important function since the names of constructions do not have a simple and transparent ordering comparable to how units in a dictionary may be listed alphabetically. New types are introduced gradually as SweCcn is employed for new purposes. This also entails that types may express quite different perspectives on language and language use, as is made clear by the following examples:

---

34. Note that the explicit linking of semantic roles to frame elements in highly abstract frames in FrameNet / SweFN implies a listing of frames inheriting from the abstract frames, for example, of all frames inheriting from Act_intentionally in the case of Agent.

- *kategori*: The construction is a category (grammatical categories are constructions and are to be included in the SweCcn).
- *sammansättning*: The construction is a compound.
- *genreberoende*: The use of the construction is dependent on genre.
- *inlärningsfokus*: The construction is problematic for non-native learners of Swedish.
- *bisats*: The construction is a kind of subordinate clause.
- *konstruktion*: The entry is a construction.

6.2.4   Keywords and common words

Some constructions have one or more lexically specific construction elements. In addition to being specified for lexical unit in the structure sketch and the construction element analysis, these elements are registered as *keywords* (roughly corresponding to the notion of *construction evoking elements* or *CEEs*; cf. Fillmore, Lee-Goldman & Rhomieux, 2012). This representation is meant to facilitate identification of constructions by lexical means.

Common words are lexical units appearing in the construction to a remarkable degree, without being proper keywords; rather, they are typical instantiations of variable CEs. In other words, they may be perceived as *collostructional* elements, although their inclusion is not based on a proper collostructional analysis (cf. Stefanowitsch, 2013) but mainly on their being strikingly frequent in the corpora employed in the construction analysis or, in some cases, known valence patterns as documented in dictionaries. Tuples of lexical units appearing together are listed inside braces. The entry for `reflexiv_resultativ` illustrates that tuples often consists of verbs and arguments frequently appearing together:

- {äta[1] 'eat' : mätt[1] 'full'}
- {supa[1] 'drink' : full[1] 'drunk'}
- {skrika[1] 'scream' : hes[1] 'hoarse'}
- springa[1] 'run'

The verb *springa* 'run' is often used in the construction, making it a common word, but it is not associated with any particular result in the same way that *skrika* 'scream' is associated with *hes* 'hoarse'. For this reason, it is added as a singleton. In general, singletons are more common than tuples among the common words listed in SweCcn.

Both keywords and common words are linked to their corresponding entries in the SALDO lexicon (and thereby other lexical resources, cf. 2.2.2), thus providing additional lexical information and also making the construction entry accessible from the lexical infrastructure Karp (cf. 2.2.3).

6.3   Relating and commenting on the constructions

This section discusses parts of the SweCcn entry with information about the internal organization of SweCcn (inheritance), the relation to frames in FrameNet and SweFN (frame), related constructions in the English FrameNet Constructicon (Berkeley ID), remarks on the analysis (comment), and references to papers and other sources (reference).

6.3.1   *Inheritance*

Sometimes a construction clearly is a more specific variant of a more general one, from which it can be said to inherit some of its properties, and in some cases a construction may even have several obvious mothers. Such relations are to be noted in the inheritance field, thus transforming a long list of entries into a network of constructions. Out of practical concerns, however, information about inheritance is added gradually, and seldom when an entry is first created. An inheritance link represents default inheritance: properties of the more general construction are inherited by the more specific one, except when an explicitly stated property in the specific construction overrides a conflicting property at a more abstract level. Inheritance and other network relations are discussed in Section 8.2 below.

6.3.2   *Frame and Berkeley ID*

The field Frame is used for linking constructions to frames in the Swedish framenet, SweFN, and indirectly to the original FrameNet in Berkeley (see Section 5). As mentioned in Section 6.2.2, links to frames are also established through our treatment of semantic roles, which are defined in terms of frame elements.

The Berkeley ID field contains the names of closely corresponding constructions in the English FrameNet Constructicon. This field is a bridge connecting entries in SweCcn to constructicons for other languages (cf. Bäckström, Lyngfelt & Sköldberg, 2014; Lyngfelt, Torrent et al., this volume).

6.3.3   *Comment and reference*

The Comment is used for remarks on the analyses intended to be visible for the actual user of the SweCcn (there is also a hidden comment field for internal use only). It is often used as a pointer to related constructions or to aspects on the entry that may be reworked in the future, as in this comment from the entry for `reflexiv_resultativ`:

> Det finns också en variant där resultatet uttrycks med prepositionsfras, t.ex. "träna sig i form".

> 'There is also a variant where the result is expressed in a prepositional phrase, for example "exercise oneself in shape".'

The comments may also include information of constructional variation beyond what is recorded in the definition and structure sketch (cf. 4.2).

7.   Using the constructicon

In this section we turn to the user's perspective. Since our general aim is a resource of wide applicability, SweCcn is not specifically adapted to any single kind of usage. Such a general purpose tends to lead to abstraction, and therefore the constructicon is probably best suited for users with better than average knowledge of linguistics. For example, our concern with L2 relevance does not mean that SweCcn is designed for direct use by L2 learners; rather, we have their teachers in mind.

Below, we first present the interfaces and search options available (7.1) and then briefly comment on potential (secondary) applications (7.2).

7.1   Interface

SweCcn is available online both independently and as a part of the lexical infrastructure Karp. In the following, we will focus on the independent interface. It offers two main ways to access construction descriptions: selecting from a *list* or entering a *search query*. The user may also choose between two display options: simple (which is default) and extended construction descriptions. Other features include a user's guide, a pedagogic toolkit, and a SweCcn bibliography.

The starting page is shown in Figure 14:
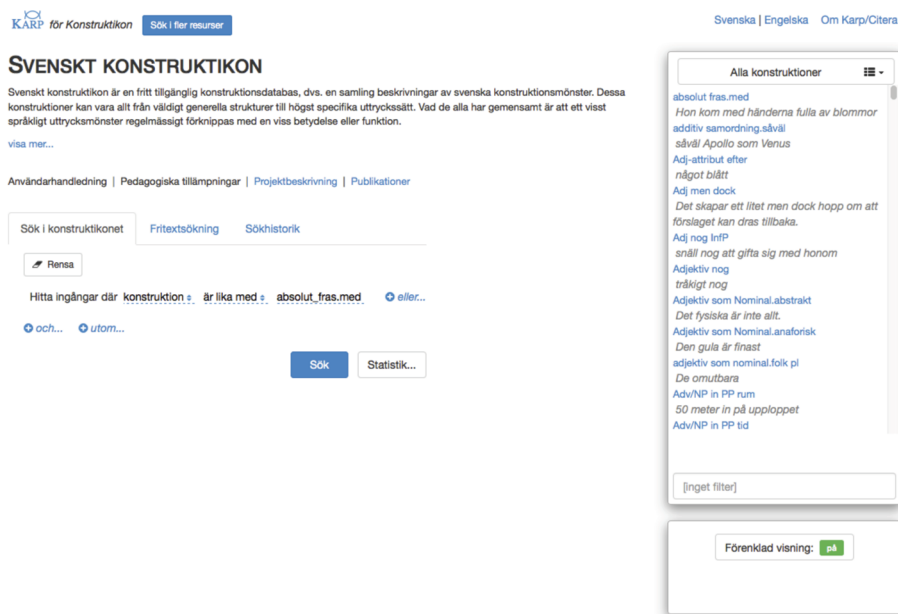
Figure 14. The SweCcn interface

On the right hand side is a *list* of constructions, presented by name and illustration (a short example). The user may choose a construction by clicking on its name, after which the construction description is displayed in the bottom left. In the default setting the list contains all construction entries in SweCcn, but it may also be restricted to show only constructions of a certain category or type.

To the left, the user may enter a *search query* for the constructions of interest, either for particular constructions or groups with a certain property. A major benefit with this option, compared to the list, is the possibility to display more than one construction description at a time, typically grouped by category and/or type.[35] One may for instance access all reflexive constructions, all PP constructions, or combine properties to show e.g. all causal constructions of category VP or all constructions expressing either contrast or comparison.

Search queries may also be based on other properties, such as keywords or common words, certain expressions occurring in the definitions, etc. One may also target specific expressions in a free text search, which does not discriminate whether a certain word is a keyword or simply occurs in one of the example sentences.

Above the search query, there are links to a user's guide, a pedagogic toolkit, a project description, and a SweCcn bibliography, including links to publications available online. In the low right box, finally, the user may choose between a simple and an extended display mode:

– *Simple:* in addition to name and illustration, this mode of display includes the definition, structure sketch, and annotated examples.
– *Extended:* the full constructicon entry is shown, including sections for construction elements, inheritance relations, FrameNet links, etc. (where applicable; empty information categories are not displayed), according to the format described in Section 6 above.

In addition to this interface, SweCcn is also accessible within Språkbanken's lexical infrastructure Karp (see Section 2.2.3 above), in which most of the same functionality is available. The main difference is that the Karp interface is the same for all resources in the infrastructure and, hence, not optimized specifically for the constructicon. Thus, it contains more functions than the exclusive SweCcn interface, but many of these are not relevant for the constructicon and therefore they obscure the functions that are. A major benefit with the Karp interface, however, is the possibility to access several resources at once. Thus, for any given lexical unit, one may display entries in one or more lexical resources in combination with all construction entries to which it is linked.

---

35. As mentioned in Section 3.4, types can refer to form (passive, genitive, reduplication) or meaning (aspect, motion, time expressions), indicate usage functions (interaction) or particular relevance (learner focus), etc.

7.2   On potential applications

While SweCcn is primarily a basic linguistic resource, it is also intended for application to lexicography, language technology, and language pedagogy. Such applications are not within the scope of the SweCcn project, but rather a potential for which we build a foundation. However, to make the database applicable it is necessary to bear likely applications in mind and adapt the design to accommodate them.

A key feature in this regard is compatibility with the Språkbanken infrastructure on the one hand and framenet/constructicon resources for other languages on the other, in the latter case working towards an international infrastructure for framenets and constructicons. A major longterm goal is interlingual constructicon application. In this endeavor we may partly draw on previous interlinking between framenets across languages, at least as regards constructions linked to FrameNet frames (cf. Bäckström, Lyngfelt, & Sköldberg, 2014). For an extensive discussion of the development of interlingual constructicography, see Lyngfelt, Torrent et al. (this volume).

Making SweCcn compatible with LT resources at Språkbanken has two purposes. One is integration with the lexical infrastructure and interlinking with various lexical resources, as described in Section 2.2 above. The other concerns LT application in general. On the one hand, information about constructions may be incorporated in various NLP systems; for example, parts of SweCcn have been adapted to the GF system (Grūzītis et al., 2015), which is employed for, e.g., computer aided translation.[36] On the other hand, SweCcn may be used to develop tools for automatic identification of constructions in authentic texts. The latter is a highly desirable research objective and in fact one of the initial motivations for the high involvement of LT in the SweCcn project. To this end, we have conducted experiments with automatically generated construction candidates, which have also benefitted the growth of SweCcn itself (see Section 3.2.1 above). An obvious goal for future development is to be able to identify not only potential construction patterns but also specific constructions.

SweCcn is also intended as a resource for language education, especially regarding Swedish as a second or foreign language. Although the constructicon is not designed as a direct tool for language learning, the database provides information that can be used in textbooks and classroom exercises. To encourage and facilitate such usage, the SweCcn website contains a pedagogic toolkit, including sample exercises, and we also present pedagogical possibilities of SweCcn in various talks and publications (e.g. Loenheim et al., 2016).

---

36. GF is short for Grammatical Framework. For an introduction, see Ranta (2011).

Obviously, such different areas of application place different demands on the constructicon. While NLP applications depend on formal precision, language pedagogy requires user-friendliness from a human perspective. This conflict of interests is less troublesome than it might seem, since not all formal specifications in the database have to be visible in the user interface. Thus, the availability of different display options is one way of making SweCcn more user-friendly. However, because the basic database format is to some extent a compromise between different demands, the adaptability is somewhat limited. This does not preclude, say, further development into a construction-based learner's tool, but such a step is probably best treated as a partly independent spin-off product.

8. Discussion and outlook

In this chapter we have been concerned with the emerging practice of constructicography, as carried out in the Swedish constructicon (SweCcn). A central theme has been the combination and accommodation of principles of construction grammar and lexicography in practical constructicon development. We have stressed the grounding of SweCcn in its local context, and its influence by and dependence on the Språkbanken environment in particular, as well as in the international CxG tradition. The process of construction description from selection to finished entry has been illustrated, and special attention has been given to the treatment of constructional variation, linking SweCcn to FrameNet, and the description format as such. We have also addressed SweCcn from the user's perspective and briefly commented on some potential applications.

Since constructicons are a new kind of linguistic resource one cannot fully anticipate what will be required for their development. Therefore, a recurring characteristic of the workflow in SweCcn is what may be called organic development. Rather than specifying features and relations in advance, before knowing what may be required to account for the constructions encountered, we have allowed a fairly wild growth of notational practice – followed by regular sessions of pruning and harmonization. This data-driven strategy benefits descriptive adequacy and adaptability to new kinds of constructions, at the cost of an increasing amount of inconsistencies in the database. Although these inconsistencies are regularly reduced in editorial rounds, where we adapt older constructicon entries to the newer standards, it is virtually impossible to eliminate them altogether. On the other hand, had we been too restricted by preset descriptive tools, that would have forced other kinds of inconsistencies whenever those tools had failed to adequately cover new phenomena.

Likewise, although we collaborate closely with the constructicon projects for other languages, aiming for maximal mutual compatibility, it is inevitable that different projects come up with different solutions to related problems. While this is a source of inconsistencies regarding interlingual applicability of the constructicons, it may also bring methodological benefits: since the differences also highlight the benefits and drawbacks of each approach, we can thus learn from each other. In any case, it is clear that difficulties to link the constructicon resources cross-linguistically not only concern differences between the languages involved but also methodological differences (cf. Lyngfelt, Torrent et al., this volume; Bäckström, Lyngfelt, & Sköldberg, 2014).

8.1   Relations between constructicon and FrameNet

In particular, SweCcn stands out as regards the relation to FrameNet. Whereas most other constructicons are part of the framenet development for their respective languages, using FrameNet annotation software etc., SweCcn is essentially independent from SweFN. This difference affects both what patterns are included in the constructicon and how they are distinguished and characterized. In FrameNet Brasil, for example, there is a general aim for complementary distribution between its framenet and constructicon components and an explicit algorithm is employed to decide which patterns go where, thereby restricting the constructicon to linguistic patterns that do not fit in framenet (Torrent et al., 2014, p. 42; Torrent et al., this volume). In SweCcn, by contrast, constructions are selected without regard to SweFN, and so there happens to be some partial overlap between the two resources, mostly in the sense that SweFN includes lexical instantiations of more general SweCcn constructions.[37]

Furthermore, there is a tendency in the English FrameNet Constructicon to assume semantic distinctions in accordance with FrameNet, whereas constructions in SweCcn are distinguished solely with respect to construction-particular properties. This is reflected, e.g., by the different treatment of `rate` constructions, as illustrated in Section 5.2.2 above (see also Bäckström, Lyngfelt, & Sköldberg,

---

37. There also some cases where a partially schematic construction in SweCcn is treated as a lexical unit in SALDO, for instance the *ju_desto* construction (see Section 4.2, Example (8) above). Such overlap is only partially redundant, even after links has been established, since the resources have different foci, contain partially different kinds of information, and may each be employed in external applications without direct access to the other. Notably, a lexical unit in SALDO does not cover the variable elements of the construction. The reason why it is still meaningful to reduce a partially schematic construction to a lexical unit is that the variable construction elements are then effectively considered valence elements.

2014, pp. 21–22). Where the English constructicon assumes four constructions, corresponding to four different frames, SweCcn employs only two, since the frame distinctions do not reflect the constructional differences in Swedish, neither regarding formal properties nor constructional meaning.

We do nonetheless link constructions to corresponding frames where applicable, as described in Section 5 above, but only after the construction has been characterized. A benefit with the SweCcn approach is that the construction analyses are less biased towards (English) FrameNet analyses, and should therefore give more accurate accounts of Swedish constructions. An obvious drawback is reduced compatibility with both FrameNet and constructicons developed in closer accordance with FrameNet. This drawback applies even to the relation between SweCcn and SweFN, since the latter resource, with a few exceptions, assumes the same frames as its English counterpart (cf. Friberg Heppin & Toporowska Gronostaj, 2014).

8.2 From construction dictionary to construction network

> The grammar of a language can be seen as a repertory of constructions, plus a set of principles which govern the nesting and superimposition of constructions into or upon one another.   (Fillmore, 1988, p. 37)

> The totality of our knowledge of language is captured by a network of constructions: a 'construct-i-con'.   (Goldberg, 2003, p. 219)

A constructicon resource like SweCcn is essentially a practical instantiation of the first part of the above Fillmore quote: "a repertory of constructions", for which Jurafsky (1991, p. 18) coined the term *constructicon*. It is thus a somewhat narrower concept than the "totality of our knowledge" notion in the Goldberg quote, as the principles for combining the cxns (the second part of the Fillmore quote) are not considered part of the constructicon *per se*. Nevertheless, the constructicon and the combinatory principles are of course highly dependent on each other, and any attempt to model the former would clearly benefit from a clear notion of the latter.

The notion of a constructicon as a repertory of cxns has been a central idea in the CxG tradition for decades, typically pictured as an inheritance network (Fillmore & Kay, 1996; Kay & Fillmore, 1999; Sag, 2012; Goldberg, 2013; see also Lyngfelt, this volume). The internal structure of this network, however, is still vastly understudied – and the combinatory principles even more so. There is ongoing work to model cxn networks computationally in both Fluid CxG (cf. Steels, 2013) and Embodied CxG (cf. Bergen & Chang, 2013) and some interesting studies on how nodes and relations in the network may be established, strengthened, and weakened on the basis of usage data (e.g. van Trijp & Steels, 2012; Van de Velde, 2012; Hilpert, 2015). There are also a number of case-studies of small sub-hierarchies of cxns. But

large-scale descriptive accounts are still largely lacking, and how the repertory of cxns may be structured remains an open question.

A resource such as SweCcn is a natural testing ground for implementing a construction network on a somewhat larger scale. Modeling such a network is also a central objective of the project, as a system for the internal structure of the database as well as for theoretical reasons. Although relations between constructions is still among the less developed aspects of SweCcn, we are currently in the process of remedying that situation. There are two main sides to this work: on the one hand, it requires descriptions of more general constructions as central nodes in the network; on the other hand, it is a matter of establishing the inter-constructional relations as such. Regarding the latter aspect, the fundamental question is what kinds of relations to posit, both vertically (inheritance) and horisontally.

In the CxG literature, and elsewhere, there are two basic approaches to inheritance relations: complete inheritance (Fillmore & Kay, 1996; Sag, 2012) and default/normal inheritance (Goldberg, 1995, 2013). On complete inheritance, "When one construction inherits another, the first contains all the information of the second and – in the non vacuous case – more" (Kay & Fillmore, 1999, p. 7). On default/normal inheritance, the construction inherits all properties "that do not conflict with its own specifications" (Goldberg, 1995, p. 70; cf. Lakoff, 1987). Both approaches allow for multiple inheritance, that is, for a construction to inherit properties from more than one other construction.

Complete inheritance captures taxonomic relations. For example, the constructions `jämförelse` 'comparison' and `adjektiv_som_nominal` 'adjective_as_nominal' each have a number of subtypes, more specific cxns which inherit all the properties of the mother cxn. Similarly, the category label of the cxn entries is essentially an inheritance link, since every cxn with the category VP is a subtype – or *instantiation* – of a general VP cxn. This also applies to cxn elements (CEs); a CE labeled AP would be an instance of a general `adjective_phrase` cxn, Pn_{refl} of a general `reflexive_pronoun` cxn etc.[38]

So far, most inheritance relations recognized in SweCcn happen to comply with complete inheritance, but not all of them do. The `adjektiv_som_nominal` construction, for instance, is an NP construction but differs from ordinary NPs in lacking a nominal head (34).[39]

---

38. Many of these general constructions have yet to be defined in SweCcn.

39. In SweCcn, as well as in the English constructicon (cf. Fillmore, Lee-Goldman, & Rhomieux, 2012, pp. 357–359), three subtypes of this construction are distinguished: abstract (34a), anaphoric (34b), and people.plural (34c).

(34) a. *Konsten söker* **det** *okända*.
    art.DEF seek.PRS the.SG.NEUT unknown.DEF
    'Art seeks the unknown'

  b. *Hon valde* **den** *blåa*.
    she choose.PST the.SG blue.DEF
    'She chose the blue one'

  c. *Kvinnor är överrepresenterade bland* **de** *fattiga*.
    woman.PL be.PRS over-represented.PL among the.PL poor.PL
    'Women are over-represented among the poor'

As a kind of noun phrase, this construction inherits properties from a general NP construction but, lacking a nominal head, it is clearly not a case of complete inheritance.[40] There is nothing extraordinary about this, however. On the contrary, many linguistic patterns have properties that not only restrict but also deviate from more general patterns of which they are intuitively perceived as subtypes. Therefore, SweCcn assumes default inheritance, and complete inheritance is simply the default case.

Turning to horizontal relations (Van de Velde, 2012), also called subpart links (Goldberg, 1995; Hilpert, 2015), some types follow inherently from the feature-based system for defining construction elements (CEs) in SweCcn, since categories, semantic roles, grammatical functions, and lexical units are all globally defined. This means that any and all constructions containing CEs that share a certain feature, or combination of features, are connected in the database through those CEs. In terms of network structure, such horizontal relations connect cxns whose CEs inherit properties from the same source; for example, all reflexive cxns include a reflexive pronoun, i.e. contain a CE that instantiates the `reflexive_pronoun` construction. Thus, such horizontal relations between cxns primarily connect CEs rather than the cxns containing them (except from the viewpoint that these CEs are also cxns of their own).

As for global properties shared between constructions, many follow from constructional inheritance whereas others might not. This is partly a matter of definition, partly a question of which kinds of relations to represent. Do we, for instance, wish to relate infinitival relative cxns to (finite) relative clause cxns? Or intransitive cxns to their corresponding transitive cxns? Constructions applying to the same domains? Should we indicate metaphorical relations?

---

40. In this particular case, complete inheritance might be saved if we assume that a nominal head is an optional property of NPs. Apart from this assumption being counter-intuitive, it would have unwanted consequences for the rest of the network of NP constructions. One might also consider classifying adjective_as_nominal as an AP construction, but then it would differ functionally from other APs in what grammatical functions it serves.

In the case of the different relative constructions, it seems reasonable that they all inherit from a general relative clause cxn. At the same time the infinitival relatives also inherit from a general infinitival clause/VP cxn, which means that the system requires multiple inheritance (see e.g. Sag, 2012). Regarding the other examples it is less obvious what the superordinate cxns would be. We may still, however, represent a connection through the type attribute (see Sections 3.4 and 6.2.4).

As previously mentioned, types may be distinguished by any salient property shared by a set of constructions, many of which do not correspond to inheritance relations.[41] For example, the types *kontrast* and *polaritet* indicate cxns for expressing contrasts and cxns marked for polarity, respectively, properties for which there are no basic general cxns in the system, not to mention the type for special L2 relevance. Types indicating a domain, such as *tidsuttryck* 'time expressions', are in some cases mirrored by links to corresponding frames. In any case, the current set of types does not follow from any systematic principle other than a subjective recognition of (potential) relevance. For that very reason, this rather motley crew may, in addition to covering relations not covered by inheritance or the feature system, also serve as a useful base for future development of the construction network.

It should be noted, however, that many kinds of relations between cxns are not to be treated as network links. One such case is argument alternations, for instance between transitive and intransitive, as in (35). Although the constructions in question are arguably connected both by type (as argument structure constructions) and by a number of subpart links, neither would capture the particular combinations of similarities and differences in (35).

(35)   a.   *The ice melted.*
       b.   *The sun melted the ice.*
       c.   *Kim ate.*
       d.   *Kim ate an apple.*

Instead of network links, these alternations may be handled by (derivational) constructions. The ergative alternation in (35a)–(b) would correspond to a so-called pumping construction (cf. Fillmore, Lee-Goldman & Rhomieux, 2012), and the implicit object in (35c) can be accounted for by a null instantiation construction.

---

41. There are also several types that do correspond to inheritance relations (for example, the type *bisats* 'subordinate clause') or to feature sharing between CEs (such as the type *reflexiv* 'reflexive'), making the system somewhat redundant. This is partly because we started using types before addressing the network structure in earnest, but the main reason is that the types provide a user-friendly searching and sorting devise. Thus, regardless of other network relations, the type attribute is a useful feature in itself.

The kind of constructional relation that most strikingly falls outside the constructicon, in its sense of a repository of constructions, are the principles by which constructs (instances of constructions) are combined to form utterances (Fillmore, 1988, p. 37; see the beginning of this subsection). Such a set of principles is an equally vital part of a construction grammar as the construction repository itself. Although the combinatory principles are by definition external to the constructicon proper, they should still be taken into consideration in constructicon development, for at least two reasons (which are two sides of the same thing). First, if the constructs licensed by the constructions are to fit together, the construction descriptions must be designed to be compatible with each other. Due to so-called coercion or mismatch phenomena, this task is less trivial than it might first seem. Second, an essential aspect of all clausal and phrasal constructions is the combination of construction elements, which in turn correspond to constructions. Thus, the combinatory possibilities are to some extent integral to the constructions themselves. This is particularly pressing in the case of the more central nodes in the network.

In other words, designing a comprehensive constructicon may require a satisfactory account of a corresponding combinatory system. In combination, the two tasks amount to writing a complete construction grammar. This, of course, is well beyond the scope of SweCcn. Not only does the sheer magnitude of such a project make it unrealistic,[42] our aim for a relatively simple description format precludes a sufficient level of detail. Nonetheless, to the extent of coverage attained, consistency and coherence of the system require that the combinatory properties of constructions be taken into account.

Acknowledgments

---

42. In principle, a full account of all the constructions in a language is not only practically but also theoretically impossible. Hence, 'comprehensive' presupposes a certain amount of idealization, comparable to the limitations of a major dictionary. Even so, a corresponding degree of coverage is not attainable within the foreseeable future.

References

Abrahamsson, N., & Hyltenstam, K. (2009). Age of Onset and Nativelikeness in a Second Language: Listener Perception versus Linguistic Scrutiny. *Language Learning*, 59(2), 249–306. doi:10.1111/j.1467-9922.2009.00507.x

Ahlberg, M., Forsberg M., & Hulden, M. (2014). Semi-supervised learning of morphological paradigms and lexicons. In *Proceedings of EACL 2014* (pp. 569–578). Gothenburg: ACL.

Allén, S. (1970). Vocabulary data processing. In H. Benediktsson (Ed.), The Nordic languages and modern linguistics: Proceedings of the International conference of Nordic and general linguistics (pp. 235–261). Reykjavík: Visindafélag íslendinga.

Andersson, P. & Ahlberg, M. (2013). Towards automatic tracking of lexical change: linking historical lexical resources. In *Proceedings of the workshop on computational historical linguistics at NODALIDA 2013*. Oslo: NEALT.

Atkins, B. T. S., & Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. Oxford & New York: Oxford University Press.

Bäckström, L., Borin, L., Forsberg, M., Lyngfelt, B., Prentice, J., & Sköldberg, E. (2013). Automatic identification of construction candidates for a Swedish constructicon. In *Proceedings of the workshop on lexical semantic resources for NLP at NODALIDA 2013* (pp. 2–11). Oslo: NEALT.

Bäckström, L., Lyngfelt, B., & Sköldberg, E. (2014). Towards Interlingual Constructicography. On correspondence between constructicon resources for English and Swedish. *Constructions and Frames*, 6(1), 9–32. doi:10.1075/cf.6.1.02bak

Bergen, B., & N. Chang (2013). Embodied Construction Grammar. In Th. Hoffmann, & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 168–190). Oxford & New York: Oxford University Press.

Blensenius, K. (2015). *Progressive constructions in Swedish*. (Doctoral dissertation). Dept. of Swedish, University of Gothenburg.

Boas, H. C. (2009). Semantic frames as interlingual representations for multilingual databases. In H. C. Boas (Ed.), *Multilingual FrameNets in Computational Lexicography: Methods and Applications* (pp. 59–100). Berlin & New York: Mouton de Gruyter. doi:10.1515/9783110212976.1.59

Boas, H. C. (Ed.). (2010). *Contrastive Studies in Construction Grammar*. Amsterdam/Philadelphia: John Benjamins. doi:10.1075/cal.10

Borin, L. (2010). Med Zipf mot framtiden – en integrerad lexikonresurs för svensk språkteknologi ['With Zipf towards the future – an integrated lexical resource for Swedish language technology']. *LexicoNordica*, 17, 35–54.

Borin, L. (2012). Core vocabulary: A useful but mystical concept in some kinds of linguistics. In D. Santos, K. Lindén, & W. Ng'ang'a (Eds.), *Shall we play the Festschrift game? Essays on the occasion of Lauri Carlson's 60th birthday* (pp. 53–65). Berlin: Springer. doi:10.1007/978-3-642-30773-7_6

Borin, L., Allwood, J., & de Melo, G. (2014). Bring vs. MTRoget: Evaluating automatic thesaurus translation. In *Proceedings of LREC 2014* (pp. 2115–2121). Reykjavík: ELRA.

Borin, L., Comrie, B., & Saxena, A. (2013). The Intercontinental Dictionary Series – a rich and principled database for language comparison. In L. Borin & A. Saxena (Eds.), *Approaches to measuring linguistic differences* (pp. 285–302). Berlin: De Gruyter Mouton. doi:10.1515/9783110305258.285

Borin, L., Dannélls, D., Forsberg, M., & McCrae, J. P. (2014). Representing Swedish lexical resources in RDF with lemon. In *Proceedings of the ISWC 2014 Posters & Demonstrations Track a track within the 13th International Semantic Web Conference* (ISWC 2014) (pp. 329–332).

Borin, L., Dannélls, D., Forsberg, M., Toporowska Gronostaj, M., & Kokkinakis, D. (2010). The past meets the present in Swedish FrameNet++. In *14th EURALEX International Congress* (pp. 269–281). Ljouwert: Fryske Akademy.

Borin, L., & Forsberg, M. (2009). Something old, something new: A computational morphological description of Old Swedish. In *LREC 2008 workshop on language technology for cultural heritage data* (LaTeCH 2008) (pp. 9–16). Marrakech: ELRA.

Borin, L., & Forsberg, M. (2011). A diachronic computational lexical resource for 800 years of Swedish. In C. Sporleder, A. van den Bosch, & K. Zervanou (Eds.), *Language technology for cultural heritage* (pp. 41–61). Berlin: Springer. doi:10.1007/978-3-642-20227-8_3

Borin, L., & Forsberg, M. (2014). Swesaurus; or, The Frankenstein approach to wordnet construction. In *Proceedings of the Seventh GlobalWordNet Conference* (GWC 2014) (pp. 315–223). Tartu: GWA.

Borin, L., Forsberg, M., & Lönngren, L. (2013). Saldo: a touch of yin to WordNet's yang. *Language Resources and Evaluation*, 47(4), 1191–1211. doi:10.1007/s10579-013-9233-4

Borin, L., Forsberg, M., & Lyngfelt, B. (2013). Close encounters of the fifth kind: Some linguistic and computational aspects of the Swedish FrameNet++ project. *Veredas*, 17(1), 28–43.

Borin, L., Forsberg, M., Olsson, L. -J., Olsson, O.; & Uppström, J. (2013). The lexical editing system of Karp. In *Proceedings of the eLex 2013 Conference* (pp. 503–516). Tallin: Eesti Keele Instituut.

Borin, L., Forsberg, M., Olsson, L. -J., & Uppström, J. (2012). The open lexical infrastructure of Språkbanken. In *Proceedings of LREC 2012* (pp. 3598–3602). Istanbul: ELRA.

Borin, L., Forsberg, M., & Roxendal, J. (2012). Korp – the corpus infrastructure of Språkbanken. In *Proceedings of LREC 2012* (pp. 474–478). Istanbul: ELRA

Borin, L., Nieto Piña, L., & Johansson, R. 2015. Here be dragons? The perils and promises of inter-resource lexical-semantic mapping. Semantic resources and semantic annotation for Natural Language Processing and the Digital Humanities. Workshop at NODALIDA, May 11, 13–18 2015 (pp. 1–11). Vilnius. Linköping: Linköping Electronic Conference Proceedings.

Bybee, J. (2010). *Language, Usage and Cognition*. Cambridge & New York: Cambridge University Press. doi:10.1017/CBO9780511750526

Chomsky, N. (1957). *Syntactic Structures*. Mouton: The Hague.

Croft, W. (2001). *Radical Construction Grammar. Syntactic theory in typological perspective*. Oxford & New York: Oxford University Press. doi:10.1093/acprof:oso/9780198299554.001.0001

Dooley, S. (2014). The Swedish Comparative Correlative Construction: Ju … Desto … and Variations. *Constructions* 2014(4).

Dowty, D. (1991). Thematic proto-roles and argument selection. *Language*, 67, 547–619. doi:10.1353/lan.1991.0021

Ehrlemark, A. (2014). *Ramar och konstruktioner – en kärlekshistoria* ['Frames and constructions – a love story']. (GU-ISS 2014-01) Dept. of Swedish, University of Gothenburg.

Fanselow, G., Féry, C., Vogel, R., & Schlesewsky, M. (Eds.). (2006). *Gradience in Grammar: Generative perspectives*. Oxford & New York: Oxford University Press. doi:10.1093/acprof:oso/9780199274796.001.0001

Fellbaum, C. (Ed.). (1998). *WordNet: An electronic lexical database*. Cambridge, MA: MIT Press.

Fillmore, C. J. (1988). The mechanisms of 'Construction Grammar'. *Berkeley Linguistic Society*, 14, 35–55.

Fillmore, C. J. (2003). *Form and Meaning in Language: Volume 1, Papers on semantic roles*. Stanford: CSLI Publications.

Fillmore, C. J. (2008). Border Conflicts: FrameNet Meets Construction Grammar. In E. Bernal, & J. DeCesaris (Eds.), *Proceedings of the XIII EURALEX International Congress* (pp. 49–68). Barcelona: Universitat Pompeu Fabra.

Fillmore, C. J., & Kay, P. (1996). *Construction Grammar Coursebook*. Manuscript. Dept. of linguistics, University of California, Berkeley.

Fillmore, C. J., Kay, P., & O'Connor, M. C. (1988). Regularity and idiomaticity in grammatical constructions. The case of *let alone*. *Language*, 64, 501–538. doi:10.2307/414531

Fillmore, C. J., Lee-Goldman, R., & Rhomieux, R. (2012). The FrameNet Constructicon. In H. C. Boas, & I. A. Sag (Eds.), *Sign-Based Construction Grammar* (pp. 309–372). Stanford: CSLI Publications.

Forsberg, M., Johansson, R., Bäckström, L., Borin, L., Lyngfelt, B., Olofsson, J., & Prentice, J. (2014). From construction candidates to constructicon entries. An experiment using semi-automatic methods for identifying constructions in corpora. *Constructions and Frames*, 6(1), 113–134. doi:10.1075/cf.6.1.07for

FrameNet. <https://framenet.icsi.berkeley.edu/fndrupal/>

Francopoulo, G. (Ed.). (2013). *LMF: Lexical Markup Framework*. London/Hoboken, NJ: ISTE/ Wiley. doi:10.1002/9781118712696

Friberg Heppin, K., & Toporowska Gronostaj, M. (2014). Exploiting FrameNet for Swedish: Mismatch? *Constructions and Frames*, 6(1), 52–72. doi:10.1075/cf.6.1.04hep

Goldberg, A. E. (1995). *Constructions. A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.

Goldberg, A. E. (2003). Constructions: A new theoretical approach to language. *Trends in Cognitive Sciences*, 7(5), 219–224. doi:10.1016/S1364-6613(03)00080-9

Goldberg, A. E. (2013). Constructionist Approaches. In Th. Hoffmann, & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 15–31). Oxford & New York: Oxford University Press.

Grūzītis, N., Dannélls, D., Lyngfelt, B., & Ranta, A. (2015). *Formalising the Swedish Constructicon in Grammatical Framework*. In *Proceedings of the Grammar Engineering Across Frameworks (GEAF) 2015 Workshop*, (pp. 49–56). Beijing, China. doi:10.18653/v1/W15-3307

Hilpert, M. (2015). From *hand-carved* to *computer-based*: Noun-participle compounding and the upward strengthening hypothesis. *Cognitive Linguistics*, 26, 113–147. doi:10.1515/cog-2014-0001

Hoffmann, Th. & Trousdale, G. (Eds.). (2013). *The Oxford Handbook of Construction Grammar*. Oxford & New York: Oxford University Press. doi:10.1093/oxfordhb/9780195396683.001.0001

Hovmark, H. (2012). Betydningsbeskrivelse og prototypteori ['Meaning description and prototype theory']. *LexicoNordica*, 19, 59–78.

ISO 2008. Language resource management – Lexical markup framework (LMF). International Standard ISO 24613:2008.

Jackendoff, R. (1997). Twistin' the Night Away. *Language* 73, 534–559. doi:10.2307/415883

Jurafsky, D. (1991). *An On-line Computational Model of Human Sentence Interpretation: A Theory of the Representation and Use of Linguistic Knowledge*. (Doctoral dissertation). Dept. of Electrical engineering and computer sciences, University of California, Berkeley.

Kay, P. (2013). The Limits of (Construction) Grammar. In Th. Hoffmann, & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 32–48). Oxford & New York: Oxford University Press.

Kay, P. & Fillmore, C. J. (1999). Grammatical constructions and linguistic generalizations. The *What's X doing Y?* construction. *Language*, 75, 1–34. doi:10.2307/417472

Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What categories reveal about the mind*. Chicago: University of Chicago Press. doi:10.7208/chicago/9780226471013.001.0001

Langacker, R. W. (1987). *Foundations of Cognitive Grammar: Theoretical Prerequisites* (Vol. 1). Stanford: Stanford University Press.

Lau, J. H., Clark, A., & Lappin, S. (2014). Measuring Gradience in Speakers' Grammaticality Judgements. In Proceedings of the 36th Annual Conference of the Cognitive Science Society (pp. 821–826).

Lindström, J. & Linell, P. (2007). Roli å roli. X-och-x som samtalspraktik och grammatisk konstruktion ['Fun 'n fun. X-and-x as conversational practice and grammatical construction']. In E. Engdahl, & A. -M. Londen (Eds.), *Interaktion och kontext* (pp. 19–89). Lund: Studentlitteratur.

Linell, P. & Norén, K. (2009). Vågar vågar ni väl men … – en reaktiv konstruktion i svenskan ['Dare dare you, but … – a reactive construction in Swedish']. *Språk och stil*, 19, 72–104.

Loenheim, L., Lyngfelt, B., Olofsson, J., Prentice, J., & Tingsell, S. (2016). Constructicography meets (second) language education. On constructions in teaching aids and the usefulness of a Swedish constructicon. In S. de Knop, & G. Gilquin (Eds.), *Applied Construction Grammar* (pp. 327–355). Berlin: De Gruyter Mouton.

Lyngfelt, B. (2007). Mellan polerna. Reflexiv- och deponenskonstruktioner i svenskan ['Between the poles. Reflexive and deponent constructions in Swedish']. *Språk och Stil*, 17, 86–134.

Lyngfelt, B., Borin, L., Forsberg, M., Prentice, J., Rydstedt, R., Sköldberg, E., & Tingsell, S. (2012). Adding a Constructicon to the Swedish resource network of Språkbanken. In *Proceedings of KONVENS 2012* (LexSem 2012 workshop) (pp. 452–461). Vienna.

Lyngfelt, B, Magnusson Petzell, E., & Wide, C. (2017). Forskning om språksystemet – olika traditioner med olika syften ['Grammatical research – different traditions with different purposes'] In D. Håkansson, & A. -M. Karlsson (Eds.), *Varför språkvetenskap?* (pp. 129–152). Lund: Studentlitteratur.

Lyngfelt, B., & Sköldberg, E. (2013). Lexikon och konstruktikon – ett konstruktionsgrammatiskt perspektiv på lexikografi ['Lexicon and constructicon – a constructionist perspective on lexicography']. *LexicoNordica*, 20, 75–91.

Margolis, E., & Laurence, S. (Eds.). (1999). *Concepts: Core readings*. Cambridge, MA: MIT Press.

Murphy, G. L. (2002). *The big book of concepts*. Cambridge, MA: MIT Press.

Ohara K. H. (2014). Relating Frames and Constructions in Japanese FrameNet. In *Proceedings of LREC 2014* (pp. 2474–2477). Reykjavík: ELRA.

Pedersen, B., Borin, L., Forsberg, M., Kahusk, N., Lindén, K., Niemi, J., Nisbeth, N., Nygaard, L., Orav, H., Rögnvaldsson, E., Seaton, M., Vider, K., & Voionmaa, K. (2013). Nordic and Baltic wordnets aligned and compared through "WordTies". In *Proceedings of NODALIDA 2013* (pp. 147–162). Oslo: NEALT.

Petruck, M. R. L. & Ziem, A. (2014). Semantic Relations for Frames and Constructions: Perspective_on. Talk at the 8th International Conference on Construction Grammar (ICCG-8). September 3–6, 2014. Osnabrück University.

Pollard, C. & Sag, I. A. (1994). *Head-Driven Phrase Structure Grammar*. Chicago: University of Chicago Press.

Pustejovsky, J. (1995). *The Generative Lexicon*. Cambridge, MA: MIT Press.

Ranta, A. (2011). *Grammatical Framework. Programming with Multilingual Grammars*. Stanford: CSLI Publications.

Ruppenhofer, J., Ellsworth, M., Petruck, M. R. L., Johnson, C. R., & Scheffczyk, J. (2016). *FrameNet II: extended theory and practice*. Berkeley: ICSI. Retrieved from https://framenet2.icsi.berkeley.edu/docs/r1.7/book.pdf

Ruppenhofer, J. & Michaelis, L. A. (2010). A Constructional Account of Genre-Based Argument Omissions. *Constructions and Frames*, 2(2), 158–184. doi:10.1075/cf.2.2.02rup

Rydstedt, R. (2012). *En matchningsdriven semantisk modell. Mellan ordboken och den interna grammatiken* ['A match-driven semantic model. Between the dictionary and the internal grammar']. (Doctoral Dissertation). (GNS 19). Dept. of Swedish, University of Gothenburg.

Sag, I. A. (2012). Sign-Based Construction Grammar: An Informal Synopsis. In H. C. Boas, & I. A. Sag (Eds.), *Sign-Based Construction Grammar* (pp. 39–170). Stanford: CSLI Publications.

Sag, I. A., Baldwin, T., Bond, F., Copestake, A., & Flickinger, D. (2002). Multi-word expressions: A pain in the neck for NLP. In A. Gelbukh (Ed.), *Computational Linguistics and Intelligent Text Processing. Proceedings of CICLING-2002* (pp. 1–15). Berlin: Springer. doi:10.1007/3-540-45715-1_1

Sjögreen, C. (2015). *Kasta bort bollen och äta bort sin huvudvärk. En studie av argument-strukturen i kausativa bort-konstruktioner* ['The argument structure of Swedish causative bort-constructions']. (Doctoral dissertation.) Dept. of Scandinavian Languages, Uppsala University.

Sköldberg, E., Bäckström, L., Borin, L., Forsberg, M., Lyngfelt, B., Olsson, L. -J., Prentice, J. R., R. Rydstedt, S. Tingsell, & J. Uppström, (2013). Between Grammars and Dictionaries: a Swedish Constructicon. In *Proceedings of eLex 2013* (pp. 310–327). Tallinn: Eesti Keele Instituut.

Steels, L. (2013). Fluid construction Grammar. In Th. Hoffmann, & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 153–167). Oxford & New York: Oxford University Press.

Stefanowitsch, A. (2013). Collostructional Analysis. In Th. Hoffmann, & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 290–306). Oxford & New York: Oxford University Press.

Svensén, B. (2009). *A Handbook of Lexicography. The Theory and Practice of Dictionary-Making*. Cambridge & New York: Cambridge University Press.

Teleman, U., Hellberg, S., & Andersson, E. (1999). *Svenska Akademiens grammatik* ['The Swedish Academy Grammar']. Stockholm: Norstedts.

Torrent, T. T., Lage, L. M., Sampaio, T. F., da Silva Tavares, T., & da Silva Matos, E. E. (2014). Revisiting border conflicts between FrameNet and Construction Grammar: Annotation policies for the Brazilian Portuguese Constructicon. *Constructions and Frames*, 6(1), 33–50.

van Trijp, R., & Steels, L. (2012). Multilevel alignment maintains language systematicity. *Advances in Complex Systems* 15(3–4), 1–30.

Van de Velde, F. (2012). Degeneracy: The maintenance of constructional networks. *Leuven Working Papers in Linguistics* 1(2), 24–61.

Van Valin, R. D., & LaPolla, R. J. (1997). *Syntax: Structure, Meaning, and Function*. Cambridge: Cambridge University Press. doi:10.1017/CBO9781139166799

Wray, A. (2008). *Formulaic language: Pushing the boundaries*. Oxford & New York: Oxford University Press.

Towards continuity between the lexicon and the constructicon in FrameNet Brasil

Tiago Timponi Torrent, Ely Edison da Silva Matos, Ludmila Lage, Adrieli Laviola, Tatiane Tavares, Vânia Gomes de Almeida and Natália Sigiliano

This chapter presents the theoretical discussion, analytical procedures, and corresponding computational implementations carried out by FrameNet Brasil to take the principle of the continuity between grammar and the lexicon to the computational domain by deeply integrating two frame-based resources under development for Brazilian Portuguese: a lexicon and a constructicon. To achieve this goal, we start by discussing the continuity principle in the Construction Grammar paradigm, that is, the founding idea that both the lexicon and what is usually referred to in most traditions as grammar can be accounted for in terms of constructions. After, we present the computational solutions that led to the development of a new common database structure for the FrameNet Brasil Lexicon and Construction: FN-Br 2.0. Finally, we present the analyses of two constructions in Brazilian Portuguese: the Dative with Infinitive Construction, which illustrates currently available features of FN-Br 2.0, and the Inceptive Aspect Construction, which imposes a series of new challenges to the model.

Keywords: FrameNet Brasil, constructicon, lexicon, database structure, construction-to-construction relations, construction-to-frame relations

1. Introduction

The development of a Constructicon in FrameNet Brasil followed the path inaugurated in 2008 by the Beyond the Core Project, whose main purpose was to design a resource that, in complementation to the Berkeley FrameNet Lexicon, could account for those phenomena located beyond the semantic and syntactic affordances of lexical units (Fillmore, Lee-Goldman & Rhomieux, 2012, p. 311–314; Petruck & Lee-Goldman, this volume). Hence, building on the analyses of Brazilian Portuguese constructions carried out by various researchers at the Federal

University of Juiz de Fora (see Miranda & Salomão, 2008 for a collection of those analyses), the Brazilian Portuguese Constructicon initiated in 2010, by analyzing the `para_infinitive` Family of Constructions (Torrent, 2015).

At first, we attempted to deploy an adapted version of the same software used by Berkeley FrameNet to create both the Lexicon and the Constructicon: the FrameNetDesktop. We started both resources by not only expanding the Berkeley FrameNet database for frames into Brazilian Portuguese, but also by adapting the labels in the annotation tool to the morphological and syntactic characteristics of our language (Torrent & Ellsworth, 2013). Inasmuch as the work advanced, however, some questions related to basic assumptions of Construction Grammar arose:

1. How to model the fact that constructions may evoke frames?
2. How to model the fact that constructions are related to each other in a network?
3. How to model productive constraints on which kind of lexical material can fill a given slot in a construction?
4. How to model the continuity between grammar and the lexicon?

Those issues led us to rethink some aspects of our analytical approach, which, in turn led to the need of revising the database structure and software tools used to model the analyses. The solutions adopted in this process and the new software apparatus developed to model them – the FN-Br 2.0 – are the main topics of this chapter.

We start by presenting some FrameNet Brasil basics in Section 2. In Section 3, we discuss the implications of the continuity approach to the development of the lexicon and the constructicon, and present the computational effort carried out to cope with those implications: FN-Br 2.0. Section 4 brings two sample analyses of Brazilian Portuguese constructions that explore features of FN-Br 2.0 presented in the previous section, and present new challenges to the same model. The last section brings our conclusions.

2.   FrameNet Brasil[1]

FrameNet Brasil started in 2007 as a research project in the Graduate Program in Linguistics at the Federal University of Juiz de Fora. Because of that specific characteristic, the work in the Brazilian branch of framenet has always been deeply connected to the development of M.A. theses and PhD dissertations discussing the analytical solutions adopted by framenet to deal with phenomena that have been the focus of Cognitive Semantics and Construction Grammar for a long time.

---

1.   http://www.framenetbr.ufjf.br

From 2010 on, the former project – now a lab housing several research initiatives – evolved also into different directions, among which, we highlight three:

1. expanding the Berkeley FrameNet lexical database into Brazilian Portuguese;
2. creating a repertoire of Brazilian Portuguese constructions;
3. developing domain specific multilingual applications of framenet to non-specialist users.

Those initiatives led so far to the development of two databases: the FN-Br database, containing a general vocabulary lexicon and a constructicon; and the m. knob database, containing a multilingual – Brazilian Portuguese, English, Spanish and French – lexicon covering the vocabularies of Tourism and Sports (Torrent, Salomão, Campos et al., 2014; Costa & Torrent, 2017). This second database supports the FrameNet Brasil World Cup Dictionary, a web app designed to help tourists during the 2014 FIFA World Cup (Torrent, Salomão, Matos et al., 2014), and the Multilingual Knowledge Base, an app providing travel recommendations and sentence translations for tourists (Paiva & Torrent, 2017).[2]

The next sections present the FN-Br database in more detail, since this is the one in which the constructicon is included.

2.1 The FN-Br lexicon

The FN-Br lexicon is being expanded from the Berkeley FrameNet data release 1.7. The expansion process into Brazilian Portuguese includes:

1. adapting the annotation tool to the specificities of this language;
2. translating the names and descriptions of frames and frame elements;
3. adapting frames and frame elements to Brazilian Portuguese, in cases where it is necessary;
4. populating the database with the lexical units.

The adaptation of the annotation tool to Brazilian Portuguese led to the definition of the labels to be used in the annotation of the grammatical functions (GFs) and phrase types (PTs) of the linguistic material instantiating the frame elements (FEs) that manifest in the local context of target lexical units (LUs) (see Petruck and

---

2. The m.knob Lexicon comprises 87 trilingual frames, 70 of which did not exist in the Berkeley FrameNet data release 1.7. A total of 5,251 LUs are associated to the frames: 1,669 for Brazilian Portuguese, 2,551 for English, 930 for Spanish and 101 for French. The m.knob database has more than 13,000 annotation sets. The FrameNet Brasil World Cup Dictionary can be accessed at http://www.dicionariodacopa.com.br and the Multilingual Knowledge Base at http://mknob.com.

Lee-Goldman, this volume, for a concise explanation of the lexicographic annotation process in framenet). In this process, documented in Torrent & Ellsworth (2013), the properties of some labels were changed, while other labels were created, such as the Indirect Object label, for example.

Also, the criteria for applying the labels concerning Null Instantiations required adaptations. Although Fillmore (2007, p. 147–148) defines the difference between Definite (DNI) and Indefinite (INI) Null Instantiations primarily in terms of the kind of informational status of the FEs marked with these labels, both in his work and in that by Ruppenhoffer et al. (2016, p. 28–29), there is a clear proposal of treating this difference in terms of properties of the LUs. In other words, the main difference between a DNI and an INI is that, while the first is a zero anaphora to which it is possible to identify a referent, the latter is an existential omission that does not call for the identification of a referent in the context. However, FrameNet treats specially the indefinite type as a valence property of the LU. Hence, some verbs such as *eat* and *bake* would license INIs of the FEs INGESTIBLES and HEATING_INSTRUMENT, respectively.[3]

For example, with *eat*, INIs would be licensed in sentences such as the one presented in (2) as an answer to the question in (1), meaning that the speaker has already eaten something else before and, therefore, will not try the cake. Note that the thing eaten by the speaker is not the cake being offered, since, if it were, the answer would be the one presented in (3).

   (1)   Would you like to try some of this delicious cake?

   (2)   No, thanks, I already ate.

   (3)   No, thanks, I already ate it.

The same does not hold for Brazilian Portuguese, since the answer in (4) would be suitable for both scenarios.

   (4)   Não, obrigado, eu já   comi.
         No   thanks   I   already eat.PAST.1SG
         No, thanks, I already ate./No, thanks, I already ate it.

Therefore, due to the fact that verbs in Brazilian Portuguese – in general – license Direct Object omissions where there is either an anaphoric reference or an existential one, the informational status of the omitted FE is the only criterion taken into consideration, with no further attempt of capturing LU-specific properties

---

3.   Frame Element, as well as Construction Element names are represented in SMALL CAPS in this chapter.

concerning NIs: when it is possible to identify a specific referent, the DNI label is applied; when not, an INI label is used.

The last adaptation concerns the Constructional Null Instantiations. In FrameNet, the CNI label is used for those cases in which a grammatical construction licenses the omission of the constituent that would be assigned a FE label. Because English requires verb subjects to be overtly expressed in declarative sentences, and Brazilian Portuguese doesn't, CNIs in FrameNet Brasil include omitted subjects.

Processes 2, 3 and 4 in the expansion are performed simultaneously. As new LUs are added to the FrameNet Brasil database, frames and FEs – both their names and descriptions – are translated and/or adapted into Brazilian Portuguese. Linkage to the original English database is maintained both by the use of the same IDs for each expanded frame and FE in the FN-Br database – so that machines can track which Brazilian Portuguese frame is linked to which English frame –, and by the maintenance of the English names for each frame and FE in the report, next to the translated names – so that non-speakers of Portuguese can understand the general structure of the frames. So far, the FN-Br lexicon comprises 472 frames, 2,896 LUs and 2,386 annotated sentences.

Except for the points discussed above, the annotation process in the FN-Br database broadly follows the same guidelines defined by Ruppenhoffer et al. (2016) for both lexicographic and full-text annotation.

2.2   The FN-Br constructicon

Built in parallel with the FN-Br Lexicon, the FN-Br Constructicon followed, from the beginning, the directions of the Beyond the Core Project, since it also started by computationally representing constructions that had already been studied by Construction Grammarians. An example of such kind of construction is the `dative_with_infinitive` cxn (Torrent, 2015). Sentence (5) provides an instance of a construct licensed by the `dative_with_infinitive` cxn highlighted in bold.

(5) Ela deu   [**dinheiro**$_{\text{HEAD}}$] [**para mim   viajar**$_{\text{PARA\_SINF}}$]
     She give.PAST.3SG money   to   me.DAT travel.INF
     She gave the money for me to travel.

A traditional Berkeley Construction Grammar representation of such a construct is presented in Figure 1.
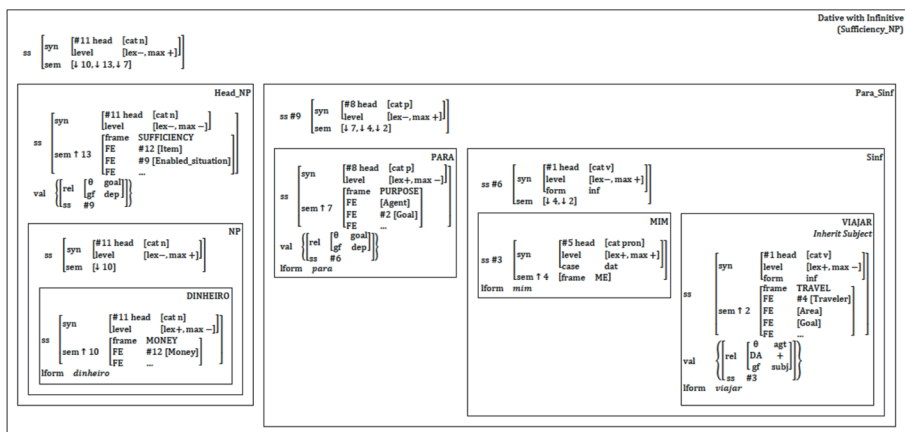
Figure 1. Berkeley-construction-grammar-style diagram for *"dinheiro para mim viajar"*

Each box in Figure 1 indicates one constituent in the construct. Three kinds of information may be provided for each constituent: its syntactic-semantic features (ss), its valence requirements (val) and its lexical form (lform). Except for the lform, the properties of each constituent are specified in terms of an Attribute Value Matrix (AVM). Values may be expressed as binary features (+/−), as items in a closed list (v, dat, subj, Traveler …), or go unspecified (…). Numbers 1 to 13 are unification indices, when preceded by an upward of downward arrow, those numbers indicate that some semantic import of the constituents is projected up to a daughter sign or to the whole construct.

Hence, by reading Figure 1 one may state that:

1. The construction licensing this construct is composed of two daughter signs, a HEAD_NP and an infinitival sentence headed by *para* (PARA_SINF), a preposition generally used in Portuguese for indicating purpose and direction. The HEAD_NP evokes the Sufficiency frame, which is composed of two core Frame Elements (FE), the ENABLED_SITUATION and the ITEM enabling it.[4]
2. The PARA_SINF sign is, in turn, composed of two daughters, its head, the preposition PARA (#8) and an infinitival sentence (SINF), which fulfills the valence requirements of PARA, as indicated by the unification index #6. The syntax and semantics of SINF unify with that of VIAJAR (#1, #2), the infinitival head verb in this sentence.

---

4. In FrameNet, FEs may be assigned three different coreness statuses: (a) core, when the frame does not exist without the FE; (b) peripheral, when the FE adds circumstantial information relevant, but not necessary, to the frame; and (c) extra-thematic, when the circumstantial information is not dependent on the frame at any level, and is usually introduced by a construction that happens to occur in the sentence.

3. In turn, the PARA_SINF sing fulfills the valence requirements of the HEAD_NP (#9). Semantically, it unifies with the ENABLED_SITUATION FE, while the ITEM FE is mapped to the MONEY FE in the Money frame evoked by DINHEIRO (#12).

Since Figure 1 brings the representation of a construct – i.e. of a piece of language licensed by a construction (Fillmore et al., 2012) – it does not match entirely with the representation of the dative_with_infinitive cxn, which is more generic: virtually any infinitival verb may occupy the head slot of the PARA_SINF sign, as well as any noun that can serve as a resource enabling a situation may be the head of the NP. Also, the subject of the infinitival verb may be left unspecified, because there are apparently no restrictions that apply to this slot. The […] notation indicates that those features are left unspecified in the construction. On the other hand, the preposition *para* must be always present.

Moreover, although the lexical constructions filling the slots are not present in the representation of the dative_with_infinitive cxn itself, the unification restrictions proposed in Figure 1 are still valid, since they are dependent on the whole construction, not on each specific lexical possibility on its own. In other words, because the Sufficiency frame is evoked by the construction, the FE ITEM will be mapped to whichever Noun heads the NP, while the FE ENABLED_SITUATION will be described in terms of whichever infinitival verb in the PARA_SINF daughter.

The nested boxes diagram of the construction itself is presented in Figure 2.
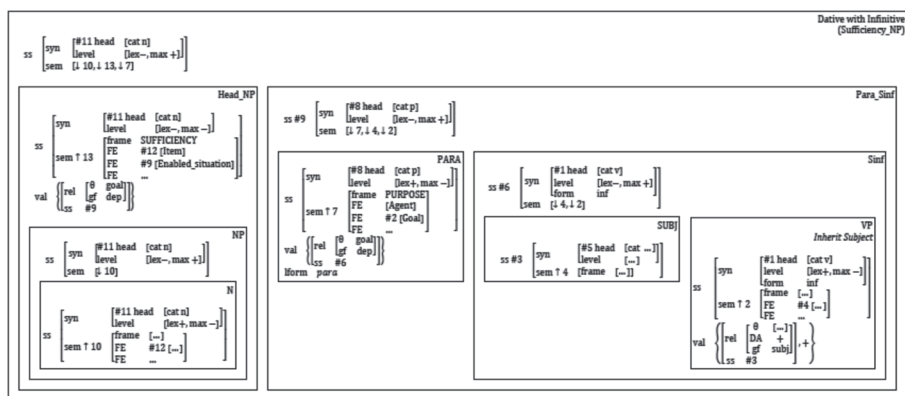
Figure 2. Berkeley-construction-grammar-style diagram for the dative_with_infinitive cxn

The challenge of developing a constructicon is that of computationally representing the restrictions and possibilities of constructions such as those depicted in Figure 2, which includes, for instance, how the elements in the construction relate to one another, or how the formal pole relates to the semantic pole. Hence, computational models of constructions must address, from the beginning, both constituency and

unification. In the Berkeley FrameNet Constructicon, constituency is addressed by the creation of Construction Elements (CEs), the daughter signs of the construction. When the construction evokes a frame, the CEs refer to the FEs in the frame evoked by the construction. Unification is handled by annotating instances of the construction in a multilayer fashion. During this annotation process, labels specifying the Grammatical Function (GF) and Phrase Type (PT) of the valents in the construct are aligned to the CEs (Fillmore et al., 2012, p. 321–324).

However, in the FN-Br Constructicon, construction modeling is somehow different to the one carried out in the original Berkeley Constructicon. The most prominent of those distinctions were formalized by Lage's (2013) annotation policies, which aim to provide objective criteria for both construction creation and annotation.

In regards to the creation of constructions and their daughter signs, the FN-Br Constructicon creates CEs based more on formal aspects than on functional/semantic features. Hence, if a construction evokes a frame, instead of creating CEs that refer to the FEs in the evoked frame, in FN-Br we create CEs such as HEAD_NP or PARA_SINF, and then link those CEs to the FEs in the appropriate frame, as it will be shown in Section 4. In other words, instead of addressing the unification of syntactic and semantic features only in annotation and in the prose description of the constructions and CEs, we also formalize it via relations in the database when such unification is part of the general properties of the construction. In this sense, besides being different from the Berkeley Constructicon, the process of construction creation in FN-Br also differs form that of the Swedish Constructicon, in which constructions feature generic semantic roles (such as Agent, Patient and so on).

For instance, in the case of the `dative_with_infinitive` cxn, a construction to frame relation in the database maps the CEs HEAD_NP (NP_NÚCLEO) and PARA_SINF to the FEs ITEM and ENABLED_SITUATION (SITUAÇÃO_HABILITADA) in the `Sufficiency` (Suficiência) frame, respectively. When it comes to annotating an instance of the construction, as in (5), such a relation automatically assigns the relevant FE labels as those of the CE are applied to the construct, as can be seen in Figure 3.

Figure 3. An annotated example of the `dative_with_infinitive` cxn in the FN-Br web tool

Although the annotation depicted in Figure 3 captures the form-meaning unification processes that are constant in the `dative_with_infinitive` cxn, it does not capture all the processes that are relevant to the meaning of the sentence. There are other aspects of such a meaning that are contributed by the lexical constructions in the sentence and their valence affordances. In practical terms, it means the annotation of a sentence such as (5) would have FEs that are evoked by the LUs *dar.v* 'give', *dinheiro.n* 'money', *para.prep* 'for', *viajar.v* 'travel' and some others evoked by the `dative_with_infinitive` cxn. Figure 4 shows how sentence (5) is fully annotated in FN-Br.

Figure 4. The complete annotation of *Ela deu dinheiro para mim viajar*
in the FN-Br web tool

As it may be seen in Figure 4, there are several layers of annotation associated with the sentence:

1. Evoked by the verb *dar* 'give', the core FEs Donor (Doador), Theme (Tema) and Recipient (Recipiente) in the Giving (Dar) frame are assigned, respectively, to *Ela* 'she', *dinheiro* 'money' and *para mim viajar* 'for me to travel', together with the relevant labels for the GFs and PTs for each valent.

2. In the case of the LU *dinheiro* 'money', evoking the `Money` (`Dinheiro`) frame, the FE Money is incorporated by the LU, meaning its realization is in the LU itself, not in its local syntactic context. Again, a non-core FE, this time the Inherent_purpose (Uso), is instantiated by the infinitival sentence headed by *para*.[5]

3. The `Purpose` (`Finalidade`) frame, evoked by *para* 'for', has as core FE the Goal (Finalidade) represented by the infinitival sentence.

4. Last, the `Travel` (`Viagem`) frame, evoked by *viajar* 'travel', is instantiated via the core FE Traveler (Viajante), linguistically manifested by *mim* 'me'.[6]

Although each LU, as well as the `dative_with_infinitive` cxn, evokes a particular frame and contributes to the meaning of the sentence in a specific way, the frames evoked interact with one another. Moreover, although LUs and constructions may be referred to as different units of language, such a distinction is more due to the historical development of Berkeley FrameNet (as a lexicon to which a constructicon is being added) than to a conceptual difference between those types of units: both kinds of units may be treated similarly when it comes to annotation, and, most importantly, both kinds of units are constructions. Hence, the improvements FrameNet Brasil has been making in its analytical tools, such as the web annotation tool depicted in Figures 3 and 4, aims to allow for deeper integration between the two kinds of annotation (lexical and constructional) and between the databases derived from each of them. In other words, our current effort is to adequately model the continuity between Brazilian Portuguese grammar and lexicon by developing a database system in which Lexicon and Constructicon interact. Let's turn now to this specific issue.

---

5.   We will return to the question about whether the Imposed_purpose and Inherent_purpose FEs should be included in the annotation for the LUs *dar* 'give' and *dinheiro* 'money' by the end of this chapter.

6.   All frames in this example are the same for both English and Portuguese, and have been expanded from Berkeley FrameNet into FrameNet Brasil with no adaptations to the structure.

3.   Modeling the continuity between grammar and the lexicon

Fillmore, Kay & O'Connor (1988), while introducing the model sustaining their analysis of the *let alone* constructions, state one of the most recognizable pillars of Construction Grammar: the continuity between grammar and the lexicon. Such a statement not only set a clear differentiation between the constructionist approaches and the preceding derivationist paradigm, but also allowed each Construction Grammar to develop one single set of analytical tools to account for lexical items, idioms, argument structures and so on. As the constructionist approach that later came to be known as Berkeley Construction Grammar (BCG) developed, the analyses of constructions, from the fully lexical to the highly schematic, were proposed in terms of complex feature structures that neither derived from assumed deep structures, nor presented empty categories. Unification of such features acts as the main formal operation in this model – as demonstrated in Figures 1 and 2 – and no transformation rules are predicted. Hence, as one of the most famous slogans of Construction Grammar would state it: "What you see is what you get" (Fillmore, 2013, p. 111–113).

Together with feature structures and unification, valence descriptions are also an important part of BCG analyses. According to Fillmore (2013, p. 118–119), in BCG, valences include both obligatory and optional valents, that is, both arguments and adjuncts, in syntactic terms. There is, nevertheless, a distinction between core and peripheral valents. Fillmore adds that the complete valence of a lexical construction includes specification about the semantic roles, grammatical functions and phrase types for each valent.

The resemblance between such a proposal and the valence descriptions found in framenets is neither coincidental, nor due to the fact that Fillmore wrote his 2013 chapter on Berkeley Construction Grammar after Berkeley FrameNet was already a mature research initiative. The idea of matching semantic information to syntactic behavior is present in Fillmore's work at least since 1968 – in *The Case for Case* – and more prominently in his 1977 paper *The Case for Case Reopened*. FrameNet lexicographic annotation is, thus, inspired by the same kind of epistemological background that led to the development of BCG.

Nevertheless, advocating in favor of the continuity between grammar and the lexicon, by assuming that lexical items, as well as morphemes, are constructions themselves does not mean to abandon the differentiation among those, idioms and clause-level constructions, or, as Langacker (2008, p. 6–7) phrases it:

> Overlap among lexicon, morphology, and syntax does not prevent us from defining them and drawing useful distinctions, any more than the absence of a precise boundary between green and blue condemns us to seeing only grue – a gradation does not imply undifferentiated homogeneity.

In a similar direction, Goldberg (1995, p. 7) also points out that by not positing a strict division between grammar and the lexicon, constructionists do not deny the existence of distinctly lexical and syntactic constructions.

Those theoretical claims of Construction Grammar set the ground for the development of a computational resource that, at one time, is capable of bridging the gap between a frame-based lexicon and a constructicon, while still maintaining some methodological boundaries when it comes to deciding whether a piece of language should be accounted for in the lexicon, in the constructicon or in both (see Torrent, Lage, Sampaio, Tavares & Matos, 2014a for a discussion of such a methodology).

Given the theoretical background briefly presented in the paragraphs above, modeling the continuity between grammar and the lexicon presents itself as a feasible task in a framenet, provided that some minor aspects of the database structure be incremented in order to promote a gradual interconnection among lexical units, constructions and their semantic import: the frames. We will, in the next two sections of this chapter, (1) present the requirements for such a model and (2) show how the original FrameNet database structure was changed so as to accommodate those requirements in one possible implementation.

3.1   Requirements for modeling the continuity between grammar and the lexicon

The first step in defining the requirements for a database featuring deeper integration between the constructicon and the lexicon was to analyze the properties of the three main entities in such a database: frames, lexical units and constructions.

Frames are defined by Fillmore (1982, p. 111) as "any system of concepts related in such a way that to understand any one of them you have to understand the whole structure in which it fits". When this seminal concept was "translated" into the idea of FrameNet (Fillmore & Atkins, 1992; Fillmore, Petruck, Ruppenhofer & Wright, 2003) and interconnected frames became the basis for a lexical resource, they were modeled as having a name, a definition usually followed by a set of examples, a list of participants and props (the FEs) involved in the scene being described, and a set of both internal and external relations. As for the FEs, they also feature a definition, sometimes followed by examples, a semantic type and a set of relations as well. Although the frame reports only show the internal relations in which FEs take part, every frame-to-frame relation is also a FE-to-FE relation. As an example, consider the `Travel` frame in Figure 5.

Travel

Definition:

In this frame a Traveler goes on a journey, an activity, generally planned in advance, in which the Traveler moves from a Source location to a Goal along a Path or within an Area. The journey can be accompanied by Co-participants and Baggage. The Duration or Distance of the journey, both generally long, may also be described as may be the Mode_of_transportation. Words in this frame emphasize the whole process of getting from one place to another, rather than profiling merely the beginning or the end of the journey.

> Ellen JOURNEYED to Europe with five suitcases.

> Samantha JOURNEYED 2500 miles with her family by sea to China.

> The Osbournes *took* a TRIP from Beverly Hills to London on the Concorde.

FEs:

Core:

Area [Area] **Semantic Type:** Location	This is the Area in which the traveling takes place. This frame element describes the enclosed area inside which travelling, of unspecified Source, Path or Goal takes place. We TRAVELLED in Europe.
Direction [dir] **Excludes:** Area	The direction in which the Traveler goes. They began their ODYSSEY north.
Goal [Goal] **Semantic Type:** Goal **Excludes:** Area	The Goal is the location where the travelers end up.
Mode_of_transportation [MoT]	The Mode_of_transportation expresses how the motion of the Traveler is effected, by their body or by a vehicle which holds and conveys the Traveler. Vehicles can move in any way and in any medium. They are usually expressed obliquely with 'in' or 'by'. Barney used to TRAVEL by bus a lot Strom TRAVELLED on foot to see the Pope
Path [Path] **Semantic Type:** Path **Excludes:** Area	The Path is the route along which the travel takes place.
Source [Src] **Semantic Type:** Source **Excludes:** Area	The Source is the starting point of the trip.
Traveler [Trav] **Semantic Type:** Sentient	This is the living being which travels. Normally, the Traveler is expressed as an external argument.

Non-Core:

Baggage [Bag]	The Baggage are the items necessary for travel that accompany the Traveler. Ellen JOURNEYED to Europe with five suitcases.
Co-participant [co-p] **Semantic Type:** Sentient	The Co-participant is the person or persons who accompany the Traveler on the journey.
Depictive [Depict] **Semantic Type:** State	The state of the Traveler during the journey. We TRAVELED around unencumbered.
Descriptor []	A characterisitic of the traveling event.

Distance [Dist]   This FE dientifies the Distance traveled.
Semantic Type: Quantity
Duration [Dur]   This FE identifies the Duration of time during which the trip occurs.
Semantic Type: Duration
Explanation [Exp]   The Explanation for which the travel is undertaken.
Semantic Type: State_of_affairs
Frequency [Freq]   The Frequency with which the Traveler makes the journey.

Iterations [Iter]   The number of times the trip is traveled by the Travelers.

Manner [Manr]   The Manner in which the traveling takes place.
Semantic Type: Manner   The Brownie troop TRAVELLED hurriedly.

Means [Mns]   The action done by the Traveler by which the traveling is accomplished.
Semantic Type: Human_act   I TRAVELED to London by hitchhiking and walking.

Period_of_iterations [peri]   The Time throughout which the traveling repeatedly takes place.

Place []   The Place is the location of the motion. It describes the area in which motion (with a
Semantic Type: Locative_relation   specified Source, Path or Goal) takes place.

Purpose [Purp]   This FE identifies the Purpose for which the Traveler travels, i.e. the future state of affairs
Semantic Type: Human_act   that will hopefully come about.

Result [Res]   The affect of travel on a Traveler.
Semantic Type: Event   We JOURNEYED to exhaustion.

Speed [Spd]   The Speed is the rate at which the travel takes place.
Semantic Type: Speed   The Simpsons TRAVELLED at 60 km per hour.

Time [Time]   When the traveling occurs.
Semantic Type: Time
Travel_means [Trvl_mns]   This FE applies to expressions that indicate documents, money, tickets, etc. which allow
                                   people to engage in travel.
                                       He TRAVELED to San Diego on a standby ticket.

FE Core set(s):

{Direction, Goal, Path, Source}, {Mode_of_transportation, Traveler}

Frame-frame Relations:

Inherits from: Self_motion
Is Inherited by:
Perspective on:
Is Perspectivized in:
Uses:
Is Used by:
Subframe of:
Has Subframe(s): Setting_out
Precedes:
Is Preceded by:
Is Inchoative of:
Is Causative of:
See also:

Figure 5. The Travel frame in Berkeley FrameNet

Definitions and examples, as well as the background colors used in the reports are mostly meant to help human users in their experience with framenet data. We will thus focus on the status and semantic type of FEs and on the relations.

Each FE may be assigned a Semantic Type, whose function is to capture commonalities among the frame-specific semantic roles. The Sentient Semantic Type, for example, applied to the FE Traveler in the Travel frame, indicates that the entity instantiating this FE must be in control of its will to carry out actions. Types may also be assigned to frames, indicating whether they are evoked by lexical items or not, and to LUs, indicating their semantic polarity, for example.

Because the annotation of non-core – that is, peripheral and extra-thematic – FEs is not mandatory, they do not take part in frame-internal relations, which are meant to model the fact that FEs that are necessary for the frame to be instantiated may be necessary in three different ways. The first kind of frame-internal relation is Excludes. In such a relation, the instantiation of a given FE precludes the others related to it from being instantiated. In the Travel frame, the FE Area excludes Goal, Path and Source. Requires is the opposite of the Excludes relation and holds between two FEs that must be instantiated together (Ruppenhofer et al. 2016).

The third kind of frame internal relation is the Core set. Core sets occur when the presence of a given core FE makes the instantiation of the other FEs in the set optional (Ruppenhofer et al. 2016). In the Travel frame, the Goal, the Path and the Source are in a Core set, meaning that one can mention one, two or all of those FEs, as shown in (6)–(8), respectively.

(6)   [I$_{\text{Traveler}}$] traveled^{Target} [to Carmel$_{\text{Goal}}$].

(7)   [I$_{\text{Traveler}}$] traveled^{Target} [from San Francisco$_{\text{Source}}$] [to Carmel$_{\text{Goal}}$].

(8)   [I$_{\text{Traveler}}$] traveled^{Target} [from San Francisco$_{\text{Source}}$] [to Carmel$_{\text{Goal}}$] [along Hwy 1$_{\text{Path}}$].

In regards to frame-to-frame relations, Berkeley FrameNet defines eight of them: Inheritance, Using, Perspective_on, Subframe, Precedes, Causative_of, Incohative_of and See_also. The Travel frame inherits from Self_motion, meaning that traveling is a kind of self-propelled motion. In computational terms, it means that every core and peripheral FE in the Self_motion frame must be mapped to a FE in the Travel frame. Travel also has a subframe, Setting_out, meaning that the latter is a separate event that happens inside the first. For the Subframe relation, as well as for the other ones, there's no such strict mapping requirements as the one needed for Inheritance relations. A diagram with frame-to-frame relations involving the Travel frame is presented in Figure 6. The FE-to-FE relations sustaining the Inheritance between Travel and Self_motion are also shown.
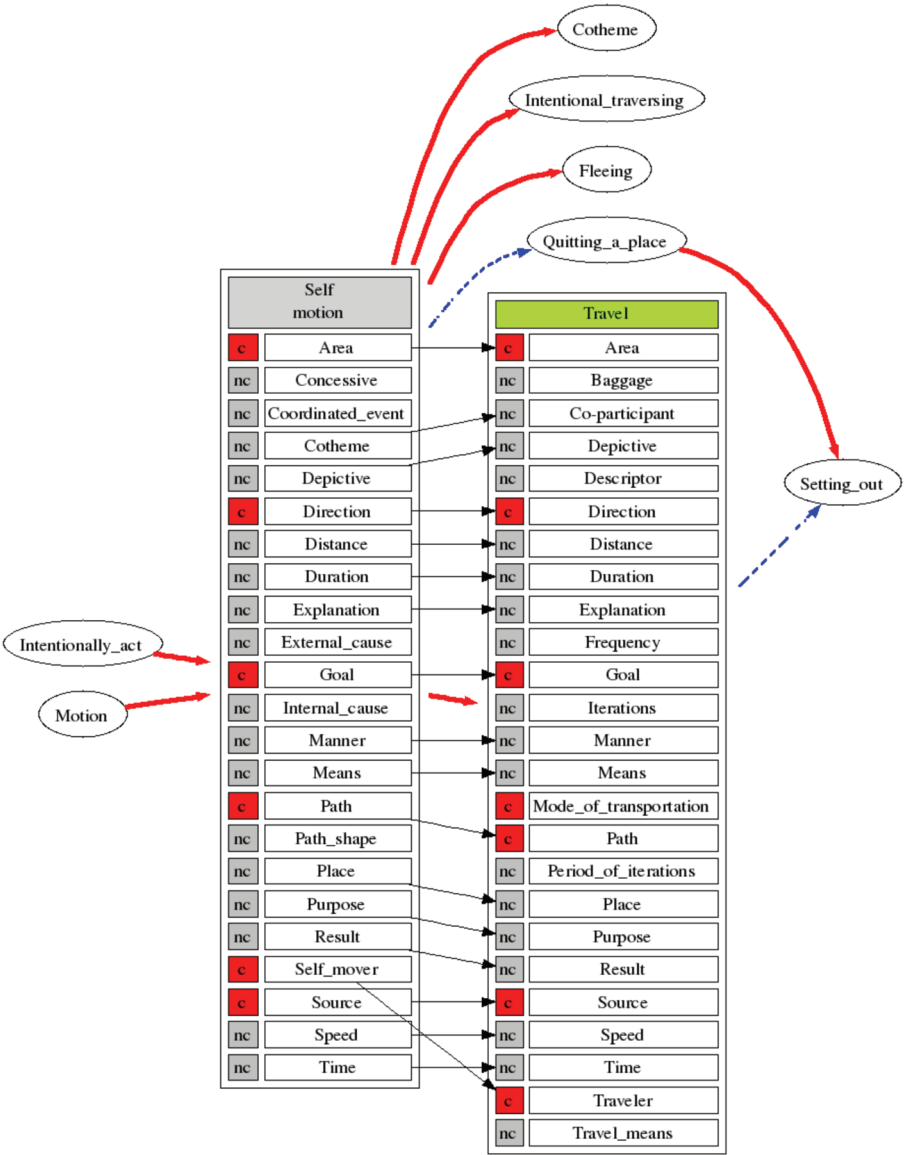
Figure 6. The frame-to-frame relations involving the `Travel` frame as shown in Berkeley FrameNet's FrameGrapher.

The LUs evoking the `Travel` frame, according to the FrameNet database for English are: *commute.v, excursion.n, expedition.n, getaway.n, jaunt.n, journey.n, journey.v, junket.n, odyssey.n, peregrination.n, pilgrimage.n, safari.n, tour.n, tour.v, travel.n, travel.v, traveler.n, trip.n, voyage.n, voyage.v*. For each one of these LUs, there will be (a) an associated lemma, which, in turn, will have one or more associated lexemes with their inflection possibilities – that is, their word forms; (b) a part of speech; (c) a definition; (d) annotation sets exemplifying the instantiation patterns of the FEs in the local syntactic context of the LU; and (e) valence patterns derived from annotation.

Last but not least, constructions share similarities with both frames and lexical units. On the frame side, constructions also have an elaborate internal structure of constituents, the CEs, although they also have an external syntax that does not correlate to any property of frames. They also are related to each other in a network of construction-to-construction relations, at the same time that their daughter signs may be related to each other in different ways. On the LU side, constructions may also evoke frames, and may require the presence of some specific lexical material.

Given the properties of frames, LUs and constructions just presented, plus the fact that frames can be evoked by both LUs and constructions, as shown in Figures 1–2, the requirements for modeling those three entities in one single integrated framenet database may be summarized as shown in Figure 7.
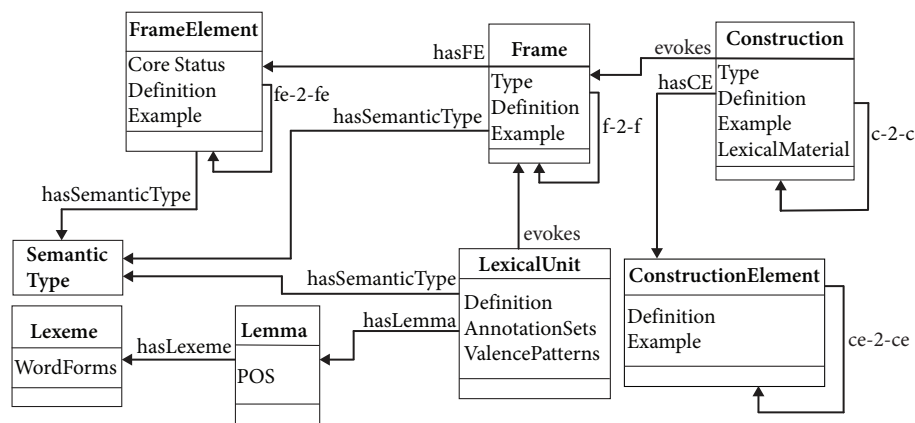
Figure 7. Requirements specification for modeling frames, LUs and constructions in a framenet

Since the requirements specification is outlined, we now move to presenting the implementation effort carried out to model the theoretical issues presented so far.

3.2   FN-Br 2.0

FN-Br 2.0 is the computational implementation developed to model the theoretical issues presented so far. The relational model, the same as the one used in Berkeley FrameNet (Baker et al., 2003), was preserved for the database so as to make FN-Br 2.0 easier to align with other framenets. Also following Berkeley FrameNet, the Relational Database Management System (RDBMS) used is MySQL.

The database structure was modeled based on four premises:
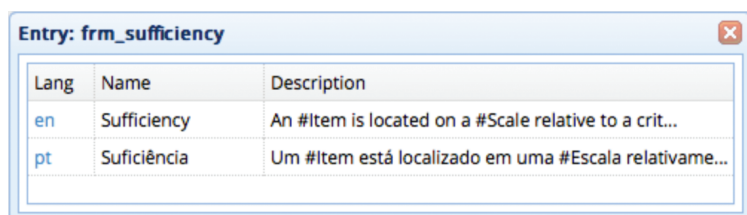
1. Enhance data consistency and integrity;
2. Support multilinguality;
3. Facilitate the creation of relations between the entities in FrameNet Brasil;
4. Reduce the number of auxiliary tables.

Data consistency is incremented by the use of Foreign Keys (FKs), a resource present in most RDBMSs (Elmasri & Navathe, 2010). FKs help maintain the integrity of the relations between tables in the database because a given record can only refer to records in other tables if the Primary Key (PK) that is referenced actually exists. Such a feature precludes records being referenced to by other entries to be removed from the database by some user's mistake or a flaw in the system. In FN-Br 2.0, FKs are also indexed, reducing the time necessary to access the data in join operations between tables.

Multilinguality support was restructured as well. In the first version of the database supporting the FrameNet Brasil World Cup Dictionary (Torrent, Salomão, Campos et al., 2014), each language-specific representation of a frame was a separate record in the database. Hence, working with three different languages demanded the creation of three records in the Frame table. A new frame-to-frame relation (Translation) had to be created to associate the three frames to one another. Although this is a fully functional solution – it was used for the FrameNet Brasil World Cup Dictionary –, it generates two problems: the complexity in information retrieval (through SQL queries) is augmented, and the occurrence of inconsistencies becomes more frequent, because the FE-to-FE relations must also be repeated every time a Translation relation is posited.

In FN-Br 2.0 a new approach was adopted. Fields requiring translations – such as Name, Description and so on – are shared by many components of the framenet model. Those fields were grouped in the Entry table together with the fields Entry and Language. The Entry table is shared by every component that might need a multilingual representation. In the case of frames, for example, one single record is created for each frame. This record has an Entry field referencing the Entry table. For each language, only one join operation is needed between the Frame and Entry

tables so as to generate a language-specific representation for that frame. Beyond initiatives that are multilingual from the very beginning – such as the m.knob project –, this feature allows for a more controlled expansion of the Berkeley FrameNet database into Brazilian Portuguese, since all the correspondences – them being partial or complete – between English and Brazilian frames are managed through the Entry table. Figure 8 illustrates this process, with the `Sufficiency` frame in the FN-Br 2.0 database.

Entry: frm_sufficiency			
Lang	**Name**	**Description**	
en	Sufficiency	An #Item is located on a #Scale relative to a crit...	
pt	Suficiência	Um #Item está localizado em uma #Escala relativame...	

Figure 8. Entries for the `Sufficiency` frame in FN-Br 2.0

As shown in Section 3.1, modeling the continuity between grammar and the lexicon implies the establishment of relations (and self-relations) among the components in the Lexicon and in the Constructicon. In the relational model, such relations are implemented as associations between tables – more specifically, between the records in the tables –, through the use of FKs. Nevertheless, there are also cases in which a given record in Table A, for example, may be associated to many records in Table B, and vice-versa, in a many-to-many relation. In those cases, relational models require the creation of additional tables to represent associations. In a framenet, the typical case is that of frame-internal relations. Take the `Travel` frame depicted in Figure 5, for example. As it may be seen, the FE AREA excludes the FEs SOURCE, PATH and GOAL. To model these relations, an additional table is created every time a relation is posited between AREA and the other three FEs, as well as additional tables are created for every FE-to-FE relation structuring the core set. A greater number of tables augments the complexity of the model and demands constant maintenance of the programs accessing the database.

FN-Br 2.0 adopted a different strategy. An *Entity* table was created, representing the components in the model at a higher level of abstraction. Each component that may be involved in a many-to-many relation with another component is considered to be a type of *Entity* – in fact, such components inherit from an *Entity*. In other words, each table representing a component is associated to the *Entity* table. Because each component is an entity, a relation between components can be abstracted as a relation between entities. The extant relations between entities are

stored in the *EntityRelation* table. Each relation has a specific type, stored in the *RelationType* table. The relation types are also clustered in groups, allowing for the distinction between, for instance, frame-to-frame relations and FE-to-FE relations.

This strategy fits well with the discussion presented in Section 3: there is a gain in terms of generality – blurring the limits between lexicon and grammar – at the same time that the specialization of each entity is allowed through the definition of different and specific types of relations. Moreover, the creation of new types of relations between components becomes more flexible, without the need to change the database structure. Currently, the following components are treated as entities: *Construction, ConstructionElement, Frame, FrameElement, GenericLabel, Label, LayerType, LU, SemanticType, Property, SubCorpus, Template, TypeInstance* and *POS*.

Finally, in relational models, *Status* and *Type* tables are common. *Status* tables store records indicating the statuses of other records in the database. For instance, an *AnnotationSetStatus* table may store the possible statuses for an *AnnotationSet*. *Type* tables, on the other hand, store records indicating the association of a given type to another record. For example, the *InstantiationType* table can store the types of null instantiations of a FE in a sentence. Usually, such tables feature a small fixed number of records, which must be related to the main table. The creation of new tables like these two leads to a structural change in the model.

FN-Br 2.0 does not use tables for specific types and statuses. Two other tables – *Type* and *TypeInstance* – are used for this function. Each record in the *Type* table corresponds to a specific type (*CoreType, InstantiationType, FramalType, LexiacalType, StatusType, AnnotationStatusType*). Values for each specific type are stored in the *TypeInstance* table. Each type is associated with its values through the *hasType* and *hasStatus* relations. Therefore, new types can be easily created (as well as new values can be added to the existing types) without the need to change the structural model of the database.

Because of the implementations presented in this section, FN-Br 2.0 provides the computational environment needed for the development of analyses that embrace the continuity between grammar and the lexicon. In Section 4 we will present two sample analyses that demonstrate the potential of FN-Br 2.0 for accounting for the relation between frames and constructions.

4.   Sample analyses

In this section, we analyze two constructions in Brazilian Portuguese, the `dative_with_infinitive` cxn, deploying the features made available by FN-Br 2.0, and the `inceptive_aspect` cxn, which presents new challenges that must be addressed by the model.

4.1 The dative with infinitive construction

Two aspects of the model proposed for the `dative_with_infinitive` cxn will be analyzed in the following subsections: first, we show how frame evocation is treated in FN-Br 2.0; second, we demonstrate how inheritance relations between constructions are accounted for.

4.1.1 *Constructions may evoke frames*

As shown in Section 2.2, the `dative_with_infinitive` is a frame-bearing construction, evoking the `Sufficiency` frame. Moreover, there is a one-to-one relation between the CEs in the construction and the FEs in the relevant frame. The former FrameNet Brasil software apparatus and database structure, which were derived from those of Berkeley FrameNet, did not support the creation of construction-to-frame relations such as the one needed to properly account for the semantic import of the `dative_with_infinitive` and many more constructions in Brazilian Portuguese.

Since both frames and constructions are entities in FN-Br 2.0, a new relation type was created in the database: the Evoking relation. In this relation, a construction is mapped to the frame it evokes and, in case there is a CE-to-FE correspondence, this information is also stored in the database, allowing the annotation tool to automatically assign the relevant FE labels to the linguistic material instantiating the CEs once they are annotated. Figure 9 shows the Relation Editor tool in FN-Br 2.0. Note that the CE HEAD_NP (NP_NÚCLEO) in the `dative_with_infinitive` cxn maps to the FE ITEM in the `Sufficiency` frame, while the CE PARA_SINF maps to the FE ENABLED_SITUATION (SITUAÇÃO_HABILITADA). The existence of such a mapping allows for the automatic annotation of the FEs shown in Figure 3, Section 2.2.
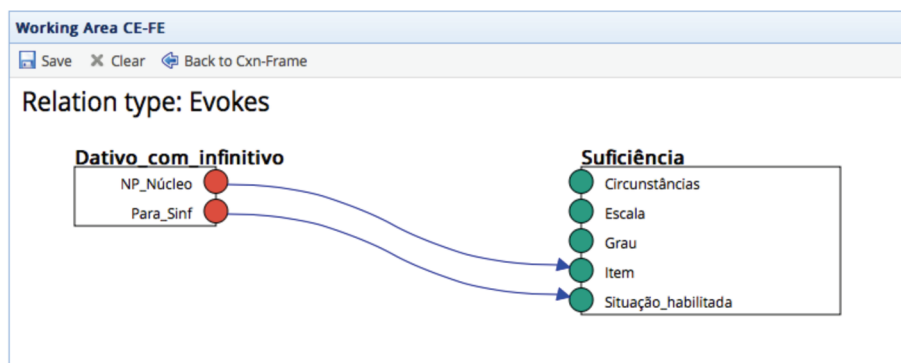
Figure 9. The evoking relation in FN-Br 2.0

Although having a similar name, the relation represented in Figure 9 is not the same as the *evokes* keyword in Embodied Construction Grammar (ECG). First, in ECG, the meaning pole of a construction is represented in terms of embodied schemas (Bergen & Chang, 2005, p. 151), which may be frame-like, but also include cognitive structures such as executing-schemas (x-schemas) and image schemas (Dodge, 2010, p. 43–44; Bergen & Chang, 2013, p. 177–178). Second, in ECG, the *evokes* keyword is not meant to relate a construction to a schema in terms of how the formal pole of each daughter sign of the construction maps to its meaning pole as defined in terms of a given frame, but, rather, to indicate the relation between one schema and the background schema(s) against which it is to be defined (Bergen & Chang, 2005, p. 152).

Dodge (2010, p. 47–50) provides an example of the use of the *evokes* keyword in ECG for the definition of the meaning import of the lexical constructions for the prepositions *in* and *out*. In the representation depicted in Figure 10, the form constraints for each construction specify the orthographic form of each preposition. As for the meaning, it is shown that both prepositions have their meaning defined in terms of the Trajector-Landmark (TL) schema, although perspectivized differently in each case. Such a difference in perspective is accounted for by stating that in both cases, the Bounded Object (BO) schema is evoked as a background against which the TL schema is to be defined: while in the case of *in*, the profiled area role of the TL schema is mapped to the interior role of the BO schema, in the case of *out*, the same role of TL is mapped to the exterior role of BO.

construction IN1	construction OUT1
subcase of LocativePreposition	**subcase of** LocativePreposition
form	**form**
constraints	**constraints**
self.f.orth ← "in"	self.f.orth ← "out"
meaning: TL	**meaning**: TL
evokes BoundedObject as bo	**evokes** BoundedObject as bo
constraints	**constraints**
self.m.landmark ⟷ bo.whole	self.m.landmark ⟷ bo.whole
self.m.profiledArea ⟷ bo.interior	self.m.profiledArea ⟷ bo.exterior

Figure 10. The representation of the constructions IN1 and OUT1 in ECG (Dodge, 2010, p. 49)

The kind of relation explicated above may also be useful for the FN-Br Constructicon, as it has proven to be so in a computational representation of ECG, the ECG Analyzer (Bryant, 2008). However, further studies are still needed in order to define the scope, constraints and application of such a relation. For now, the Evoking relation in the FN-Br Constructicon has the sole purpose of formalizing the mapping between the internal structure of constructions with that of frames.

More than allowing the partial automation of the annotation process, such a mapping may lead to a reassessment of some of the decisions made by Berkeley FrameNet in the process of frame creation. As an example, let us compare (9), (10) and (11):

(9) Eu recebi um dinheiro para pagar as contas.
   I receive.PAST.1SG a money to pay.INF the bills
   I received some money to pay the bills.

(10) Eu comprei um livro para estudar pra prova.
    I buy.PAST.1SG a book to study.INF to test
    I bought a book to study for the test.

(11) Eu tenho dinheiro para pagar as contas.
    I have.PRES.1SG money to pay.INF the bills
    I have the money to pay the bills.

If we consider the verbs *receber* 'to receive', *comprar* 'to buy' and *ter* 'to have' as lexical targets, (9), (10) and (11) can be annotated for the `Receiving`, `Commerce_buy` and `Possession` frames, respectively, as shown in (9a)–(11a):

(9a) [Eu$_{\text{RECIPIENT}}$] *RECEBI^{Target}* [um dinheiro$_{\text{THEME}}$] [para pagar as contas$_{\text{PURPOSE\_OF\_THEME}}$]

(10a) [Eu$_{\text{BUYER}}$] *COMPREI^{Target}* [um livro$_{\text{GOODS}}$] [para estudar pra prova$_{\text{IMPOSED\_PURPOSE}}$]

(11a) [Eu$_{\text{OWNER}}$] *TENHO^{Target}* [dinheiro para pagar as contas$_{\text{POSSESSION}}$]

Note, first, that there is an asymmetry in the annotation, with the infinitival sentence headed by *para* being annotated as a separate FE in (9a) and (10a), and as part of the direct object in (11a). In fact, the annotation proposed for (11a) does not account properly for the constituent structure of this sentence. Since *para pagar as contas* could be easily moved to left of the sentence, it is not likely to be a part of the NP headed by *dinheiro*. The reason why (11a) is annotated as such is due to the fact that there is no non-core FE that could be assigned to the *para* sentence.

One possible solution would be creating such an FE, maybe named PURPOSE_OF_POSSESSION. However, there is an alternative analysis that represents a gain in generality: instead of proposing frame-specific non-core FEs to deal with instances of the same kind of infinitival sentence, one could annotate (9)–(11) also for the `dative_with_infinitive` cxn, thus assigning the ENABLED_SITUATION FE of the `Sufficiency` frame automatically to all instances. Such an analysis would capture the general fact that the money and the book in sentences (9)–(11) are the ITEMS whose `Sufficiency` enables the first person to pay for the bills and study for the test.

When more frame-bearing constructions are added to FN-Br 2.0 and analytical generality reaches domains other than purpose, the role of non-core FEs

may be deeply revised, i.e., instead of creating non-core FEs for each frame whose LUs are attested in sentences featuring these circumstantial elements, one could create more general constructions evoking circumstantial frames, and restrict the lexical annotation to those FEs that are truly defined by the valence properties of the lexical item.

Importantly, no claim is being made towards using some kind of argument-adjunct distinction to draw a line between FEs that should be created and those that shouldn't. The claim being made here still respects the basic BCG assumption that the minimal valence of lexical items includes both core and peripheral valents. However, as Fillmore (2013, p. 132) points out:

> In addition to 'core' and 'periphery' (…), there are also constructions that introduce into clauses various 'extrathematic' subordinate structures that are not directly a part of the semantic frame of the syntactic head of the clause.

Such a distinction is also valid for Berkeley FrameNet (Ruppenhoffer et al., 2016), although it seems not to be consistently applied for the specific examples annotated in (9a)–(11a): Berkeley FrameNet classifies the PURPOSE_OF_THEME FE in the `Receiving` frame as peripheral, while the IMPOSED_PURPOSE FE in the `Commerce_buy` frame is defined as extra-thematic.

The integrated approached in favor of which we advocate, would also, thus, enhance the consistency of the FN-Br Lexicon, since a richer constructicon has the potential for providing an additional more reliable criterion for the differentiation between peripheral and extra-thematic FEs, removing the need for creating the latter in several different frames.

4.1.2   *Constructions may inherit from other constructions*

Besides being related to the frames they evoke, constructions are also related to other constructions in a network (or lattice) of inheritance relations. Inheritance is usually approached in two different ways in Construction Grammar. As Kay (2005) points out, cognitively inspired approaches, such as Goldberg's (1995, 2006), adopt multiple inheritance link types, such as Polysemy, Instance, Subpart and Metaphor (Goldberg, 1995, p. 75–81), while monotonic approaches, such as Kay & Fillmore's (1999), adopt only one type of inheritance link.

In the first approach – the so-called normal mode of inheritance (Goldberg, 1995, p. 73–74) –, links between constructions are defined as cognitive objects and are meant to capture how the mother construction motivates the daughter. No strict constraints on how much of the information in the mother construction is transferred to the daughter are posited, as long as the daughter construction does not conflict with the mother. Hence, subregularities, exceptions and partial generalizations are allowed.

In the latter approach – the complete mode of inheritance –, links between constructions are meant to account for the generalities observable across the network of constructions. All the information in the mother construction must be equally or more specifically present in the daughter (Kay & Fillmore, 1999, p. 7).

Beyond the differences in regards to the kinds of constraints that must be satisfied when positing a link, the adoption of either the normal or the complete mode of inheritance may lead to proposing a completely different set of constructions for a language. Kay (2005) demonstrates this fact, by proposing an alternative monotonic approach to Goldberg's (1995) account for the ditransitive cxn. In her analysis, Goldberg (1995, p. 75–77) proposes five Polysemy links connecting the central sense of the ditransitive cxn – 'X causes Y to receive Z' – to five extensions of this sense, which include X enabling Y to receive Z, X causing Y not to receive Z or X intending to cause Y to receive Z. Kay (2005) argues that, instead of positing the existence of Polysemy links connecting six different ditransitive constructions, the grammar of English should feature one `abstract_recipient` cxn, which adds a recipient argument to the minimal valence of a predicator, and three maximal subconstructions: the `direct_recipient` cxn, the `intended_recipient` cxn and the `modal_recipient` cxn. By exclusively using complete inheritance and the same kind of unification processes between frames used for the `dative_with_infinitive` construction in Figures 1 and 2, Kay (2005) shows the variation in the senses of a construction can be accounted for by the interaction of frames and their elements inside the construction.

Building on that and similar analyses (Kay & Fillmore, 1999; Fillmore, 1999), our first attempt to model construction-to-construction relations in the FN-Br Constructicon adopts the complete mode of inheritance. Hence, when a daughter construction inherits from its mother in the resource, all the CEs in the mother must map to an equal or more specific CE in the daughter. Multiple inheritance is allowed, meaning that a construction may inherit structure from more than one mother.

The approach we adopted to inheritance is very similar to the one used in ECG with the *subcase of* keyword (Dodge, 2010, p. 51). In ECG, this keyword is used for modeling both construction and schema inheritance relations. In FN-Br 2.0, inheritance also holds between constructions and between frames, with almost the same kinds of constraints. Let us now return to the `dative_with_infinitive` cxn and see how its network of inheritances is formed.

As pointed out by Laviola (2015), the combination of a head_np showing an augmented valence that requires a para_sinf may yield two different readings in Brazilian Portuguese: that of enablement/`Sufficiency` already discussed in Section 2, and one of obligation, as exemplified in (12).

(12) Eu tenho   provas pra corrigir.
    I   have.PRES.1SG tests   to   correct.INF
    I have some tests to grade.

Laviola (2015) has also shown that, for the obligation reading to be available, the HEAD_NP, *provas* in the case of (12), must instantiate what would be a core FE of the frame evoked by the infinitival verb. Moreover, such a core FE is preferably an undergoer-like function. In (12), the frame evoked by *corrigir* is `Assessing`, and *provas* would be assigned the FE for the PHENOMENON being assessed by the first person ASSESSOR.

Therefore, two `dative_with_infinitive` constructions seem to exist in Brazilian Portuguese, since both the meaning and the unification constraints of the constructions licensing (5) and (12), the enablement and the obligation readings, respectively, are different. Nevertheless, both constructions are a more specific type of an `infinitival_relative` cxn whose infinitival sentence is headed by *para*. Since Brazilian Portuguese also admits other prepositions acting as heads of `infinitival_relative` constructions, such as *de* 'of' for example, the `para_infinitival_relative` cxn would be a more specific type of the former, which in turn, is a more specific type of a general `relative` cxn.

Because the presence of *para* brings to the `para_infinitival_relative` cxn a subjacent purpose reading that is not present in the `de_infinitival_relative` cxn, the former also inherits from the `purpose_adjunctive_clause` cxn. The inheritance network just described is shown in Figure 11, which depicts the working area of the FN-Br 2.0 web tool where such relations are modeled.

Note that there are two inheritance paths leading to the `dative_with_infinitive` constructions: the one in the right shows that both are a specific type of
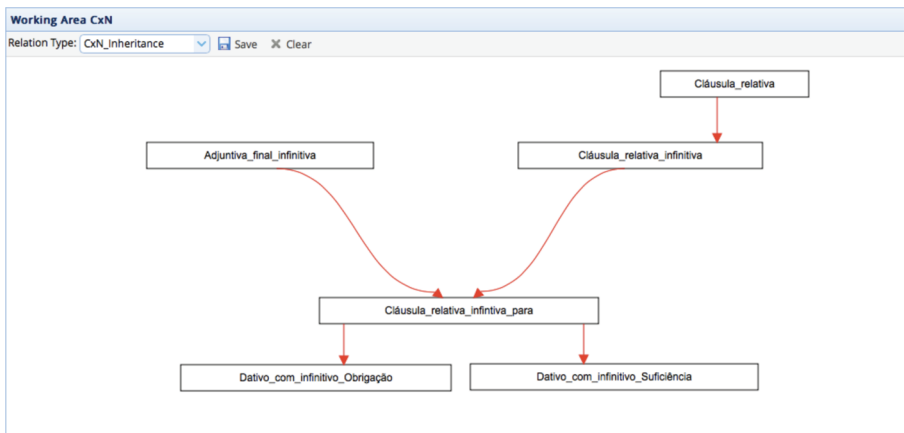
Figure 11. The inheritance network of the `dative_with_infinitive` constructions

relative_clause (cláusula_relativa), while the one in the left shows that they also incorporate structure from a purpose_infinitival_adjunctive clause (adjuntiva_final_infinitiva). Both inheritance paths converge to a para_infinitival_relative_clause (cláusula_relativa_final_infinitiva), an abstract construction whose structure is shared by the two dative_with_infinitive constructions. The relative_clause inheritance path has one additional level, the one featuring an abstract infinitival_relative_clause (cláusula_relativa_infinitiva), whose structure is shared by both the *para*-headed and the *de*-headed infinitival relatives.

Inheritance relations modeled by FN-Br 2.0 also take into consideration the internal constituency of the constructions involved, meaning that the daughter signs of the constructions are mapped to each other. Such a mapping may occur in either a one-to-one or a many-to-one fashion.

The first case, observable in the relative_clause inheritance chain, is straightforward: because the dative_with_infinitive constructions are ultimately a type of relative clause, the HEAD_NP CEs in the daughter constructions of this chain are mapped to the HEAD_NP CEs in the mothers all the way up. In turn, the PARA_SINF CEs in the bottom two levels of the chain are mapped to a SINF CE in the infinitival_relative cxn – in which the head preposition is left unspecified – and to a SREL CE in the relative_clause cxn, in which the kind of VP in the relative sentence is left unspecified.

The second case requires deeper explanation. The reason for positing an inheritance link between an adjunctive clause and (ultimately) a relative clause may seem obtuse. However, Torrent (2009, 2015) has shown that the PARA_SINF CE in the dative_with_infinitive cxn inherits the structure and constraints of the purpose_adjunctive cxn, specially the fact that, unlike most typical relative clauses, Para-infinitival ones can be fronted. Hence, an inheritance link connecting these constructions is posited. Since the purpose_adjunctive cxn does not share the same internal structure of a relative_clause, this inheritance link is of a many-to-one kind: the two CEs of the mother construction – the preposition *para* and the infinitival sentence – are both mapped to the PARA_SINF CE, as shown in Figure 12.
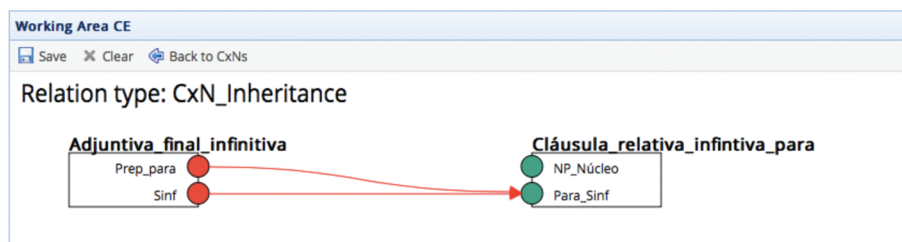
Figure 12. The CE-to-CE mapping sustaining the inheritance relation between the purpose_adjunctive cxn and the para_infinitival_relative cxn in FN-Br 2.0

4.2   The inceptive aspect construction

The analyses presented in the previous section demonstrate the advances already achieved by FN-Br 2.0. However, other important aspects of constructions and their relations with frames are still to be modeled. In this section, we present another construction in Brazilian Portuguese, the `inceptive_aspect` cxn (Sigiliano, 2011).

This construction has two daughter signs, one aspectual marker in the finite form followed by an infinitival VP, which may or may not be headed by a preposition. The first sign may be instantiated by several different verbal stems, some of which are canonical inceptive aspectual markers, such as *começar* and *iniciar* 'start', while others are not typically aspectual, such as *danar* 'harm', *desatar* 'untie', *entrar* 'enter' and *romper* 'break'. Sentences (13)–(17) are constructs licensed by this construction.

(13) Maria começou   a estudar   logo cedo.
      Maria start.PAST.3SG to study.INF soon early
      Maria started studying early in the morning.

(14) Maria danou   a reclamar   do   irmão.
      Maria harm.PAST.3SG to complain.INF of the brother
      Maria started to (iteratively) complain about her brother.

(15) Maria desatou   a falar   mal do   emprego.
      Maria untie.PAST.3SG to talk.INF badly of the job
      Maria started to (iteratively) complain about her job.

(16) Maria rompeu   a chorar.
      Maria break.PAST.3SG to cry.INF
      Maria burst into tears.

(17) Entrou   a chover.
      enter.PAST.3SG to rain.INF
      It started raining.

As it may be seen from the examples, the different aspectual markers yield slightly different inceptive readings.

As for the infinitival verb, Sigiliano (2011, p. 131) points out that 13 different semantic types of verbs can occur in the constructs licensed by this construction, as shown in Table 1.

Sigiliano (2011) shows that the combination of the aspectual marker with the V_INF is not free of constraints, and, also, that metaphors are key in the definition of such restrictions. We will look into each of these two aspects more in detail in the next two sections.

Table 1. Semantic types of Vᴵɴꜰ adapted from Sigiliano (2011)

Semantic type of Vᴵɴꜰ	Examples
Requesting	*ordenar* 'give orders', *persuadir* 'persuade'
Desiring	*querer* 'want', *desejar* 'desire'
Perception	*ver* 'see', *ouvir* 'listen'
Cognitive process	*saber* 'know', *entender* 'understand'
Believing	*achar* 'think', *acreditar* 'believe'
Communication	*dizer* 'say', *falar* 'talk'
Motion	*ir* 'go', *vir* 'come'
Action	*fazer* 'do', *pegar* 'take'
State	*ser* 'be', *permanecer* 'remain'
Natural phenomenon	*chover* 'rain', *nevar* 'snow'
Change of state	*secar* 'become dry', *passar* 'pass'
Feeling	*amar* 'love', *odiar* 'hate'
Emotion expression	*chorar* 'cry', *soluçar* 'hiccup'

4.2.1 *Constructions specify slot-filling constraints*

Sigiliano (2011, p. 132), in a corpus-based diachronic study, demonstrates that, while some of the non-canonical aspectual markers take almost any type of Vᴵɴꜰ, such as *entrar* 'enter', attested with 12 of the 13 semantic types listed in Table 1, others, such as *romper* 'break', take only four types. Also, for the case of *romper*, 68% of the licensed constructs involve an infinitival verb of emotion expression, 18% a verb of action, 7,8% one of motion and 6,2% a verb of communication.

In addition, the combination of *romper* with other semantic types of Vᴵɴꜰ sounds rather infelicitous in Brazilian Portuguese, as it is shown, for instance, in (18a)–(18d).

(18) a. *Maria rompeu   a persuadir   os colegas.
       Maria break.ᴘᴀsᴛ.3sɢ to persuade.ɪɴꜰ the colleagues
       **Maria burst into persuading the colleagues.*

   b. *Maria rompeu   a querer   um carro.
       Maria break.ᴘᴀsᴛ.3sɢ to want.ɪɴꜰ a   car
       **Maria burst into wanting a car.*

   c. *Maria rompeu   a ouvir   os colegas.
       Maria break.ᴘᴀsᴛ.3sɢ to listen.ɪɴꜰ the colleagues
       **Maria burst into listening to her colleagues.*

   d. *Maria rompeu   a saber   a verdade.
       Maria break.ᴘᴀsᴛ.3sɢ to know.ɪɴꜰ the truth
       **Maria burst into knowing the truth.*

Hence, in some cases, the attested constructs indicate a strong correlation between the aspectual marker and the semantic type of the Vinf, indicating the existence of constraints regarding the filling of the auxiliary and Vinf slots of the construction. In terms of modeling, this and other problems must be addressed.

First, the `inceptive_aspect` cxn may evoke either the `Activity_start` or `Process_start` frames, depending on the valence affordances of the Vinf: when the Vinf valence requirements involve an agentive external argument, the frame evoked is `Activity_start`; when it doesn't, the frame evoked is `Process_start`, provided that the difference between these two frames is precisely the lack of an Agent FE in the latter. Such a distinction in the `X_start` frames in Berkeley FrameNet is related to the fact that, in the lattice of frames, `Activity_start` inherits from `Process_start`, adding the Agent FE and turning the Event FE in `Process_start` into a more specific Activity FE.

This first problem could be solved, in principle, by following the same solution adopted for the two readings of the `dative_with_infinitive`, that is, by creating two inceptive constructions in the Constructicon. An alternative solution would be to create only one construction and to model the constraints according to which the existence or absence of an agentive FE in the frame evoked by the vinf would define whether the frame evoked by the construction is `Process_start` or `Activity_start`.

Despite the fact that the first solution is already feasible in FN-Br 2.0, we claim that the infrastructure for the kind of constraint modeling proposed in the second solution needs to be included in FN-Br 2.0. If dealing with this difference in the frames evoked by the construction is not a good enough reason for that, it would still be needed to allow the model to account for the collocational restrictions that characterize the second problem to be addressed in regards to the `inceptive_aspect` cxn: how to tell the system that sentences like (16) are possible, while sentences like (18a)–(18d) are not?

Again, the frame evoked by the vinf plays a key role in the definition of this constraint: almost 75% of the attested examples presented by Sigiliano (2011) featuring *romper* as the aspectual marker have a vinf evoking either the `Communication` or the `Communication_noise` frames, which inherits from the former. Such a high percentage is not coincidental: according to Sigiliano (2011), the reason behind such a correlation is grounded on the diachronic principle of persistence (Hopper, 1991), because some residual semantics of *romper,* still present in its aspectual use, poses metaphorically grounded constraints to the types of infinitival verbs that may co-occur with it.

Such a claim leads us to approach the second challenge to FN-Br 2.0, which we will discuss in the next section.

4.2.2 *Constructions may instantiate metaphors*

In order to explain why the `inceptive_aspect` cxn featuring *romper* as the aspectual marker licenses mostly constructs with verbs of communication and emotion expression, while also blocking most of other semantic types, Sigiliano (2011) adopts Talmy's (2001) Force Dynamics. According to her analysis, aspectual markers such as *romper* 'break' and *desatar* 'untie' preserve the residual semantics of their use as main verbs, which can be represented by the force-dynamics pattern presented in Figure 13.
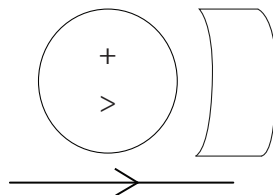
Figure 13. Force-dynamics pattern (Talmy, 2001 apud Sigiliano, 2011, p. 137)

In this pattern, the Agonist – represented by the circle – has an intrinsic force tendency to go towards (>) the Antagonist – represented by the concave figure. Since the Agonist is the strongest entity in this pattern (+), the resultant of the force interaction – represented by the arrow in the bottom of the diagram – is motion towards – and actually through – the Antagonist.

Hence, while, as a main verb, *romper* indicates that a force – the Agonist – moves through a barrier – the Antagonist – by destroying it, in sentences like (19); as an aspectual marker, *romper* maintains an image-schematic version of this meaning, indicating that an activity metaphorically goes through a barrier and begins. The ACTIONS ARE SELF-PROPELLED MOVEMENTS metaphor (Lakoff, 1979, p. 220) provides the basis for this semantic extension.

(19) A força   da   água   rompeu   a barragem.
       The force of the water break.PAST.3SG the dam
       The force of the water destroyed the dam.

Sigiliano (2011) moves on to explicate that the reason why verbs of emotion expression tend to occur in instances of the `inceptive_aspect` cxn featuring *romper* is due to the fact that aspectual meaning provided by the combination of this marker with the construction is that of an abrupt start. According to the author, another metaphor – EMOTIONS ARE LIQUIDS IN A CONTAINER – is also brought into play for those cases: since emotions are conceived as liquids, when the container is broken these emotions are no longer contained and tend to spread (Sigiliano, 2011, p. 139).

In a nutshell, in order to model the slot-filling constraints for the `inceptive_ aspect` cxn it would be necessary not only to provide a means to automatically

evaluate whether the frame evoked by the VINF features or not an agentive FE – and, hence, define whether the construction evokes `Activity_start` or `Process_start` –, but, also, to model a metaphorically based constraint that relates the frame that would be evoked by the aspectual marker – if it was used as a main verb – and the one evoked by the VINF.

FN-Br 2.0 is unable to approach these issues for now, specially because it does not contain a metaphor repository. In the future, when metaphor repositories are made available, we can assess the possibility of incorporating them to our database.

5.   Conclusions

In this chapter we showed that the continuity between grammar and the lexicon, a core principle of Construction Grammar, can be approached computationally by integrating framenet-based lexicons and constructicons in one single relational database. We demonstrated the feasibility of this task by providing exemplar analyses of the `dative_with_infinitive` constructions involving (1) the unification of the information associated with the various annotation layers, (2) the evoking relation between a construction and a frame, and (3) the inheritance relation between constructions. By showing an analysis of the `inceptive_aspect` cxn, we also pointed to new challenges that must be faced in the expansion of the analytical capacity of FN-Br 2.0, specially the modeling of frame-based and metaphor-based constraints.

Acknowledgments

References

Baker, C. F., Fillmore, C. J., & Cronin, B. (2003). The structure of the FrameNet database. *International Journal of Lexicography*, 16(3), 281–296. doi: 10.1093/ijl/16.3.281

Bergen, B. K. & Chang, N. (2005). Embodied Construction Grammar in Simulation-Based Language Understanding. In J. -O. Östman & M. Fried (Eds.), *Construction Grammars: Cognitive Grounding and Theoretical Extensions* (pp. 147–190). Amsterdam: John Benjamins. doi: 10.1075/cal.3.08ber

Bergen, B. K. & Chang, N. (2013). Embodied Construction Grammar. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 168–190). New York: Oxford University Press.

Bryant, J. E. (2008). *Best-fit Constructional Analysis*. Ph.D. Dissertation in Computer Sciences. University of California Berkeley. Berkeley.

Costa, A. D., & Torrent, T. T. (2017). A Modelagem Computacional do Domínio dos Esportes na FrameNet Brasil [The Computational Modeling of the Sports Domain in FrameNet Brasil)]. In *Proceedings of the 11th Brazilian Symposium in Information and Human Language Technology* (pp. 201–208). Uberlândia: Universidade Federal de Uberlândia.

Dodge, E. K. (2010). *Constructional and Conceptual Composition*. Ph.D. Dissertation in Linguistics. University of California Berkeley. Berkeley.

Elmasri, R. & Navathe, S. (2010). *Fundamentals of Database systems*. Boston: Addison-Wesley.

Fillmore, C. J. (1968). The Case for Case. In E. Bach & R. Harms (Eds.), *Universals in Linguistic Theory* (pp. 1–89). New York: Holt, Rinehart and Winston.

Fillmore, C. J. (1977). The Case for Case Reopened. In P. Cole & J. Sadock (Eds.), *Syntax and Semantics 8: Grammatical Relations* (pp. 59–82). New York: Academic Press.

Fillmore, C. J. (1982). Frame Semantics. In The Linguistic Society of Korea (Ed.), *Linguistics in the Morning Calm* (pp. 111–137). Seoul: Hanshin Publishing Company.

Fillmore, C. J. (1999). Inversion and Constructional Inheritance. In G. Webelhuth, J. P. Koenig & A. Kathol (Eds.), *Lexical and Constructional Aspects of Linguistic Explanation* (pp. 113–128). Stanford: CSLI.

Fillmore, C. J. (2007). Valency Issues in FrameNet. In T. Herbst & K. Götz-Votteler (Eds.), *Valency: Theoretical, Descriptive and Cognitive Issues* (pp. 129–160). Berlin and New York: Mouton de Gruyter.

Fillmore, C. J. (2013). Berkeley Construction Grammar. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 111–132). New York: Oxford University Press.

Fillmore, C. J. & Atkins, B. T. (1992). Toward a Frame-Based Lexicon: the Semantics of RISK and its Neighbors. In A. Lehrer & E. F. Kittay (Eds.), *Frames, Fields and Contrasts* (pp. 75–102). New York: Routledge.

Fillmore, C. J., Kay, P. & O'Connor, M. C. (1988). Regularity and Idiomaticity in Grammatical Constructions: The Case of Let al.ne. *Language* 64(3), 501–538. doi:10.2307/414531

Fillmore, C. J., Lee-Goldman, R. & Rhomieux, R. (2012). The FrameNet Constructicon. In H. C. Boas & I. A. Sag (Eds.), *Sign-Based Construction Grammar* (pp. 309–372). Stanford: CSLI Publications.

Fillmore, C. J., Petruck, M. R. L., Ruppenhofer, J. & Wright, A. (2003). FrameNet in Action: The Case of Attaching. *International Journal of Lexicography* 16(3), 297–332. doi:10.1093/ijl/16.3.297

Goldberg, A. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: The University of Chicago Press.

Goldberg, A. (2006). *Constructions at Work: The Nature of Generalization in Language*. New York: Oxford University Press.

Hopper, P. (1991). On some principles on grammaticalization. In E. Traugott & B. Heine (Eds.), *Approaches to Grammaticalization*, vol. I (pp. 17–36). Amsterdam: John Benjamins. doi:10.1075/tsl.19.1.04hop

Kay, P. (2005). Argument Structure Constructions and The Argument-Adjunct Distinction. In M. Fried & H. Boas (Eds.), *Grammatical Constructions: Back to the Roots* (pp. 71–98). Amsterdam: Benjamins. doi:10.1075/cal.4.05kay

Kay, P. (2013). The Limits of (Construction) Grammar. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 32–48). New York: Oxford University Press.

Kay, P. & Fillmore, C. J. (1999). Grammatical Constructions and Linguistic Generalizations: The What's X Doing Y? Construction. *Language* 75, 1–33. doi:10.2307/417472

Lage, L. M. (2013). *Frames e construções: A implementação do constructicon na FrameNet Brasil*. [Frames and constructions: The implementation of the constructicon in FrameNet Brasil]. M.A. Thesis in Linguistics. Universidade Federal de Juiz de Fora, Juiz de Fora.

Lakoff, G. (1979). The Contemporary Theory of Metaphor. In A. Ortony (Ed.), *Metaphor and Thought* (pp. 202–251). New York: Cambridge University Press.

Langacker, R. W. (2008). *Cognitive Grammar: a basic introduction*. New York: Oxford University Press. doi: 10.1093/acprof:oso/9780195331967.001.0001

Laviola, A. B. (2015). Construções em Contraste: uma Análise Comparativa Português-Inglês no Tangente à Implementação de Constructicons. [Constructions in Contrast: a Portuguese-English Comparative Analysis Regarding the Implementation of Constructicons]. M.A. Thesis in Linguistics. Universidade Federal de Juiz de Fora. Juiz de Fora.

Miranda, N. S. & Salomão, M. M. M. (2008). *Construções do Português do Brasil*. Belo Horizonte: Editora da UFMG.

Paiva, V. M. R. L., & Torrent, T. T. (2017). Uma Proposta Metodológica para a Categorização Automatizada de Atrações Turísticas a partir de Comentários de Usuários em Plataformas Online [A Methodological Proposition for the Automatic Categorization of Touristic Attractions from User Comments in Online Platforms]. In *Proceedings of the 11th Brazilian Symposium in Information and Human Language Technology* (pp. 232–239). Uberlândia: Universidade Federal de Uberlândia.

Ruppenhoffer, J., Ellsworth, M., Petruck, M. R. L., Johnson, C. R. & Scheffczyk, J. (2016). *FrameNet II: Extended Theory and Practice*. Berkeley: ICSI, available at: http://framenet. icsi.berkeley.edu.

Sigiliano, N. S. (2011). Persistência e Extensão na Construção Inceptiva em [V1fin (PREP) V2inf] do Português [Persistence and Extension in the [V1fin (PREP) V2 inf] Inceptive Construction in Portuguese]. *Letras & Letras*, 27(1).

Talmy, L. (2001). *Toward a Cognitive Semantics*, vol. I. Cambridge: The MIT Press.

Torrent, T. T. (2009). *A Rede de Construções em Para (SN) Infinitivo: Uma Abordagem Centrada no Uso para As Relações de Herança e Mudança Construcionais*. [The Para (NP) Infinitive Construction Network: a Usage-Based Approach to Construction Inheritance and Change]. Ph. D. Dissertation in Linguistics. Universidade Federal do Rio de Janeiro. Rio de Janeiro.

Torrent, T. T. (2015). On The Relation Between Inheritance and Change: The Constructional Convergence and The Construction Network Reconfiguration Hypotheses. In J. Barðdal, E. Smirnova, L. Sommerer & S. Gildea (Eds.), *Diachronic Construction Grammar* (pp. 175–214). Amsterdam: John Benjamins. doi: 10.1075/cal.18.06tor

Torrent, T. T. & Ellsworth, M. J. (2013). Behind the Labels: Criteria for Defining Analytical Categories in FrameNet Brasil. *Veredas* 17(1), 44–65.

Torrent, T. T., Lage, L. M., Sampaio, T. F., Tavares, T. S., & Matos, E. E. S. (2014). Revisiting border conflicts between FrameNet and Construction Grammar: Annotation policies for the Brazilian Portuguese Constructicon. *Constructions and Frames*, 6(1), 34–51. doi: 10.1075/cf.6.1.03tor

Torrent, T. T., Salomão, M. M. M., Matos, E. E. S., Gamonal, M. A., Gonçalves, J., Souza, B. C., Gomes, D. S. & Peron-Correa, S. R. (2014). Multilingual Lexicographic Annotation for Domain-Specific Electronic Dictionaries. *Constructions and Frames* 6(1), 73–91. doi: 10.1075/cf.6.1.05tor

Torrent, T. T., Salomão, M. M. M., Campos, F. A., Braga, R. M., Matos, E. E. S., Gamonal, M. A., Gonçalves, J., Souza, B. C., Gomes, D. S. & Peron-Correa, S. R. (2014). Copa 2014 FrameNet Brasil. In Proceedings of COLING 2014, the 25th International Conference on Computational Linguistics: System Demonstrations (pp. 10–14). Dublin: Dublin City University.

Relations between frames and constructions

A proposal from the Japanese FrameNet constructicon

Kyoko Ohara

This chapter discusses relations between frames and constructions, based on the constructicon-building project within the Japanese FrameNet (JFN) project. The aims are: to clarify distinctions between a framenet lexicon and a constructicon; and to contribute to the on-going discussion on whether all constructions are "meaning-bearing." I will argue that a framenet analysis involves annotating frame-based syntactic/semantic structures of words (simple words and multiwords), while a constructicon annotation pertains to describing the internal and external syntax/semantics of linguistic objects that have complex structures. While maintaining that all constructions are meaning-bearing, I will point out that meaning structures of some constructions may not involve frames and propose a frame-based classification of constructions. Finally, I will suggest that a constructicon annotation needs both semantic frames and interactional frames.

Keywords: FrameNet, constructicon, Frame Semantics, Construction Grammar, syntax-lexicon continuum, semantic frame, interactional frame, Japanese, annotation

1. Introduction

This chapter discusses relations between frames and grammatical constructions,[1] based on the constructicon-building project within the Japanese FrameNet (JFN) project (Ohara, 2013), the umbrella for building both the lexical resource and the constructicon.[2] In addition to the existing lexical-resource-buidling projects of framenets for a range of languages, several constructicon-building projects have been underway for languages other than English, such as Japanese, Swedish, and Brazilian

---

1. Hereafter, I will use "constructions" to mean "grammatical constructions" in the sense of Construction Grammar (e.g. Fillmore 2013, inter alia).

2. http://jfn.st.hc.keio.ac.jp

Portuguese. These include the English FrameNet[3] Constructicon (Lee-Goldman & Petruck, this volume), the Japanese FrameNet Constructicon (Ohara, 2013, 2014), the Swedish Constructicon (Sköldberg et al. 2013; Lyngfelt, Bäckström et al., this volume), the FrameNet Brasil Constructicon (Torrent and Lage, 2014; Torrent et al., this volume), the Russian Constructicon (Janda et al., this volume), and the German Constructicon (Boas & Ziem, this volume).

As pointed out in Lyngfelt (this volume), two meanings of "constructicon" exist in the literature. One meaning pertains to a theoretical concept, that is, a structured network of grammatical constructions (Fillmore, 1988; Jurafsky, 1991). The other involves an actual instantiation of construction descriptions (Fillmore, Lee-Goldman & Rhomieux, 2012). This chapter focuses on the latter sense of constructicon.

At the same time, both constructicons and framenets are practical implementations of the theories of Construction Grammar and Frame Semantics (cf. Boas, 2010), respectively. In the two theories, grammatical constructions are defined as form-meaning pairs. Furthermore, the two theories assume the syntax-lexicon continuum rather than the dictionary-and-grammar model, since dividing speakers' knowledge of vocabulary from that of grammar is impossible, as apparent from the existence of many productive idiomatic expressions at different levels and at varying degrees in a language (Fillmore, Kay, and O'Connor, 1988; Hilpert, 2014, pp. 3–8). In discussing relations between frames and constructions in this chapter, I will attempt to preserve the syntax-lexicon continuum in the JFN Lexicon and the JFN Constructicon in two ways: (1) avoiding an arbitrary distinction between words and linguistic objects that are not words; and (2) ensuring that the information in the JFN Lexicon and the JFN is Constructicon as parallel as possible (cf. Torrent et al., this volume).

The two-fold aims of this chapter are: (1) clarify distinctions between a framenet lexicon and a constructicon; and (2) contribute to the on-going discussion on whether or not all constructions should be seen as "meaning-bearing". Regarding the first aim, the need for a constructicon in addition to a FrameNet-style lexicon may not be obvious to everyone. Moreover, relations between a framenet annotation and a constructicon annotation have not been discussed much in the literature and variations seem to exist in understanding the relations between the two among the current projects as well.[4] In this chapter, I will distinguish between the two as follows: framenet annotation involve annotating frame-based syntactic

---

3.   FrameNet (FN) is used as the name of the lexical resource and also as the name of the project.

4.   Fillmore (2008, p. 59) calls the two types of annotation "FN lexicographic annotation" and "constructional annotation". In this chapter, I will use "framenet annotation" and "construction annotation" for reasons discussed in Section 3.

and semantic structures of words (simple words and multiwords); constructicon annotation requires describing the *internal* and *external syntax* and *internal* and *external semantics* of linguistic objects that have complex structures.

With respect to the second aim, there have been discussions on whether all constructions should be seen as "meaning-bearing" or not (Fillmore, Lee-Goldman, & Rhomieux, 2012, pp. 325–328; Goldberg, 2006, pp. 166–182; Hilpert, 2014, pp. 51–57). I maintain that all constructions are meaning-bearing and instead regard what the researchers have been calling "non meaning-bearing" constructions to be constructions that do *not* evoke frames (Section 4). I will then propose classification of constructions based on the notion of frames (see also Lyngfelt, Bäckström et al., this volume, for a discussion on frame-bearing and non frame-bearing constructions in Swedish in SweCcn, from a slightly different perspective[5]). Finally, I will suggest that for a constructicon annotation, we need interactional frames in addition to semantic frames.

The discussion in this chapter will be based on analyses of Japanese constructions, which are being carried out for the purpose of building a prototype of a constructicon as part of the JFN project. In building the JFN Constructicon, we have been concentrating on grammatical constructions that cannot be annotated in the JFN Lexicon (cf. Section 2). Since the databases and other resources of the overarching JFN project are compatible with those of the FN project, in which the lexicon and the constructicon are integrated parts of the same resource, the JFN lexical data and the JFN Constructicon data allow cross referencing. As the time of writing, however, the JFN Constructicon data has not been linked to the JFN Lexicon data yet.

Frame Semantics and framenets, i.e., its practical implementations, pertain to linguistically anchored frames. Frames refer to "any of the organized packages of knowledge, beliefs, and patterns of practice that shape and allow humans to make sense of their experiences" (Fillmore & Baker, 2010, p. 314). Frame elements (FEs) are the aspects and components of individual frames (Fillmore & Baker, 2010, p. 321). There are two very important notions in Frame Semantics and consequently in framenets: frame evocation and frame invocation. The former is defined as "a cognitive experience on the part of an interpreter that comes about by the interpreter's responding to language-specific associations connecting linguistic signs with particular frames" and in Frame Semantic terms a given linguistic sign evokes a linguistically anchored frame that contributes to interpreting the passage (Fillmore & Baker, 2010, p. 316). The latter notion of frame invocation is defined as "a cognitive act that the interpreter (possibly quite unconsciously) performs to

---

5.   However, their analysis of Swedish constructions and the analysis of Japanese constructions presented in this chapter share many insights and thus are essentially compatible.

make sense of some incoming information" and in this theory it is viewed that the interpreter invokes a cognitive frame that enables the experience to make sense (Fillmore & Baker, 2010, p. 316). Typically words are frame evoking elements (FEEs), or targets, but as we will see in Section 4 linguistic objects other than words may also evoke frames.

In Construction Grammar, grammatical constructions are pairings of a linguistic form with a meaning. Constructs are actual structures licensed by one or more constructions and construct elements (CEs) are components of constructions (Fillmore, Lee-Goldman & Rhomieux, 2012, p. 321). Construction evoking elements (CEEs) are lexically-limited material (if any) (Fillmore, Lee-Goldman & Rhomieux, 2012, p. 323).

The rest of the chapter is organized as follows. Sections 2 and 3 deal with distinctions between framenet lexicons and constructicons: Section 2 discusses why we need constructicons in addition to framenets, by focusing on the targets of annotation needed to produce the two kinds of resources; and Section 3 compares and contrasts information added in a framenet annotation and a constructicon annotation. Sections 4 and 5 have to do with the issue of whether all constructions are meaning-bearing: Section 4 discusses "non meaning-bearing" constructions, which have also been called constructions "without meanings" and/or "semantically null" constructions, from a Frame-Semantic point of view; and Section 5 proposes a five-way classification of constructions based on whether or not constructions evoke frames. Finally, Section 6 summarizes the discussions.

2.   The need for constructicons

Why do we need constructicons in addition to framenets? One might argue that since grammatical constructions in Construction Grammar are defined as form-meaning pairs, which include linguistic objects at word, phrase, and clause levels, we do not need a lexicon and a constructicon separately, especially if we want to maintain the syntax-lexicon continuum, which is one of the most important assumptions in Construction Grammar just like in the other approaches in Cognitive Linguistics. In this section, I will argue that a constructicon is needed in addition to a framenet lexicon, since there are linguistic objects that cannot be annotated within the framework of a framenet lexicon.

First of all, limitations on lexicographic annotations in FN and JFN have become apparent: the two projects originally had the purely lexicographic purposes of (1) characterizing the main distributional properties of verbs, nouns, and adjectives based on valences and (2) identifying the requirements that lexical units

(LUs), pairings of a lemma and a frame, might impose on their dependents, using the concept of frames (cf. Fillmore, Lee-Goldman & Rhomieux, 2012, p. 310–313; Ohara, 2013, p. 21–22). This means that the kinds of grammatical structures that purely lexicographic annotation of LUs is recording are more or less limited to relations of predication, modification, and complementation (Fillmore, Lee-Goldman & Rhomieux, 2012, p. 312). Furthermore, Fillmore, Lee-Goldman and Rhomieux (2012) point out that linguistic objects "that function *as units* while at the same time having a describable internal structure", in other words, linguistic objects with internal and external structures are difficult to describe in framenets (Fillmore, Lee-Goldman & Rhomieux, 2012, pp. 12–313). At the same time, purely grammatical patterns with no reference to any lexical items cannot be dealt with by framenet annotation. Consequently, there are many sentences whose semantic and syntactic organizations cannot be fully annotated in framenets. Therefore, in order to describe the meaning of various kinds of sentences, the FN project and later the JFN project began to engage in construction-annotating activities additionally.

Currently, the JFN Constructicon is being built as a prototype of a Japanese constructicon, for the purpose of finding out what would be needed in language resources that can be regarded as practical implementations of the theories of Frame Semantics and Construction Grammar (cf. Section 1). We are thus focusing on expressions that cannot be annotated in JFN.

Let us examine an example of a grammatical structure that cannot be recorded by framenet annotation. The sentence pattern is often called "internally headed relativization" (Ohara, in press). The following is taken from a voice mail.

(1) *[[kinoo <u>ringo</u> o okuttekudasatta] no] ga kyoo tukimasita*
yesterday apple ACC send-HON-PST NMLZ NOM today arrived
Literal translation: '[That (you) sent me apples yesterday] arrived today.'
Intended: '(You) sent me apples yesterday, and I received (them) today.'

Here, the main clause asserts the fact that you sent me apples yesterday. The subject of the main verb *tsukimashita* "arrived" is semantically construed as *ringo* "apple". However, it is inside the nominalized clause (i.e. the relative clause) with no syntactic identification. Hence the name internally headed relativization. This is a purely grammatical pattern and involves no lexical item. This sentence pattern is known for its narrative-advancing function and each of the two clauses has various semantic and pragmatic constraints. In particular, there are many internal and external semantic restrictions on the relative clause and on the main clause. That is, in the `internally_headed_relativization` construction, no single lexical item evokes a frame; rather, the entire sentence pattern functions as a unit, and the sentence pattern includes a describable internal structure. Thus, describing

the `internally_headed_relativization` construction requires a constructicon annotation, not a framenet annotation.

There are many other grammatical constructions in Japanese like the `internally_headed_relativization` construction with internal and external structures. Consequently, semantics and the JFN project is thus creating a prototype of the JFN Constructicon to record and annotate such constructions in Japanese.

3.   Framenet annotations and constructicon annotations

This section will clarify distinctions between framenet lexicons and constructicons by focusing on the kinds of annotation conducted to produce the two types of language resources. I will first bring up terminological issues and then discuss the information added in the two kinds of annotation.

So far I have been using "framenet annotation" and "constructicon annotation," instead of "lexicographic/frame annotation" and "construction(al) annotation" (see also Footnote 3). There are at least three reasons. First, there are variations among linguists and projects in what information to include in "frame annotation" and "construction(al) annotation." For example, by "constructional annotation" some focus on annotating constructions with frame names and FE labels, while others emphasize identifying constructions' constructs and CEs. (2) and (3) are examples of constructicon entries given in Boas (2010, p. 71) and Fillmore, Lee-Goldman and Rhomieux (2012, p. 39), respectively. In (2b) the sentence structure is annotated with semantic role labels such as Agent, Patient, and Recipient and so is the example sentence in (2d).

(2)   Constructicon entry in Boas (2010, p. 71)
       The `ditransitive` construction
       a.   Description: A volitional AGENT successfully transfers a PATIENT to a willing RECIPIENT, who receives the PATIENT.
       b. $[\text{NP}_1/\text{Subj}]_{\text{AGENT}}$ verb^{tgt} $[\text{NP}_2/\text{Obj}_1]_{\text{RECIPIENT}}$ $[\text{NP}_3/\text{OBJ}_2]_{\text{PATIENT}}$
       c.   List of LUs that evoke the `ditransitive` construction: v.Giving, signal. Communication, tell.Telling, v.Cooking_creation, …
       d.   Annotated example sentence for each LU that evokes the `ditransitive` construction: [*Miriam*]$_{\text{AGENT}}$ *passed* [*Joe*]$_{\text{RECIPIENT}}$ [*the salt*]$_{\text{PATIENT}}$

On the other hand, in (3) an interpretation of the construction is given but the example sentence is not annotated with FE names.

(3) Constructicon entry in Fillmore, Lee-Goldman and Rhomieux (2012, p. 339)

{[AUX-INITIAL:COND] [AUX] [SUBJ] [PRED] }

Name	`aux_initial:conditional`
M	Inverted finite clause
D1	Auxiliary verb, either *had, should,* or *were*
D2	NP, the subject of D1
D3	Predicate (verbal or otherwise), selected by D1, shares subject with D1
Interpretation	A conditional clause, which interpretation varying with the identity of D1.

{[AUX-INITIAL:COND] [AUX *had*] [SUBJ *you*] [PRED *arrived on time*] }

Here, following the convention of Sign-Based Construction Grammar, outer brackets '{ }' are used to enclose the entire structure of the construction; and inner brackets '[ ]' indicate the individual CEs. "M" stands for the whole (the Mother), that is, the external structure; and "D1" through "D3" for parts (Daughters), namely the internal structure (Fillmore, Lee-Goldman & Rhomieux 2012, p. 331). The numbering on the Ds indicates the order of the CEs.

Second, I chose to use the terms "framenet annotation" and "constructicon annotation" instead of "lexicographic annotation" and "construction(al) annotation" to avoid misunderstanding that framenets only annotate "words" and that constructicons exclusively involve annotating non-words and "constructions." In Construction Grammar, grammatical constructions are defined as form-meaning pairs and in this definition "forms" include not only phrasal and clausal patterns but also words and morphemes. We should therefore not distinguish words from non-words outright, because doing so might suggest that we are making arbitrary distinctions among "forms". Since the JFN Lexicon contains words including both simple words and multiwords and since the JFN Constructicon is for the other types of linguistic objects with both internal and external structures, "framenet" and "constructicon," rather than "lexicographic" and "constructional," seem preferable.

Third, I prefer the terms "framenet annotation" and "constructicon annotation" to "frame annotation" and "construction(al) annotation" since annotation in framenets is not only about revealing the "meaning" of words by assigning frame names and FE labels. It also pertains to describing words' "structural properties." That is, valence patterns, which are created as a result of annotation in framenets, include phrase types (PTs) and grammatical functions (GFs) in addition to FE labels. Conversely, annotating constructions does not have to be exclusively about describing "structures" of constructions. It also involves describing their meanings and interpretations.

For the reasons given above the JFN project uses the terms "framenet annotation" and "constructicon annotation." Regarding the format of a constructicon

entry, the JFN Constructicon chose to adopt the one used by FN Constructicon, used in Fillmore, Lee-Goldman and Rhomieux (2012) and exemplified in (3) above, for the following reasons. First, since we are interested in describing internal syntax/ semantics as well as external syntax/semantics of various grammatical constructions, the latter type of notation is better suited for the purpose, since the latter includes more detailed structural information. Second, as we will see in the next section, there are constructions for which we cannot assign semantic role or FE labels to their CEs, since the constructions do not evoke frames. Third, as mentioned in Section 1 the databases of the larger JFN project are compatible with those of the FN project and it is therefore practical to maintain the same format as that of the FN Constructicon.

Next, let us examine the kinds of information added in the two types of annotation. In JFN, framenet annotation processes consist of: (1) identifying the frame evoking element (FEE); (2) annotating constituents corresponding to the FE's of the frame; and (3) annotating the constituents with the PT and GF labels. Note that processes (2) and (3) above result in creating LUs' valence patterns, which consist of FE, PT, and GF labels.

Following Fillmore, Lee-Goldman and Rhomieux (2012), the JFN Constructicon regards constructicon annotation to consist of: (1) identifying the CEE if there is one (2) identifying constructs (i.e. actual structures licensed by the construction in question); and (3) identifying the construct-internal constituents as instantiating CEs (Fillmore, Lee-Goldman & Rhomieux, 2012, pp. 321–323). In other words, while framenet annotation describes syntax and semantics, that is, valence patterns of FEEs, constructicon annotation describes constructions' internal syntax and semantics by identifying CEEs, constructs, and Daughter CEs; and constructions' external syntax and semantics by identifying the Mother CE and the interpretation of the whole. Note that constructicon annotation by itself does not necessarily involve annotating constructions with frame names, as will be shown in Section 5.

Table 1 summarizes framenet annotation and constructicon annotation, focusing on their targets of annotation and the kinds of information added. We can see that the syntax-lexicon continuum is guaranteed with respect to both the target of annotation and the information added. First, the distinction between the two kinds of annotation is not about words *vs.* anything other than words but rather it has to do with whether the target of annotation has both internal and external structures or not. Second, in the two types of annotation the kinds of information added are kept as parallel as possible. That is, just like in framenet annotation, in constructicon annotation information pertaining to both forms and meanings are added. The only difference is that in constructicon annotation the meaning of a construction to be annotated may or may not involve frames, as will be discussed in the next section.

Table 1. Framenet annotations and constructicon annotations

	Framenet annotation	Constructicon annotation
Targets of Annotation	– Simple words; Multiwords (phrasal verbs, verb particles, words with selected prepositional complements, support constructions, combinations, transparent nouns (Fillmore 2008, p. 55–56))	– Linguistic objects with internal and external structures
Information Added	– Frame Name – Frame-Evoking Elements (FEEs) – Frame Elements (FEs) – Phrase Types (PTs) – Grammatical Functions (GFs)	– Construction Name – Construction-Evoking Elements (CEEs) – Constructs – Construct Elements (CEs) – Interpretation – Frame Name[*] – FEEs[*] – FEs[*]

[*] If the construction evokes a semantic frame (See Section 5)

4. Constructions "without meanings" and the use of frames to represent meaning structures of constructions

Let us now turn to the second goal of this chapter, namely, to discuss the issue of whether all constructions are meaning-bearing or not. Asking whether constructions have meanings or not may sound strange to many, since one of the basic tenets of Construction Grammar is the assumption that constructions are pairings of a form with a meaning and thus all constructions should have meanings. In the Construction Grammar community, however, there have indeed been discussions on whether there are constructions "without meanings". In this section, I will propose that instead of debating whether all constructions are meaning-bearing, it would be more appropriate to ask whether meanings of all constructions can be described by the notion of frames or not. That is, assuming that all constructions are meaning-bearing I argue that there are constructions whose meanings cannot be described by frames. I will first give a brief history of previous discussions in the community and then suggest that some constructions' meaning structures do not involve frames. In Frame Semantics, relations between frames and linguistic objects are accounted for by the concept of "frame evocation" and the concept is essential in framenets and constructicons as well. I will thus explicate the notion of "frame evocation", contrasting it with the related concept of "frame invocation", another important concept in Frame Semantics.

There have been discussions on whether or not *all* constructions should be seen as meaning-bearing (Fillmore, Lee-Goldman & Rhomieux 2012, pp. 325–328). In Construction Grammar, a grammatical construction is a conventionalized pairing between a specific formal pattern and the meaning it contributes to the expressions that contain it. However, some linguists including Fillmore (1999) argue for a "semantic-free syntax" for some linguistic expressions such as `aux_initial` in English (e.g. *May we come in?* with the meaning of question; *May you have a long and fruitful marriage* as a wish; *Had I known this, I would have gone* with the meaning of condition; *Did I do something stupid last night!* as an exclamation; see also (3) in Section 3). In the `aux_initial` construction, it is not possible to assign a clear meaning in the general form of the construction that is described by the more specific constructions. Goldberg, a strong proponent of Construction Grammar, opposes such a "semantic-free syntax" view and claims that the `aux_initial` construction indeed has a generalized meaning of its own, namely, non-assertiveness (Goldberg, 2006, pp. 166–182). Unlike Goldberg, however, the current FN Constructicon policy is to assume that semantically null constructions are legitimate and the `aux_initial` construction is categorized as one of such constructions (Fillmore, Lee-Goldman & Rhomieux, 2012, pp. 325–328) (see Section 5.1.2 below).

Since I view constructicons to be practical implementations of Construction Grammar, I maintain that all constructions have meanings. I argue, at the same time, that meaning structures of some constructions do not involve frames. More specifically, I contend that what Fillmore, Lee-Goldman and Rhomieux (2012) called constructions "without meaning" and "non meaning-bearing" constructions can be paraphrased as constructions that do not evoke frames.[6] In framenets it is assumed that simple words and multiwords evoke frames but linguistic forms other than words may also evoke frames (Ohara, 2013; Sköldberg et al., 2013). In fact, Fillmore and Baker (2010) points out, "Frame Semantics is the study of how linguistic forms evoke or activate frame knowledge, and how the frames thus activated can be integrated into an understanding of the passages that contain these forms" (Fillmore & Baker, 2010, p. 317). It is also important to note that there are constructions whose meanings cannot be described by frames, that is, there are constructions that do not evoke frames. Based on this idea, the next section presents a classification of constructions according to whether their meanings pertain to frames or not.

---

6.   See pp. 76–77 of Lyngfelt, Bäckström et al. (this volume) for a similar position.

5. A five-way frame-based classification of constructions

It is possible to classify constructions based on whether they evoke frames or not. There are three types of non frame-evoking constructions and two types of frame-evoking constructions as shown in Table 2. Let us examine each of these types of constructions.

Table 2. Five-way frame-based classification of constructions

Cxn type	Frame-evoking cxn?	Sub-section	Japanese examples
[1] Compositionally interpretable	NO	5.1.1	`modifier_head` cxn: (4)
[2] Its more elaborated cxns evoke frames	NO	5.1.2	`V_te_iru` cxn: (5′)
[3] With omission of repetitive position -specific constituents	NO	5.1.3	`gapping` cxn: (6)
[4] Evoking a semantic frame	Yes	5.2.1	`comparative_inequality` cxn: (7a)
[5] Evoking an interactional frame	Yes	5.2.2	`te_linkage` cxn: (8c); `suspended_clause` cxn: (9b)

5.1 Non frame-evoking constructions

The three types of non frame-evoking constructions are: [1] compositionally-interpretable cxns (Section 5.1.1); [2] cxns whose more elaborated constructions evoke frames (Section 5.1.2); and [3] cxns with omission of repetitive position-specific constituents (Section 5.1.3).

5.1.1 *Compositionally interpretable constructions*

The first type of non frame-evoking constructions pertains to syntactic patterns with specific formal features whose interpretation depends on combining information from their constituents in a completely regular way (Fillmore, Lee-Goldman, & Rhomieux, 2012, p. 326) (Type [1] in Table 2). The `head_complement` construction (involving the structure of complementation), the `modifier_head` construction (involving the structure of modification), and the `subject_predicate` construction (involving the structure of predication) in English and Japanese are examples of such constructions. Example (4) illustrates the `modifier_head` construction in Japanese. As mentioned in Section 3 with respect to (3), here and in the other examples of constructicon entries below, outer brackets '{ }' indicate the entire expression produced by the construction and inner brackets '[ ]' are for individual

CEs. M represents the external structure of the construction and the numbers on Ds show the order of the CEs. [7] In (4), the prenominal adjective *utukusii* 'beautiful' modifies the noun *hana* 'flower' and the interpretation of the whole phrase is dependent on combining information of the two pieces in a regular way.

(4) The `modifier_head` construction
{ [MODIFIER] [HEAD] }
M   NP
D1   Modifier AP
D2   Head NP
Interpretation   An NP, with D1 modifying D2
{ [MODIFIER *utukusii*] [HEAD *hana*] }
beautiful   flower
'(A) beautiful flower'

In other words, the meanings of the constructions belonging to Type [1] are derived compositionally and no frame is involved in interpreting them.

5.1.2 *Constructions whose more elaborated constructions evoke frames of their own*

The second type of non frame-evoking constructions involves constructions that determine syntactic patterns to which separate interpretations can be given under different variations (Type [2] in Table 2). Fillmore, Lee-Goldman and Rhomieux (2012) argue that the `aux_initial` construction (see also (3) in Section 3 and the discussion in Section 4) and the `filler_gap` construction are categorized as belonging to this type (Fillmore, Lee-Goldman & Rhomieux, 2012, pp. 326–327).[8] In the `filler_gap` construction, an argument of a verb, typically the direct object, appears to the left of the verb, i.e., in a place that differs from its canonical position in a simple declarative clause (Hilpert 2014, pp. 53–54). Constructs of this construction include: *What are you reading?* (*wh*-question), *How clever I am!* (exclamative), *The coat Dan had on yesterday was new* (relative clause)*, Normally people don't buy such books, but this one they'll read* (topicalization), and *The more he criticizes the author, the more they will read* (*the X-er, the Y-er*).

---

7.   Here and in the rest of the paper, unless otherwise noted, the order of Ds is rigid.

8.   The adjective_as_nominal construction in English is another example of Type [2] constructions, since there are at least three more specific constructions that inherit from it (e.g. *The party was losing its attraction for the young* (adjective_as_nominal.Human); *I took the shortest* (adjective_as_nominal.Anaphoric); and *It was only putting off the inevitable* (adjective_as_nominal. Abstract)) (Fillmore, Lee-Goldman & Rhomieux 2012, pp. 357–360).

That is, the `filler_gap` construction is inherited by more specific constructions, namely, the *wh*-question, exclamative, `relative_clause`, `topicalization` and *the_X-er,_the_Y-er* constructions, and each of these more specific constructions has a separate meaning. The `filler_gap` construction, however, does not have a generalized meaning and thus does not evoke a frame.

The V_*te_iru* construction in Japanese is another example of Type [2] constructions. The auxiliary *te iru* functions as an aspectual marker and attaches to the stem of a verb. Depending on the aktionsart of the preceding verb, separate aspectual interpretations are given, as shown in (5a) through (5c). In (5a), *te iru* attaches to a state verb *niru* 'resemble' and the whole sentence is interpreted as expressing a state; in (5b), *te iru* attaches to an activity verb *hasiru* 'run' and the sentence expresses an activity; and in (5c), *te iru* attaches to an achievement verb *oriru* 'fall' and the sentence describes a resultant state.

(5a) *haha   to   musume wa   yoku   ni   te iru*
      mother CONJ daughter TOP much resemble ASP
      '(The) mother and (the) daughter are much alike.'

(5b) *kodomo-tati ga   hasit te iru*
      child   PL NOM run   ASP
      '(The) Children are running.'

(5c) *koi   kiri ga   numa no   ue   ni   ori te iru*
      thick fog NOM mire GEN top LOC fall ASP
      '(A) thick fog has fallen over (the) mire.'

Rather than assigning a generalized meaning to the V_*te_iru* construction, it seems preferable to recognize three constructions that inherit from the V_*te_ir* construction, each having a specific aspectual meaning, namely, that of state, activity, or resultant state. As shown in (5′) below, the V_*te_iru* construction does not evoke a frame. Instead the V_*te_iru*:state (5′a), V_*te_iru*:activity (5′b), and V_*te_iru*:resultant_state (5′c) constructions, which inherit from the V_*te_iru* construction, have more specific meanings and each evokes a frame.

(5′) The V_*te_iru* construction
      { [SUBJECT] [VERB] [AUX *te iru*] }
      M   Clause
      D1 Subject NP
      D2 Verb
      D3 Auxiliary verb *te iru*

(5a′) The V_*te_iru*:state construction
{ [SUBJECT] [VERB- STATE] [AUX *te iru*] }

M	Clause
D1	Subject NP
D2	State verb
D3	Auxiliary verb *te iru*

Interpretation Evokes the State frame
{ [SUBJECT *haha to musume wa*] *yoku* [VERB- STATE *ni*] [AUX *te iru*] }
mother and daughter TOP   much resemble   ASP
'(The) mother and (the) daughter are much alike.'

(5b′) The V_*te_iru*:activity construction
{ [SUBJECT] [VERB- ACTIVITY] [AUX *te iru*] }

M	Clause
D1	Subject NP
D2	Activity verb
D3	Auxiliary verb *te iru*

Interpretation Evokes the Activity frame
{ [SUBJECT *kodomo-tati ga*] [VERB-ACTIVITY *hasit*] [AUX *te iru*] }
child PL NOM   run   ASP
'(The) Children are running.'

(5c′) The V_*te_iru*:resultant_state construction
{ [SUBJECT] [VERB- ACHIEVEMENT] [AUX *te iru*] }[9]

M	Clause
D1	Subject NP
D2	Resultant verb
D3	Auxiliary verb *te iru*

Interpretation Evokes the *Resultant_state frame[9]
{ [SUBJECT *koi kiri ga*] *numa no ue ni* [VERB- ACHIEVEMENT *ori*] [AUX *te iru*] }
thick fog NOM mire GEN top LOC   fall   ASP
'(A) thick fog has fallen over (the) mire.'

Note that since each of the V_*te_iru*:state, V_*te_iru*:activity, and V_*te_iru*:resultant_state constructions evokes a semantic frame, the relevant frame name is recorded in the "Interpretation" section in their respective JFN Constructicon entries (5a′) through (5c′). Additionally, the JFN constructicon provides a separate layer for semantic-frame annotation for each of the constructs. It includes the bracketing formula with the FEE and the FE labels, as shown in (5a″) through (5c″) below.

---

9.   Here and in the rest of the paper, the asterisk (*) before a frame name indicates that, at the time of writing, the frame has not been defined in the FrameNet Lexicon.

(5a″)　The State frame:
- The ENTITY persists in a stable situation called STATE.
- FEs: ENTITY, STATE

[_ENTITY *haha to musume wa*]　*yoku*　[_STATE = FEE *ni te iru*]
mother and daughter TOP　　much　resemble ASP
'(The) mother and (the) daughter are much alike.'

(5b″)　The Activity frame:
- The AGENT enters an ongoing state of the ACTIVITY, remains in this state for some DURATION of TIME and leaves this state.
- FEs: AGENT, ACTIVITY, DURATION, TIME

[_AGENT *kodomo-tati ga*]　[_ACTIVITY = FEE *hasit te iru*]
child PL NOM　　　　　run ASP
'(The) Children are running.'

(5c″)　The *Resultant_state frame:
- As a result of an EVENT happening to an ENTITY, A STATE begins and continues.
- FEs: ENTITY, STATE, EVENT

[_ENTITY *koi kiri ga*]　*numa no ue ni*　　[_STATE = FEE [_EVENT *ori*] [*te iru*]]
thick fog NOM　　mire GEN top LOC　fall ASP
'(A) thick fog has fallen over (the) mire.'

As we have just seen, in the case of Type [2] constructions, it is the high level of abstraction that makes them "meaningless." Thus, it may be misleading to say that the constructions belonging to Type [2] are "without meanings." They are semantically underspecified rather than being meaningless. They do not evoke frames but constructions that inherit from them have more specific meanings and evoke frames.

5.1.3　*Constructions that omit repetitive position-specific constituents*
The third type of constructions that do not evoke frames are constructions that allow the omission of position-specific constituents that would otherwise be repetitions (Type [3] in Table 2). Gapping (e.g. *John ate an apple and Mary a peach*), stripping (e.g. *Chris plays the guitar, but not the piano*), and shared_completion (e.g. *Robin is familiar with and fond of the dog*) in English belong to this type. In gapping, typically two phrasal constituents are juxtaposed and the second one is missing a verb that is present in the first. In stripping, a full sentence is stripped of everything except one constituent. In shared_completion, two phrases share a common ending (Fillmore, Lee-Goldman & Rhomieux, 2012, pp. 334–335).

`Gapping` exists in Japanese as well.

(6)   The `gapping` construction
{ [ITEM1] [ITEM2] [ITEM1] [ITEM2] [ITEM3] }

M	A coordinate structure, whose non-final conjuncts are missing some linguistic material present in the last conjunct
D1	Appears in each conjunct
D2	Appears in each conjunct
D3	A string in the last conjunct which contains the main predicate, and which is omitted from non-final conjuncts
Interpretation	Each non-final conjunct is missing some material that is present in the final conjunct, and each conjunct is interpreted and parsed as though that missing material were present.

{ [ITEM1 *ozii -san wa*]   [ITEM2 *yama e*] ,   [ITEM1 *obaa -san wa*]
old-man HON TOP   mountain GOAL old-woman HON TOP
[ITEM2 *kawa e*] [ITEM3 *ikimasita*] }
river GOAL went

'(The) old man went to the mountain, (the) old woman to the river.'

Type [3] constructions may therefore be called elliptical constructions and even though they do not have meanings of their own, sentences licensed by the constructions carry meanings that can be calculated by processing the meanings of the component words (Hilpert, 2014, p. 55). Due to their common property of being elliptical, I agree with Fillmore, Lee-Goldman and Rhomieux (2012) and Hilpert (2014) in regarding this type of constructions as distinct from Type [1] and Type [2] constructions. As Hilpert (2014) points out, these syntactic patterns do not seem to have idiosyncratic constraints or collocational preferences, unlike other types of constructions (Hilpert, 2014, p. 56–57).

To summarize the discussion on non frame-bearing constructions, in the cases of Type [1] and Type [3] constructions it can be said that their meanings are somehow derived from the meanings of the words that make up the constructions. As for Type [2] constructions they themselves do not evoke frames but the constructions that inherit from them do evoke frames.

5.2   Frame-evoking constructions

Next, let us examine frame-evoking constructions. Frame-evoking constructions can be classified based on the kinds of frames they evoke: those evoking semantic fames (Type [4] in Table 2); and those evoking interactional frames (Type [5] in

Table 2) (Fillmore, 1982, p. 117 (p. 379 in Geeraerts (Ed.), 2006)).[10] Semantic frames are "script-like conceptual structures that describe a particular type of situation, object, or event along with its participants and props" (Ruppenhofer et al., 2016).[11] Interactional frames, on the other hand, have to do with "how we conceptualize what is going on between the speaker and the hearer, or between the author and the reader" (Fillmore 1982, p. 117 (p. 379 in Geeraerts (Ed.), 2006)). Interactional frames do not involve participants in situations and events, which correspond to FEs in semantic frames, but rather they have to do with interactions between the speaker and the hearer or between the author and the reader.

5.2.1   *Constructions evoking a semantic frame*

The comparative_inequality construction in English (e.g. *She is better than her father at chess*), which evokes the *Comparison_inequality frame (The ENTITY is compared against some STANDARD with respect to their values for some FEATURE), is an example of a Type [4] construction (cf. Hasegawa et al., 2010, p. 179–186). The comparative_inequality construction in Japanese, shown in (7a) below, also evokes the *Comparison_inequality frame and is thus categorized as a Type [4] construction.

(7a)   The comparative_inequality construction
{ [ENTITY1] [ENTITY2] [FEATURE] }

M	Clause
D1	NP. May include the phrase *no hoo* (Lit. 'the NP's side')
D2	NP, accompanied by a case marker *yori*
D3	a plain adjective
Interpretation	Evokes the comparative_inequality frame that reports inequalities between two *Entities* as arguments of a plain adjective.

{ [ENTITY1 *kore (no hoo) ga*] [ENTITY2 = CEE *are yori*] [FEATURE *nagai*] }
this GEN side NOM   that than   long
'This is longer than that.'

---

10. See Section 5.3 in Lyngfelt, Bäckström et al. (this volume) for constructions in Swedish with pragmatic functions, which are classified as non frame-bearing constructions.

11. Other terms such as *cognitive frames* (Fillmore, 1982, p. 117 (p. 379 in Geeraerts (Ed.), 2006); Fillmore and Baker 2010, p. 314), *linguistic frames* (Fillmore & Baker, 2010, p. 338) and *frames* (Fillmore and Baker 2010, p. 314) have been used to refer to the notion of semantic frames.

When a Type [4] construction evokes a semantic frame, the frame name is documented in its "Interpretation" section in the JFN Constructicon, as in (7a) above. In addition, the construct will have a separate layer of frame annotation, which includes the bracketing formula with the FEE and with the FE labels, as shown in (7b).

(7b)   The *Comparison_inequality frame:
- The ENTITY is compared against some STANDARD with respect to their values for some FEATURE.
- FEs: ENTITY, STANDARD, FEATURE

[_ENTITY_ *kore (no hoo) ga*] [_STANDARD = FEE_ *are yori*] [_FEATURE_ *nagai*]
this GEN side NOM   that than   long
'This is longer than that.'

5.2.2   *Constructions evoking an interactional frame*

The second type of frame-evoking construction involves those that evoke an interactional frame (Type [5] in Table 2). Type [5] constructions have various rhetorical constraints, and such constraints do not seem to involve ordinary semantic frames. As indicated above, instead of involving FEs in semantic frames, such rhetorical constraints address interactions between the speaker and the hearer or between the author and the writer. At the time of writing, very few interactional frames have been defined in FN. Exceptions include the Attention_getting frame. The FN Lexicon defines the frame as "(t)his frame covers terms used to get someone's attention, including INTERJECTIONS (e.g. *hey*, *yo*) and certain terms of address (ADDRESS_TERM), the latter of which may serve the function of attention getting" and LUs that evoke the frame include interjections such as *excuse me* and *hello there* and address terms such as *boy*. The FE ADDRESS_TERM may be a proper name.

In Japanese, the *te_linkage* construction is classified as Type [5]. (8a) is an instance of the *te_linkage* construction, in which two clauses are connected by a clause-linking marker *te* 'and.'

(8a)   *koosi ga   kaizyoo ni tui-te   kooen ga   hazimatta.*
lecturer NOM hall   LOC arrive-TE lecture NOM began
'The lecturer arrived at the hall, and the lecture began.'

According to Hasegawa (1996), the constraints on the use of *te_linkage* "are neither on syntactic structures alone, nor on semantic relations alone; they apply only when a particular syntagm is used to express a certain semantic relation." In this construction the speaker construes the two events presented in the two clauses connected by *te* to be somehow relevant. Thus, unlike (8a), (8b) below is unacceptable, since it is difficult to imagine a situation in which the arrival of *watasi* 'I,' who was in the audience, is relevant to the start of the lecture.

(8b) *#watasi ga kaizyoo ni tui-<u>te</u>   kooen ga   hazimatta.*
     I NOM   hall   LOC arrive-TE lecture NOM began
     (Intended) 'I arrived at the hall, and the lecture began.'

The `te_linkage` construction can thus be said to evoke an interactional frame, namely, the *`Relevancy` frame, which may be defined as "the Speaker construes the two reported events to be somehow relevant." Note that in this definition of the frame there is no reference to FEs, corresponding to participants and props in situations or events, since interactional frames do not involve participants and props in events and situations and instead they pertain to interactions between the speaker and the hearer or between the author and the reader. The `te_linkage` construction can be described as in (8c). Since there is no FE in the definition of the *`Relevancy` frame, there is no separate layer for frame annotation.

(8c)   The `te_linkage` construction
      { [CONJUNCT1] [CLAUSE-CONNECTIVE *te*] [CONJUNCT2] }

M	Bi-clausal sentence
D1	First clausal conjunct
D2	Clause-connective *te*
D3	Final clausal conjunct
Interpretation	Two clausal conjuncts report two events and the two events exhibit temporal sequentiality. The construction evokes the *`Relevancy` frame, in which the Speaker construes the two reported events to be somehow relevant.

      { [CONJUNCT1 *koosi ga kaizyoo ni tui*] [CLAUSE-CONNECTIVE = CEE *te*]
      lecturer NOM hall   LOC arrive-TE
      [CONJUNCT2 *kooen ga hazimatta*] }
      lecture NOM began
      'The lecturer arrived at the hall, and the lecture began.'

The `suspended_clause` construction in Japanese is another example of a Type [5] construction, exemplified in (9a). Here, the speaker, trying to end a conversation on the phone, first says *sore zya ne* 'that's it!' angrily and then utters *kirase te morau <u>kara</u>*, which can be translated into English as '<u>Because</u> I'm gonna hang up.' What the speaker conveys by the second sentence is a message that s/he does not want the hearer to bother him/her anymore. In Japanese a dependent clause with a clause-linking marker such as *kara* 'because' is typically followed by a main clause. In (9a), however, there is no main clause that follows the clause-linking marker *kara*. Hence the name "suspended" clause.[12]

---

12. This kind of structure is found in many languages and is often called "insubordination" (Evans and Watanabe, 2016).

(9a) *sore zya   ne. kir   -ase   te morau <u>kara</u>*
      that DAT-TOP SFP hang-up CAUS AUX because
      [On the phone] (Lit.) 'That's it! <u>Because</u> I'm gonna hang up. (i.e., I don't want
      you to bother me anymore).'

A "suspended" clause is not just an ellipsis of a main clause, since there is no need
to reconstruct the content of the "missing" main clause; and in this construction,
there is a conventionalized implicature. The speaker expects the hearer's empathy
toward the speaker's situations (Ohori 2002). Therefore, it is possible to characterize
the suspended_clause construction as evoking the *Expect_empathy frame, an
interactional frame defined as "the Speaker expects the Hearer to empathize with
the Speaker's situation." The suspended_clause construction is shown in (9b).

(9b) The suspended_clause construction
      { [CLAUSE] [CLAUSE-CONNECTIVE] }
      M   Clause
      D1   Clause
      D2   Clause-connective such as *kara* 'because,' *node* 'because,' *ga*
                       'but,' and *kedo* 'but'
      Interpretation Evokes the *Expect_empathy frame, in which the Speaker
                       expects the Hearer to empathize with the Speaker's situation.
      *sore zya   ne. { [CLAUSE *kir -ase te morau*] [CLAUSE-CONNECTIVE= CEE <u>*kara*</u>] }
      that DAT-TOP SFP hang-up CAUS AUX   because
      [On the phone] (Lit.) 'That's it. <u>Because</u> I'm gonna hang up. (i.e., I don't want
      you to bother me anymore).'

As in the case of the te_linkage construction, there is no separate layer for frame
annotation in the case of the suspended_clause construction either, since the
frame that the construction evokes, namely, the *Expect_empathy frame, is an
interactional frame and hence there is no FE involved.

In this section I have presented a five-way classification of constructions, cat-
egorizing constructions based on whether they evoke frames and based on which
type of frames is involved. In the case of constructions that evoke semantic frames,
the construction annotation has an additional layer for a frame annotation. On the
other hand, construction annotations of constructions evoking interactional frames
do not have an additional layer for a frame annotation, since interactional frames
do not have to do with FEs, that is, participants in events and situations described
in the sentence. The five types of constructions proposed above seem to be mutu-
ally exclusive, although we have yet to find out whether they are exhaustive or not.

6. Summary and conclusion

In this chapter, based on the idea that framenets and constructicons are practical implementations of the theories of Frame Semantics and of Costruction Grammar, I have discussed relations between frames and constructions. I illustrated the differences between framenet annotation and constructicon annotation. Framenet annotation involves describing frame-based syntactic and semantic structures of words, both simple words and multiwords, resulting in their valence patterns. Framenet annotation processes consist of: identifying the FEE; identifying constituents corresponding to the FE's of the frame; and annotating the constituents with the PT and GF labels. On the other hand, constructicon annotation pertains to internal and external syntax/semantics of complex linguistic objects. Constructicon annotation processes consist of: identifying the CEE if there is any; identifying the construct's span of text; and identifying the construct-internal constituents as instantiating CEs.

Instead of asking the question of whether all constructions are meaning-bearing, I proposed that it is possible to classify constructions according to whether or not they evoke a frame and based on which type of frame is evoked and suggested a five-way classification. I also pointed out that for constructicon annotation, we need interactional frames in addition to semantic frames. It may turn out to be impossible to incorporate interactional frames to framenets, since they have to do with the speaker-hearer/writer-reader interactions but not with FEs, However, it is beyond the scope of this present paper to discuss this issue.

I hope to have shown a way to describe how grammatical constructions relate to semantic and interactional frames, the two kinds of frames that were originally proposed by Fillmore in the early days of Frame Semantics.

References

Boas, H. C. (2010). The Syntax-lexicon continuum in Construction Grammar: A case study of English communication verbs. *Belgian Journal of Linguistics*. 24, 54–82. doi:10.1075/bjl.24.03boa

Evans, N., & Watanabe, H. (2016). *Insubordination*. Amsterdam: John Benjamins. doi:10.1075/tsl.115

Fillmore, C. J. (1982). Frame Semantics. In The Linguistic Society of Korea (Ed.), *Linguistics in the Morning Calm* (pp. 111–137). Seoul: Hanshin Publishing Company. Reprinted in D. Geeraerts (Ed.), *Cognitive Linguistics: Basic Readings* (pp. 373–400). Berlin & New York: Mouton de Gruyter.

Fillmore, C. J. (1988). The mechanisms of 'Construction Grammar'. *Berkeley Linguistic Society*, 14, 35–55.

Fillmore, C. J. (1999). Inversion and Constructional Inheritance. In G. Webelhuth, J. P. Koenig, & A. Kathol (Eds.), *Lexical and Constructional Aspects of Linguistic Explanation* (pp. 113–128). Stanford: CSLI.

Fillmore, C. J. (2008). Border Conflicts: FrameNet meets Construction Grammar. In E. Bernal, & J. DeCesaris (Eds.), *Proceedings of the XIII EURALEX International Congress* (pp. 49–68). Barcelona: Universitat Pompeu Fabra.

Fillmore, C. J. (2013). Berkeley Construction Grammar. In T. Hoffmann & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 111–132). New York: Oxford University Press.

Fillmore, C. J., Kay, P. & O'Connor, M. C. (1988). Regularity and Idiomaticity in Grammatical Constructions: The Case of Let Alone. *Language* 64(3), 501–538. doi:10.2307/414531

Fillmore, C. J., Lee-Goldman, R., & Rhomieux, R. (2012). The FrameNet Constructicon. In H. C. Boas & I. A. Sag (Eds.), *Sign-Based Construction Grammar* (pp. 309–372). Stanford: CSLI Publications.

Fillmore, C. J., & Baker, C. (2010). A Frames Approach to Semantic Analysis. In B. Heine, & H. Narrog (Eds.), *The Oxford Handbook of Linguistic Analysis* (pp. 313–340). Oxford: Oxford University Press.

Goldberg, A. E. (2006). *Constructions at Work: The Nature of Generalization in Language*. New York: Oxford University Press.

Hasegawa, Y. (1996). Toward a description of *te*-linkage in Japanese. In M. Shibatani & S. A. Thompson (Eds.), *Grammatical Constructions* (pp. 55–75). New York: Oxford University Press.

Hasegawa, Y., Lee-Goldman, R. R., Ohara, K. H., Fujii, S., & Fillmore, C. J. (2010). On expressing measurement and comparison in Japanese and English. In H. C. Boas (Ed.), *Contrastive Construction Grammars* (pp. 169–200). Amsterdam: John Benjamins. doi:10.1075/cal.10.08has

Hilpert, M. (2014). *Construction Grammar and its Application to English*. Edinburgh: Edinburgh University Press.

Jurafsky, D. (1991). *An On-line Computational Model of Human Sentence Interpretation: A Theory of the Representation and Use of Linguistic Knowledge*. (Doctoral dissertation). Dept. of Electrical Engineering and Computer Science, University of California, Berkeley.

Ohara, K. H. (2013). Toward Constructicon Building for Japanese in Japanese FrameNet. *Veredas* 17(1), pp. 11–27.

Ohara, K. H. (2014). Relating Frames and Constructions in Japanese FrameNet. In *Proceedings of LREC 2014* (pp. 2474–2477). Reykjavík: ELRA.

Ohara, K. H. (In Press). Internally headed relativization and its related constructions. In Y. Hasegawa (Ed.), *The Cambridge Handbook of Japanese Linguistics*. Cambridge: Cambridge University Press.

Ohori, T. (2002). (In Japanese) *Cognitive Linguistics*. Tokyo: University of Tokyo Press.

Ruppenhofer, J., Ellsworth, M., Petruck, M. R. L., Johnson, C. R. & Scheffczyk, J. (2016). *FrameNet II: Extended Theory and Practice*. Berkeley: International Computer Science Institute. Retrieved from https://framenet.icsi.berkeley.edu/fndrupal/the_book

Sköldberg, E., Bäckström, L., Borin, L., Forsberg, M., Lyngfelt, B., Olsson, L. -J., Prentice, J., Rydstedt, R., Tingsell, S., & Uppström, J. (2013). Between Grammar and Dictionaries: a Swedish Constructicon. In *Proceedings of eLex 2013* (pp. 310–327). Tallinn: Eesti Keele Instituut.

Torrent, T. T., Lage, L. M., Sampaio, T. F., Tavares, T. S., & Matos, E. E. S. (2014). Revisiting border conflicts between FrameNet and Construction Grammar: Annotation policies for the Brazilian Portuguese Constructicon. *Constructions and Frames*, 6(1), 34–51. doi:10.1075/cf.6.1.03tor

A constructicon for Russian

Filling in the gaps

Laura A. Janda[i], Olga Lyashevskaya[ii], Tore Nesset[i],
Ekaterina Rakhilina[ii,iii] and Francis M. Tyers[ii]
[i]UiT The Arctic University of Norway / [ii]National Research University
Higher School of Economics / [iii]Vinogradov Institute for Russian language

The Russian Constructicon project currently prioritizes multi-word constructions that are not represented in dictionaries and that are especially useful for learners of Russian. The immediate goal is to identify constructions and determine the semantic constraints on their slots. The Russian Constructicon is being built in parallel with the Swedish Constructicon and will ultimately model the entire Russian language in terms of constructions at all levels from morpheme to discourse. The contents of the Russian Constructicon will serve learners of the language, linguists researching both language-internal and typological phenomena, and will also serve language technology applications such as spell checkers and automated readability assessment tools.

Keywords: Russian, constructions, multi-word, semantic constraints, learners, language technology, typology

1. Introduction

The Russian Constructicon project has emerged organically from a milieu with sustained focus on a range of relevant theoretical and practical aims, including: construction grammar, lexical semantics, quantitative analysis of language data, and development of pedagogical materials for learners of Russian as well as language technology resources for users of Russian. While each of these ideas and undertakings approach Russian from a unique perspective, they all converge on a single challenge, namely the lack of an extensive inventory of Russian constructions. Certain kinds of constructions are represented in dictionaries and other reference works, but many types of constructions are not. This is not due to any shortcomings in such reference works, but due to the fact that their mission is fundamentally different:

they are not designed to deliver a full-scale inventory of the constructions that are useful for second-language learners. For example, a dictionary will not predict constructions of the type: (1) X *tak i ne* Vpast [X.NOM thus and not V.PST] 'But X didn't V after all', as in *On tak i ne ženilsja* [He.NOM thus and not marry.PST.M.SG] 'But he didn't get married after all'; and (2) *Raz i ty* … [Once and you …] '(And) before you know it you're …', as in *Raz i ty v belom plat'e* [Once and you in white. LOC.SG.N dress.LOC.SG] 'And before you know it you're wearing a white dress'.

In its present stage, the Russian Constructicon project is focused on those "missing" constructions, particularly the constructions that are most essential for learners of Russian. A prototypical characteristic of the constructions that we target is the presence of one or more "slots" (see Section 3) that can be filled with a range of words depending on the semantic restrictions of the given construction.

A concise history tracing the relationship of the Russian Constructicon project to the research agendas of its partners appears in Section 2. Section 3 details both the types of constructions that have been previously documented and those that have not, and then presents examples of the types of constructions that the project is currently collecting. Descriptions of annotation and interpretation techniques are provided in Section 4. In Sections 5 and 6 we project the benefits of the Russian Constructicon both in terms of further research it will facilitate and user applications (language technology resources) that can be built on or enhanced by the Constructicon.

2.   History and partners

In Russia Construction Grammar has a long history. An important early contribution is the Meaning Text Theory proposed by Meľčuk, Apresjan and Žolkovskij (Žolkovskij & Meľčuk, 1965), whose lexical functions became one of the basic concepts in the Moscow Semantic School Approach. This included the analysis of constructions with so-called "light verbs", intensifiers, and lexicalized evaluative expressions. Additionally, Apresjan (1967) explored verbal government as morpho-syntactic relationships motivated by semantics. Švedova (1960) developed the idea of syntactic schemas associated with words and turned attention to patterns that are significant in spoken Russian, beginning with the use of reduplication. Another important theoretical contribution was made by Zolotova in her *Syntactic Dictionary* (2006) which gave an inventory of minimal units of Russian syntax. However, all of these works aimed at constructions at a rather abstract and generic level, focusing on the basic syntax of the language. More recently, Rakhilina's group has focused on these data in the theoretical context of Construction Grammar, analyzed a number of core and non-core Russian constructions, and showed how the constructions

are organized (Rakhilina, 2010). Saj (2008, 2014; Ovsjannikova & Saj, 2014) has led research on the syntactic periphery of Russian, on the interrelationships between lexical items and slots, as well as constructions that do not conform to core syntax. Rakhilina and Letučij (2012; Letučij & Rakhilina, 2014) have focused on what they call "quasigrammatical" constructions and the ways in which they relate to various semantic fields such as time, iterativity, and quantification. Kuznetsova (2015) and Janda and Solovyev (2009) have taken a quantitative approach to the study of Russian constructions to determine the relationships between lexical items and slots.

Despite Russian's vast size (it ranks sixth in the world in terms of total number of speakers, eighth in terms of L1 speakers; https://www.ethnologue.com/statistics/size), the Russian language lags far behind English in terms of electronic resource development. A Russian Constructicon is an important component in addressing this need. Our project approaches the building of a Russian Constructicon from the complementary perspectives of native and non-native language users, and we achieve this through collaboration between Russian and foreign researchers.

The main partners in the project are linguists at the Higher School of Economics (HSE) in Moscow (https://ling.hse.ru/en/) and their counterparts at The Arctic University of Norway (UiT) in Tromsø, namely those in the CLEAR (Cognitive Linguistics: Empirical Approaches to Russian) research group (https://uit.no/forskning/forskningsgrupper/gruppe?p_document_id=344365), and SweCcn – a Swedish Constructicon research group at the University of Gothenburg (https://spraakbanken.gu.se/eng/sweccn).

The partners at HSE and UiT share three core features in their linguistic agendas: (1) the theoretical framework of cognitive linguistics, (2) focus on construction grammar, and (3) statistical analysis of linguistic data. All three of these features directly support the development of a Russian Constructicon as a natural outgrowth of established research traditions.

The Russian Meaning Text Theory and other semantic theories that have emerged in Russia are highly compatible with cognitive linguistics (Rakhilina, 2000, pp. 342–378). Linguists of Russian at UiT share the theoretical commitment to cognitive linguistics, and within that theoretical framework, both groups of linguists have consistently focused on construction grammar. Both groups of linguists have also applied quantitative methods to the study of Russian linguistics, and in both cases this has spilled over into computational approaches and applications (Janda, 2013; Lyashevskaya, 2016). The Russian partners play leading roles in the continuing development of the Russian National Corpus (the foremost linguistic database of Russian, http://ruscorpora.ru/, released in 2002 and under continuous expansion and development). HSE is a world leader in the development of open-source electronic resources for Russian, such as learner corpora, a corpus

of heritage Russian, corpora of dialectal and regional Russian, a tutorial in academic writing, and the semantic edition of Tolstoy's collected works (a digital humanities project). UiT has developed UDAR (Reynolds, 2016), the only full-scale open-source finite-state transducer morphological computational model of Russian that takes into account stress (the placement of accents of words, which can convey meaning differences, as in *dóma* [at.home] 'at home' and *dóm-a* [house-GEN.SG] 'house' vs. *dom-á* [house-NOM.PL] 'houses'), and is also engaged in the development of interactive and web-enhancement resources for learners of Russian.

In a very concrete sense, the Russian Constructicon project has evolved from traditional reference and electronic corpus resources. There are, of course, a myriad of dictionaries of Russian, but among these Zaliznjak 1980 stands out as a landmark work that for the first time detailed the morphological forms for all inflected words in Russian. Among major world languages, Russian is relatively morphologically complex, with large inflectional paradigms for nouns, adjectives, and verbs, features usually associated with minority languages (McWhorter, 2011; Trudgill, 2011). Zaliznjak set the standard for interpreting and modeling the morphological complexity of Russian, an essential component for most language technology resources. Zaliznjak 1980 is also a cornerstone of the Russian National Corpus (as well as other corpora of Russian), since the morphological analysis it uses, both in the portion of the corpus that is automatically tagged and in the portion that has been manually tagged, is based directly on Zaliznjak's model of Russian morphology.

Morphological analysis of Russian has made it possible to search for lexemes in a corpus (since all forms of a word can be associated with the appropriate lexeme), and this in turn has facilitated the creation of corpus-based dictionaries such as the frequency dictionary by Lyashevskaya and Sharoff (2009).

Corpus-based research set the stage for the systematic study of linguistic constructions, since it raised the issue of how to track and interpret units larger than single words. A number of resources have been developed to address this need, all of which give a firm basis for the building of a Russian Constructicon. One outcome of this line of work is the *Corpus Dictionary of Multi-Word Lexical Units* (2008, http://ruscorpora.ru/obgrams.html), composed of data on frequent collocations in the Russian National Corpus, with supplementary material from Rogožnikova's (2003) dictionary of collocations and the four-volume academy dictionary of Russian (Evgen'eva, 1999). The *Corpus Dictionary of Multi-Word Lexical Units* lists over 2900 such collocations, along with their frequency (in the Russian National Corpus as of 2008), and links to corpus examples. This inventory is broken down into five groups according to syntactic-semantic functions: (1) multi-word units functioning as prepositions like *vo imja X* [in name.ACC X.GEN] 'for the sake of X'; (2) adverbial and predicational multi-word units like *na vsjak-ij slučaj* [on any-ACC.

SG event.ACC.SG] 'just in case'; (3) parenthetic multi-word units like *s točk-i zreni-ja X* [from point-GEN.SG view-GEN.SG X.GEN] 'from X's point of view'; (4) multi-word units that function as conjunctions like *dlja togo čtob(y)* [for that.GEN so-that] 'in order that'; and (5) multi-word units that function as particles like *ne inače* [not otherwise] 'certainly'.

Another outcome of corpus investigations into units larger than the lexeme was the Russian FrameBank (http://framebank.ru/; Lyashevskaya & Kashkin, 2015). Analogous to FrameNet for English (Fillmore et al., 2008), the Russian FrameBank draws on Russian lexicographical traditions and traditional printed dictionaries (Apresjan & Pall, 1982; Sazonova, 2008). The result is a hybrid resource that integrates dictionary-style information about verbal government (e.g., valency, syntactic frames) with linguistic interpretation of corpus data. The Russian FrameBank is centered on 2700 high-frequency verbs in Russian and the constructions that they appear in, both in corpus data (100 corpus examples for each verb are fully parsed both semantically and syntactically and classified according to construction type), and according to dictionaries (which may list constructions in addition to those found in the 100 corpus examples). For example, for the verb *vzjat'* 'take', there are three examples among the 100 corpus examples of the S.NOM V S.ACC v + S.ACC construction, as in *Ja sam vzja-l v ruk-i mokr-uju xolodn-uju butylk-u* [I.NOM self. NOM.SG.M take-PST.M.SG in hand-ACC.PL wet-ACC.SG.F cold-ACC.SG.F bottle-ACC. SG] 'I myself took the cold wet bottle in my hands'. However, there are no corpus examples in the FrameBank sample of the type S.NOM V S.ACC S.INS like *On vzja-l ščit lev-oj ruk-oj* [He.NOM take-PST.M.SG shield.ACC.SG left-INS.SG.F hand-INS.SG] 'He took the shield with his left hand', although this type is attested in dictionaries and can be located in more extensive corpus searches. At the present time this research is also being extended to constructions associated with adjectives, yielding findings concerning constraints such as the limitation to predicative use for the Adj *na* + ACC construction in for example *On sposoben na podlost'* [He.NOM capable.M.SG. SHORT-FORM on meanness.ACC.SG] 'he is capable of meanness', and that superlative forms can have different argument structure properties than their neutral equivalents, as in *lučš-ij/sam-yj xoroš-ij v mir-e* [best-NOM.SG.M in world-LOC.SG] 'best in the world', cf. the unattested *xorošij v mire 'good in the world'.

The FrameBank hybrid between a linguistic reference work and a portal for corpus examples is the future of the dictionary as envisioned by Atkins (1992) and Kilgarriff et al. (2006), and also leads us in the direction of a dictionary of constructions, or a constructicon. However, as we detail in the following section, there remain gaps in our coverage of the syntactic-semantic peculiarities of Russian, and the Russian Constructicon project is designed to fill those gaps.

3.   Russian constructions: What's missing

Our starting point is construction grammar as outlined by Langacker (1987, 1991a–b, 2003), Croft (2001), Goldberg (1995, 2006), and Fillmore (1985; Kay & Fillmore, 1999). Although these scholars take slightly different perspectives on constructions, they all share a similar view on what constitutes a construction, namely any conventionalized pairing of form and meaning in language, at any level, from the level of the morpheme, through words and phrases, and up to the level of discourse. The meaning of each construction is emergent (Langacker, 1991b, pp. 5–6, 534), motivated by the patterns of uses of the units that appear in the construction, and also by the larger (clause- or discourse-level) constructions that a given construction appears in. Since a language is a network of interrelated constructions, a constructicon is a model of an entire language. While our ultimate goal is to create a full-scale constructicon, at present we have made strategic decisions to prioritize the types of constructions that should be collected.

The Russian Constructicon project specifically addresses the resources that are lacking for both research and pedagogy with respect to Russian constructions. Some items along the scale from morphemes to discourse are relatively well described. For example, at the word level we have traditional dictionaries, and the argument structures associated with lexemes are detailed in those sources and in the Russian FrameBank. There also exist phraseological dictionaries (Mixel'son, 1896–1912/2004; Lubensky, 1995; Bystrova, 1997; Kuz'mič, 2000; Fedosov, 2003), but these have a strong bias toward very specific types of phrases such as sayings, aphorisms, and proverbs where the relationship between the components and the semantics of the whole are particularly obscure. Take for example the phrase *kak siv-yj merin* [like gray-NOM.SG.M gelded.horse.NOM.SG], which literally means 'like a gray gelded horse', but actually describes a particularly dishonest manner of behavior (usually in relation to telling lies), roughly equivalent to 'through one's teeth' in English collocations like *He's lying through his teeth*. These are the types of entries one finds in phraseological dictionaries, but phrases like this tend to be of very low frequency – *kak sivyj merin* appears only twenty-seven times in the entire Russian National Corpus, approximately once in ten million words. Such phrases are idiosyncratic, thus rarely yielding general patterns that would be of interest to theoretical linguists, and so infrequent as to be of little use to learners of Russian as a second language. Some of these resources, while they have their merits, are themselves very skewed. For instance, Baranov et al.'s (2009) dictionary-thesaurus of Russian idioms was largely compiled from detective stories and thus overrepresents phrases used in Russian taboo expressions and swearing known as *mat*, a much stronger genre than its English correlate and decidedly inappropriate for use by second language learners in most contexts. What learners really need are

phrases that phraseological dictionaries overlook, such as *davaj ruku* [give.ɪᴍᴘ.sɢ hand.ᴀᴄᴄ.sɢ] 'give me your hand (so that we can shake hands or so that you can help me get up)' and *skol'ko možno!* [how-much possible] (lit. 'how much is possible (for someone to X)') 'oh, for crying out loud, give it a rest already!' (used to express exasperation at excessive talk about something).

Because of its inflectional morphology, pedagogical materials for learners of Russian invest heavily in teaching paradigms and grammatical endings. This is well justified since in a very real sense, one cannot even begin to speak Russian without mastering a large portion of the grammatical inflections. It is fairly easy to succeed at speaking "bad" English for example, by merely stringing together lexemes in a largely predictable order – the result will not be idiomatic, but you can get your message across. However, speaking good English is very hard. The big hurdle for learners of English comes along when they try to master the constructions, making the need for a constructicon very obvious. For learners of Russian, it is difficult even to speak badly and be understood since all words in a sentence (except for prepositions and conjunctions and a few "particles", see Endresen et al., 2016) have to be inflected. Russian grammatical morphology has to be acquired and routinized to a high degree right from the beginning. This is a huge task and for this reason the American Council of Teachers of Foreign Languages (http://www.languagetesting. com/how-long-does-it-take) ranks Russian among the "Group III Languages" in their four-point scale (with respect to difficulty for learners whose native language is English). Group I Languages (Afrikaans, Danish, Dutch, French, etc.) are learned easily and quickly, followed by Group II Languages (Bulgarian, Dari, Farsi, German, etc.). Group IV Languages (Arabic, Chinese, Japanese, Korean) are the hardest to acquire, and note that none of those are Indo-European. Russian is the Group III Language with the largest number of speakers and the only major world language in that group (other languages in Group III are: Amharic, Bengali, Burmese, Czech, Finnish, Hebrew, Hungarian, Khmer, Lao, Nepali, Filipino, Polish, Serbo-Croatian, Sinhala, Thai, Tamil, Turkish, and Vietnamese). While Russian textbooks place considerable emphasis on inflectional morphology, they tend to have at best only sporadic coverage of constructions.

However the need to master the morphology doesn't mean that speaking good Russian is any less dependent on knowing constructions. Indeed, the morphology serves an essential role in Russian constructions. But the constructional landscape of Russian is also highly complex, and we are only beginning to explore that landscape. Especially with respect to the needs of learners, Russian constructions are woefully underdescribed.

Our Russian Constructicon project currently prioritizes the constructions that are missing from other resources. This means that we are not concerned with units at the word-level (since those are represented in dictionaries), nor with the

argument structure features of word-level units (since those are represented in dictionaries and FrameBank), nor with sayings and phrases (since those are represented in phraseological dictionaries). Our focus is on the multi-word units that are not represented in other resources and particularly on those that are most useful to learners of Russian.

We use a team strategy in our approach to constructions. Native speakers are typically blind with regard to the constructions that are challenging to learners, since to them all constructions are equally comprehensible. Non-Russian team members are needed to identify the constructions that stand between learners and Russian proficiency. However, only native Russian team members have the capacity to fully interpret and annotate constructions.

The multi-word units that we target present a range of types that vary according to the presence vs. absence both of "slots" (underdetermined portions of constructions) and of constraints on those slots. On one end of the scale are fixed expressions where all the components are obligatory and unchangeable, such as *Kto tam?* [who.NOM there] 'Who's there?' (a response to a knock at the door) or *Vot ešče!* [Look still] 'No way!'. Baranov et al. (2009) call these "situational clichés", and they can be thought of as degenerate constructions. Although technically they have no slots, they still have variables, since there often has to be something that precedes or follows them, such as the knock at the door before *Kto tam?* 'Who's there?' or the specification of what is being rejected in *Vot ešče! Ja posud-u my-t' ne bud-u.* [Look still I.NOM dishes-ACC.SG wash-INF not be.FUT-1SG] 'No way! I'm not going to wash the dishes.' There are also some constructions that approach this degenerate type because they have severe restrictions on their slots, as in *Èx ty!* [Oh you.NOM] 'Shame on you! Darn!' (said when something doesn't work out). For some speakers this is a fixed expression allowing only the second person singular (intimate) or plural pronouns *ty* and *vy* 'you', while others can also admit (usually reduplicated) names for people (*Èx Vitja, Vitja*), but other fillers are excluded.

On the other end of the scale are syntactic constructions that have almost no constraints on their slots. This type is called a "schema" by Švedova (Švedova et al., 1980) and Belošapkova (1977), and can be realized as both a simple sentence like *Kakoj* X! [What.NOM.SG.M X.NOM.SG] 'What a X!' (where the adjective *kakoj* 'what kind of' needs to have the correct inflectional ending to agree with the number and gender of whatever noun goes in the slot), or as a complex sentence like *Esli* Y*, togda* Z [If Y, then Z] 'If Y, then Z'. At present we lack inventories of both the slotless degenerate type of constructions and those that are maximally open.

However, the most interesting constructions are those that lie between these two extremes, namely those with various types of restrictions on their slots, and these can include both constructions that constitute entire sentences and those that are phrases. The tendency here is that when there are lexical constants in a

construction (items that are fixed, cf. Fillmore et al., 1988), there are also greater semantic constraints on the slots. For example, the construction *Kak u vas s* X? [How by you.ɢᴇɴ with X.ɪɴs] 'What's your X situation like?' has a semantic restriction on the range of words that can go in the X slot, which most often refer to essential challenges for human beings like *pitaniem* 'food', *zdorov'em* 'health', *den'gami* 'money', *pogodoj* 'weather', *nasledstvennost'ju* 'inheritance'. It is hard to come up with contexts that would support the use of other kinds of fillers, such as features of nature like *nebom* 'sky' and deverbals like *prixodom* 'arrival' in this slot. A phrase-level example is *Let* Y [year.ɢᴇɴ.ᴘʟ Y.ɢᴇɴ] 'about Y years old, in his/her Y-ies', as in *let semidesjati* [year.ɢᴇɴ.ᴘʟ seventy.ɢᴇɴ] 'about seventy years old, in his/her seventies'. In addition, constructions can have multiple slots, all of which have semantic restrictions and some of which can be optional. For example, there is the X *v* Y [X in Y.ᴀᴄᴄ] construction used to describe patterns on clothing, as in *jubk-a v kletočk-u* [skirt-ɴᴏᴍ.sɢ in check-ᴀᴄᴄ.sɢ] 'a checkered skirt'. The X slot is filled with a noun that refers to an article of clothing (which may be singular or plural), and the case marking depends on the role of that noun in the larger sentence. The Y slot refers to a type of pattern, usually additionally marked with a diminutive suffix containing *k*, such as *poloska* 'stripe', *gorošek* 'polka dot', *cvetoček* 'flower'. An option is to also specify the color(s) of the pattern by inserting one or more adjectives designating colors before Y.

Our current task in building the Russian Constructicon is to study such syntactic fragments and their interpretations, to map out the range of Russian constructions, and work out the semantic restrictions on their slots.

4. Status of the project and examples from the Russian constructicon

Textbooks of Russian and texts that represent or approximate spoken language (for example children's stories and films as well as the prose of certain writers such as Sergej D. Dovlatov) are good sources for the type of constructions we are targeting. At present over six hundred constructions have been entered in *A Constructicon for Russian* at https://spraakbanken.gu.se/karp/#?mode=konstruktikon-rus. This site uses the same architecture as the Swedish Constructicon and thus preserves all the search and other features of that constructicon and is designed to be comparable across languages.

A full entry in the Russian Constructicon can include up to five elements: NAME, DEFINITION, STRUCTURE, EXAMPLES, and COMMENT, as shown in Figure 1.

NAME	**60_километров_в_час** – *два раза в день, 60 километров в час*
DEFINITION	Употребляется для обозначения скорости или частоты. Обозначает [расстояние]Distance, [количество квантов]Quantity [повторяющегося действия]Event , [объём вещества]Quantity или [расход денег]Cost, приходящихся на [период времени]Time.
STRUCTURE	[root NP [nummod Num] [nmod [case в] NP-Acc]]
EXAMPLES	Разрешенная [скорость]Measure [движения]Event на этом крайне опасном участке дороги — не более [[40]Quantity [км]Unit [в]cee [час]Time]60_километров_в_час.
	[Поезд будет курсировать между Миланом и Неаполем]Event со [скоростью]Measure [[300]Quantity [километров]Unit [в]cee [час]Time]60_километров_в_час.
	[Выпивай по стакану чудесного напитка]Event [[два]Quantity [раза]Unit [в]cee [день]Time]60_километров_в_час.
	А хотите, я скажу вам, как [продавать]Event [[300]Quantity [литров]Unit [в]cee [день]Time]60_километров_в_час?
	Значит, [прибавка к пенсии]Event не превысит [[88]Quantity [рублей]Unit [в]cee [месяц]Time]60_километров_в_час.
	Такие красавцы, я думаю, [рождаются]Event [[раз]Quantity [в]cee [сто лет]Time]60_километров_в_час.
COMMENT	В письменной речи обозначается также как 60 км/ч. Обратите внимание, что расход топлива обозначается с помощью другой конструкции (ср. семь литров на 100 километров).
SWECCN	AVBCLFmN9pg0dX7Prwnn

Figure 1. Example entry in the Russian constructicon

The NAME of a construction may either represent an example or be more schematic, depending on the construction. In Figure 1, the NAME is 60_kilometrov_ v_čas 'sixty kilometers an/per hour', and a second example is also supplied: *dva raz-a v den*' [two.NOM time-GEN.SG in day.ACC.SG] 'two times a/per day'. The NAME is used in the EXAMPLES section, appearing in blue with square brackets to show where the construction begins and ends.

The DEFINITION of the construction describes its semantics, with tags for the elements. In our example in Figure 1, the DEFINITION (translated from the Russian) is: Used to designate speed or frequency. Designates the [distance]$_{Distance}$, [number of units]$_{Quantity}$ of a [repeated action]$_{Event}$, the [volume of a substance] $_{Quantity}$ or the [expenditure of money]$_{Cost}$, that occurs over a [period of time]$_{Time}$. The tags appear in red and are used also in the EXAMPLES with square brackets so that is it easy to keep track of correspondences. The tags and the definition aim to capture the semantic restrictions on the slots of the construction. Thus, for example, we see that there is an Event (named outside the construction) that involves a Quantity, usually expressed with a numeral and a Unit in relation to a period of Time.

The STRUCTURE of this construction is rendered in Universal Dependency Grammar as: [root NP [nummod Num] [nmod [case v] NP-Acc]. This means that there are two noun phrases and the preposition *v* 'in' in the construction. The first noun phrase can contain a numeral and a noun phrase quantified by that numeral. The second noun phrase is governed by the preposition in the accusative case.

The EXAMPLES for this construction are as follows:

(1) *Razrešenn-aja   skorost'   dviženij-a   na èt-om*
    allowed-NOM.SG.F speed.NOM.SG movement-GEN.SG on that-LOC.SG.M
    krajne   opasn-om   učastk-e   dorog-i   ne bolee
    extremely dangerous-LOC.SG.M portion-LOC.SG road-GEN.SG not more
    40   km   v čas
    40.GEN   km.GEN.PL in hour.ACC.SG
    'The speed limit for that extremely dangerous part of the road is not more than 40 km per hour.'

(2) *Poezd   bud-et   kursirova-t' meždu   Milan-om   i*
    train.NOM.SG be.FUT-3SG shuttle-INF between Milan-INS.SG and
    Neapol-em   so   skorost'ju   300 kilometr-ov   v čas
    Naples-INS.SG with speed-INS.SG 300 kilometer-GEN.PL in hour.NOM.SG
    'The train will shuttle between Milan and Naples at a speed of 300 kilometers per hour.'

(3) *Vypivaj   po   stakan-u   čudesn-ogo   napitk-a*
    drink.IMP.SG along glass-DAT.SG marvelous-GEN.SG.M beverage-GEN.SG
    dva   raz-a   v den'
    two.ACC time-GEN.SG in day.ACC.SG
    'Drink this marvelous beverage twice a day.'

(4) *A   xoti-te,   ja   skaž-u   vam,   kak prodava-t'*
    and want-PRS.2PL I.NOM tell-PRS.1SG you.DAT how sell-INF
    300   litr-ov   v den'
    300.ACC liter.GEN.PL in day.ACC.SG
    'And if you want, I will tell you how to sell 300 liters a day.'

(5) *Znači-t,   pribavka   k pensi-i   ne*
    mean-PRES.3SG supplement.NOM.SG to pension-DAT.SG not
    prevysi-t   88   rublej   v mesjac
    exceed-PRS.3SG 88.ACC ruble.GEN.PL in month.ACC.SG
    'In other words, the pension supplement will not exceed 88 rubles per month.'

(6) *Tak-ie   krasavc-i,   ja   dumaj-u,*
    Such-NOM.PL.M handsome.man-NOM.PL I.NOM think-PRS.1SG
    roždaj-ut-sja   raz   v sto   let
    be.born-PRS.1SG-REFL time.NOM.SG in hundred.ACC year.GEN.PL
    'Such handsome men, I think, are born once in a hundred years.'

The COMMENT for this entry is: "In writing it also appears as 60 km/h. Note that the use of fuel is designated by a different construction (compare seven liters in 100 kilometers)."

Although at present we are focusing on multi-word units, our aim is to model the entire Russian language in terms of constructions. To this end, existing resources (dictionaries, argument structure information from FrameBank) will be integrated into the Russian Constructicon, and the scope of the project will be extended to include both units smaller than a word (morphemes, derivational morphology) and concatenation of constructions into larger discourse units. For now, the entries are given in Russian, though in the future users will be able to get the definitions and comments also in other languages, such as English. In keeping with the pedagogical aims of the Russian Constructicon, materials and resources for learners will also be developed and integrated into this project.

5.   Further research facilitated by the Russian constructicon

Of course a large research investment will be made in the definition of Russian constructions and the semantic restrictions on their slots, and corpus linguistic techniques will play an important role in that research. However, the Russian Constructicon itself will also serve as a research tool. There are many directions which that research may take. For example, to date there has been very little research on typological comparisons of constructions across languages. Rakhilina and colleagues (Rakhilina & Majsak, 2007; Rakhilina et al., 2012; Rakhilina & Plungian, 2013) have pioneered typological work on the lexical semantics of certain domains (aquamotion, pain, speed), but such typological comparisons could be extended both in terms of the syntax and the semantics of constructions. Following this lead, it would be possible to take an onomasiological approach, starting from general types of meanings such as negation and indefiniteness, and examine how these meanings are expressed by constructions.

Within Russian, various kinds of classifications of constructions will reveal systematic grammatical patterns and also facilitate research as well as access to examples through our interface. For example, both syntactic and semantic classifications can be developed. This will make it possible to discover the relationships among constructions in what could be called "construction families" similar to the family of `Subject-Auxiliary Inversion` constructions in English (Goldberg, 2006, Chapter 8). Some preliminary work on paradigmatic relations among constructions has been attempted (Janda & Divjak, 2008), but only at a very schematic level (specifying the grammatical case of the arguments of verbs). Constructions can be grouped paradigmatically according to the part of speech that serves as their core (nouns, verbs, etc.). To our knowledge, no systematic study of the syntagmatic co-occurrence patterns of Russian constructions has been attempted, and this is

a very complex dimension, since constructions can be nested within each other, overlap, or be contiguous, even across sentence boundaries.

Another line of research that will benefit both description and pedagogy is the behavior of grammatical categories in constructions. Many constructions have slots for verbs, and in Russian all verbs express either perfective or imperfective aspect, referring to the way in which an event is understood, roughly as either a complete whole or as an unbounded situation. The category of aspect in Russian is among the most challenging grammatical concepts for learners of Russian. It is extremely difficult both for linguists and for language teachers to explain when to select a perfective or an imperfective verb. Textbooks devote considerable space to "rules" for using aspect, but nearly all such rules admit exceptions. These rules present various "triggers" for use of aspect, such as: "use perfective aspect in the presence of *uže* 'already'", or "use imperfective in the presence of *vsegda* 'always'". However, the triggers for such rules are actually fairly rare in authentic texts: Reynolds (2016) finds that these triggers co-occur with only about 2% of verbs in a corpus. While the triggers are good indicators of aspect (yielding 98% correct guesses according to rules), they aren't plentiful enough to be useful. In other words, by focusing on a small number of coarse-grained triggers, we are failing both as linguists to fully describe the phenomenon and as instructors to give our students adequate guidance. The Russian Constructicon will make it possible to investigate the parameters of less clear-cut cases. For example, a search in the Russian National Corpus reveals that 75% of the verbs that appear directly after *čtoby* 'in order to' are perfective. While 75% is a strong trend, it is not very reliable. We need more detail on exactly what kinds of constructions and which verbs influence the choice of aspect. The Russian Constructicon already shows promise, by identifying constructions where imperfective verbs are preferred after *čtoby* 'in order that', such as X `sozdan, čtoby` [X.NOM.SG created.NOM.SG.M in.order.that] 'X was made in order to' as in *avtomobil' sozdan, čtoby na nem ezdi-t'* [automobile.NOM.SG created.NOM.SG.M in.order.that on it.LOC ride-INF] 'the automobile was made to be ridden' and *sliškom* Y, *čtoby* [too Y in.order.that] 'too Y to' as in *ja sliškom ustal, čtoby sraža-t'-sja* [I.NOM too tired.M.SG in.order.that fight-INF-REFL] 'I'm too tired to fight'). Details like these can be used to calibrate more precise rules. And this kind of research can be extended to other grammatical categories and parts of speech. In this way, the Russian Constructicon provides added value for both researchers and learners.

6.   Applications served by the Russian constructicon

In addition to serving linguistic research and pedagogical needs, the Russian Constructicon has important implications for the development of language technology applications for Russian. Many types of constructions present challenges for computational processing even in morphologically tagged corpora; this is especially true for multi-word units that can be discontinuous and contain variable slots. A full-scale inventory of Russian constructions can improve the standard resources of language technology such as spell checkers and machine translation. The density and complexity of constructions are one indicator of the readability of texts that has until now remained beyond the reach of language technology (see the overview in Vajjala, 2015). Comparison of the constructions present in texts rated for readability can serve as training material for machine learning that will make it possible to automatically and accurately gauge appropriate reading materials for both native Russian schoolchildren and second language learners. Interactive learning and web enhancement tools (Meurers et al., 2010) can also be designed to focus on the task of mastering Russian constructions.

Acknowledgments

References

Apresjan, Ju. D. (1967). *Èksperimental'noe issledovanie semantiki russkogo glagola* [Experimental research on the semantics of the Russian verb]. Moscow: Nauka.

Apresjan, Ju. D., & Pall, È. (1982). *Russkij glagol – vengerskij glagol* [Russian verb – Hungarian verb]. Budapest: Tankönyvkviadó.

Atkins, S. (1992). Tools for computer-aided corpus lexicography: the Hector project. In F. Kiefer, G. Kiss & J. Pajsz (Eds.). *Papers in Computational Lexicography: Complex'92*, Hungarian Academy of Sciences, Budapest (pp. 1–60). Budapest: Akadémiai Kiadó.

Baranov, A. N., Voznesenskaja, M. M., Dobrovol'skij, D. O., Kiseleva, K. L., & Kozerenko, A. D. (2009). *Frazeologičeskij ob"jasnitel'nyj slovar' russkogo jazyka* (Phraseological explanatory dictionary). Moscow: ÈKSMO.

Belošapkova, V. A. (1977). *Sovremennyj russkij jazyk. Sintaksis* [Contemporary Russian. Syntax]. Moscow: Vysšaja škola.

Bystrova, E. A. (1997). *Učebnyj frazeologičeskij slovar'* [Academic phraseological dictionary]. Leningrad: Prosveščenie.

Croft, W. (2001). *Radical construction grammar*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780198299554.001.0001

Endresen, A., Janda, L. A., Reynolds, R., & Tyers, F. M. (2016). Who needs particles? A challenge to the classification of particles as a part of speech in Russian. *Russian Linguistics*. doi:10.1007/s11185-016-9160-2.

Evgen'eva, A. P., Ed. (1999). *Malyj akademičeksij slovar'* [Small academy dictionary]. Moscow: Russkij jazyk.

Fedosov, I. V. (2003). *Frazeologičeskij slovar' russkogo jazyka* [Phraseological dictionary of Russian]. Moscow: JuNVES.

Fillmore, C. (1985). Syntactic intrusions and the notion of grammatical construction. *Proceedings of the Berkeley Linguistics Society* 11, 73–86.

Fillmore, C., Kay, P., & O'Connor, C. (1988). Regularity and Idiomaticity in Grammatical Constructions: The Case of *let alone*. *Language* 64, 501–538. doi:10.2307/414531

Fillmore, C. J., Bernal, E. & DeCesaris, J. (2008). Border conflicts: FrameNet meets construction grammar. *Proceedings of the XIII {EURALEX} International Congress*. Barcelona: Universitat Pompeu Fabra. http://www.icsi.berkeley.edu/proceeds/2008/speech/Euralex2008/docum/10001.pdf

Goldberg, A. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: Chicago University Press.

Goldberg, A. (2006). *Constructions at work: The nature of generalizations in language*. Oxford: Oxford University Press.

Janda, L. A. (Ed.) (2013). *Cognitive linguistics: The quantitative turn. The essential reader*. Berlin: De Gruyter Mouton. doi:10.1515/9783110335255

Janda, L. A., & Divjak, D. (2008). Ways of attenuating agency in Russian. In A. Siewierska (Ed.), *Impersonal constructions in grammatical theory*, a special issue of Transactions of the Philological Society 106, 138–179.

Janda, L. A., & Solovyev, V. (2009). What constructional profiles reveal about synonymy: A case study of Russian words for SADNESS and HAPPINESS. *Cognitive Linguistics* 20, 367–393. doi:10.1515/COGL.2009.018

Kay, P., & Fillmore, C. (1999). Grammatical constructions and linguistic generalizations: the what's X doing Y? construction. *Language* 75,1–34. doi:10.2307/417472

Kilgarriff, A., Rundell, M., & Uì Dhonnchadha, E. (2006). Efficient corpus creation for lexicography: Building the new corpus for Ireland. *Language resources and evaluation* 40, 127–152. doi:10.1007/s10579-006-9011-7

Kuz'mič, V. (2000). *Žgučij glagol – slovar' narodnoj frazeologii* [Burning verb – a dictionary of folk phraseology]. Moscow: Zelenyj vek.

Kuznetsova, J. (2015). *Linguistic profiles: Going from form to meaning via statistics* (= Cognitive Linguistics Research 53). Berlin: De Gruyter Mouton. doi:10.1515/9783110361858

Langacker, R. W. (1987). *Foundations of cognitive grammar*. Vol. I. Theoretical prerequisites. Stanford: Stanford University Press.

Langacker, R. W. (1991a). *Concept, image, and symbol: The cognitive basis of grammar*. Berlin: Mouton de Gruyter. doi:10.1515/9783110857733

Langacker, R. W. (1991b). *Foundations of cognitive grammar*. Vol. II. Descriptive application. Stanford: Stanford University Press.

Langacker, R. W. (2003). Constructional integration, grammaticization, and serial verb constructions. *Language and Linguistics* 4, 251–278.

Letučij, A. B. & Rakhilina, E. V. (2014). Načaľnaja stadija grammatikalizacii značenij glagoľnoj množestvennosti v kvazigrammatičeskix konstrukcijax [Early stage of grammaticalization of the meaning of vebal multiplicity in quasi-grammatical constructions]. In A. Moldovan & S. Tolstaja (Eds.), *Slavjanskoe jazykoznanie. XV Meždunarodnyj s'ezd slavistov. Minsk, 21–27 avgusta 2103 g. Doklady rossijskoj delegacii* [Slavic linguistics. XV International Congress of Slavists. Minsk August 21–27, 2013. Contributions by the Russian participants] (pp. 432–451). Moscow: Indrik.

Lubensky, S. (1995). Random House Russian-English Dictionary of Idioms. New York: Random House.

Lyashevskaya, O. N. (2016). *Korpusnye instrumenty v grammatičeskix issledovanijax russkogo jazyka*. Corpus tools for Russian grammar studies. Moscow: LRC Publishing House.

Lyashevskaya, O., & Kashkin, E. (2015). FrameBank: a database of Russian lexical constructions. In M. Yu. Khachay, N. Konstantinova, A. Panchenko, D. I. Ignatov, & G. V. Labunets (Eds.), Analysis of images, social networks and texts. Fourth international conference, AIST 2015, Yekaterinburg, Russia, April 9–11, 2015, Revised selected papers (= Communications in computer and information science, Vol. 542) (pp. 337–348). Springer.

Lyashevskaya O. N., & Sharoff, S. A. (2009). *Častotnyj slovar' sovremennogo russkogo jazyka (na materiale Nacional'nogo Korpusa Russkogo Jazyka)* [Frequency dictionary of modern Russian (based on the Russian National Corpus)]. Moscow, Azbukovnik.

McWhorter, J. H. (2011). *Linguistic simplicity and complexity. Why do languages undress?* Berlin: De Gruyter Mouton. doi:10.1515/9781934078402

Meurers, D., Ziai, R., Amaral, L., Boyd, A., Dimitrov, A., Metcalf, V., & Ott, N. (2010). Enhancing authentic web pages for language learners. In *Proceedings of the 5th Workshop on Innovative Use of NLP for Building Educational Applications (BEA-5) at NAACL-HLT 2010* (pp. 10–18). Los Angeles.

Mixel'son, M. I. (1896–1912/2004). *Bol'šoj tolkovo-frazeologičeskij slovar'* [Big explanatory phrazeological dictionary]. Moscow: ETS Publishing House.

Ovsjannikova, M. A. & Saj, S. S. (2014). Raznye formy, obščie mexanizmy: aľternacija odno- i dvuxob"ektnoj diatez u russkix glagolov mental'noj sfery [Different forms, common mechanisms: the alternation of single- and double-object diathesis among Russian verbs expressing mental events]. In S. S. Saj, M. A. Ovsjannikova & S. A. Oskoľskaja (Eds.), *ACTA LINGUISTICA PETROPOLITANA. Trudy Instituta lingvističeskix issledovanij RAN [Works of the Institute for Linguistic Research of the Russian Academy of Sciences] Russkij jazyk: grammatika konstrukcij i leksiko-semantičeskie podxody [Russian: construction grammar and lexico-semantic approaches]* (pp. 248–291). St. Petersburg: Nauka.

Rakhilina, E. (2000). *Kognitivnyj analiz predmetnyx imen: Semantika i sočetaemost'* [Cognitive analysis of concrete nouns: Semantics and collocation]. Moskva: Russkie slovari.

Rakhilina, E. (Ed.) (2010). *Lingvistika konstrukcij* [The linguistics of constructions]. Moscow: Azbukovnik.

Rakhilina, E. V. & Majsak, T. A. (2007). Glagoly dviženija i naxoždenija v vode: leksičeskie systemy i semantičeskie parametry (Verbs of motion and position in water: lexical systems and semantic parameters). In T. A. Kajsak & E. V. Rakhilina (Eds.), *Glagoly dviženija v vode: leksičeskaja tipologija* [Verbs of motion in water: lexical typology] (pp. 27–75). Moscow: Indrik.

Rakhilina, E. V., Reznikova, T., & Bonch-Osmolovskaya, A. (2012). Towards a typology of pain predicates. *Linguistics* 50, 421–465.

Rakhilina, E. V. & Letučij, A. B. (2012). Russkie konstrukcii s vremennym značeniem: o granicax nastojaščego vremeni [Russian constructions with temporal meaning: on the borders of present tense]. In V. I. Zabotkina (Ed.), *Präsens: Sbornik naučnyx trudov* [Präsens: Anthology of scientific works] (pp. 224–242). Moscow: OLMA Media Grupp.

Rakhilina, E. V. & Plungian, V. (2013). Time and speed: Where do speed adjectives come from? *Russian linguistics* 37, 347–359. doi:10.1007/s11185-013-9117-7

Reynolds, R. J. (2016). *Russian natural language processing for computer-assisted language learning*. Doctoral Dissertation, UiT The Arctic University of Norway.

Rogožnikova, R. P. (2003). *Tolkovyj slovar' sočetanij, èkvivalentnyx slovu* [Explanatory dictionary of collocations that are equivalent to a word]. Moscow: AST, Astreľ.

Saj, S. S. (2008). *K tipologii antipassivnyx konstrukcij: semantika, pragmatika, sintaksis* [Toward a typology of antipassive constructions: semantics, pragmatics, syntax]. PhD dissertation. St. Petersburg: Russian Academy of Sciences, Institute of Linguistic Research.

Saj, S. S. (2014). Leksičeskie mečanismy grammatičeskogo drejfa: konstrukcii s dativnym sub"ektom v russkom jazyke XVIII–XXI vekov [The lexical mechanisms of grammatical drift: constructions with a dative subject in Russian in the XVIII–XXI centuries]. In S. Ju. Dmitrenko & N. M. Zaika (Eds.), *ACTA LINGUISTICA PETROPOLITANA. Trudy Instituta lingvističeskix issledovanij RAN [Works of the Institute for Linguistic Research of the Russian Academy of Sciences] Studia typologica octogenario Victori Khrakovskij Samuelis filio dedicate* (pp. 568–610). St. Petersburg: Nauka.

Sazonova, I. K. (2008). *Tolkovo-grammatičeskij slovar' russkix pričastij* [Explanatory and grammatical dictionary of Russian participles]. Moscow: AST-PRESS.

Švedova, N. Ju. (1960). *Očerki po sintaksisu russkoj razgovornoj reči* [Outlines of the syntax of conversational Russian]. Moscow: Izdateľstvo Akademii nauk SSSR.

Švedova, N. Ju., Arutjunova, N. D., Bondarko, A. V., Ivanov, V. V., Lopatin, V. V., Uluxanov, I. S., & Filin, F. P. (1980). *Russkaja grammatika. II: Sintaksis* [Russian grammar. II: Syntax]. Moscow: Nauka.

Trudgill. P. (2011). *Sociolinguistic typology. Social determinants of linguistic complexity*. Oxford. Oxford University Press.

Vajjala, S. (2015). *Analyzing Text Complexity and Text Simplification: Connecting Linguistics, Processing and Educational Applications*. PhD thesis, University of Tübingen.

Zaliznjak, A. A. (1980). *Grammatičeskij slovar' russkogo jazyka* [Grammatical dictionary of Russian]. Moscow: Russkij jazyk.

Zolotova, G. A. (2006). *Sintaksičeskij slovar': Repertuar èlementarnyx edinic russkogo sintaksisa* [Syntactic dictionary: The repertoire of elementary units of Russian syntax]. Moscow: Editorial URSS.

Žolkovskij, A. K., & Meľčuk, I. A. (1965). O vozmožnom metode i instrumentax semantičeskogo sinteza [On a possible method and means for semantic synthesis]. *Naučno-texničeskaja informacija* [Scientific and Technical Information], 5, 23–28.

Constructing a constructicon for German

Empirical, theoretical, and methodological issues

Hans C. Boas and Alexander Ziem

University of Texas at Austin / Heinrich-Heine-Universität Düsseldorf

This chapter discusses a number of important issues underlying and motivating the development of a constructicon for German. More specifically, it presents an overview of some typologically interesting facets of German syntax such as word order, topological fields, case, and passives. Taking a contrastive view of some German constructions and their English counterparts, this chapter shows under what circumstances existing entries from the Berkeley constructicon for English can be reused to create corresponding entries in a German constructicon. Of particular interest in this context are the notions of idiomaticity, abstraction, and the continuum of constructional correspondence. Finally, this chapter introduces ongoing constructicographic efforts to create a constructicon for German. To document the current status of the project, both the methodology and the workflow guiding the German Constructicon project (GCon) are illustrated.

Keywords: annotation, constructicon, construction, construction grammar, contrastive linguistics, German

1. Introduction

This paper addresses empirical, theoretical, and methodological issues that arise in the development of a constructicon for German. By discussing a set of grammatical constructions in contemporary German and comparing them to their equivalents in English, we aim at singling out to what extent constructions in German exhibit commonalities but also idiosyncrasies that need to be taken into account when constructing a constructicon for German. On the basis of the results, we propose that the benefit of mapping English constructions, as, for example, documented in the prototype of the Berkeley FrameNet constructicon (see Boas, 2017; Lee-Goldman & Petruck, this volume), to their counterparts in German is limited to a relatively small number of constructions. Other constructions require additional treatments both in terms of their syntactic behavior as well as their grammatical realization patterns and their semantic properties including pragmatic constraints.

The remainder of the paper is structured as follows. Section 2 provides a brief overview of typologically interesting facets of German syntax that distinguishes German from other languages, most notably English, for which there already exists a prototype constructicon (Fillmore, 2008; Fillmore, Lee-Goldman & Rhomieux, 2012; for an overview: Ziem, 2014a). These include word order (Webelhuth, 1992, Kathol, 2000), topological fields (Wöllstein, 2010), the case system (Zifonun, Hoffmann & Strecker, 1997), the passive (Ackerman & Webelhuth, 1998; Lasch, 2016), and (semi-)idiomatic constructions (Oya, 1999; Boas, 2003; Engelberg et al., 2011, Ziem & Staffeldt, 2011), among others (see also Boas & Ziem, in press a; in press b). The goal of this section is to highlight the particularities of a specific range of grammatical phenomena of German that have important consequences for the architecture of a German constructicon, with particular reference to its reliance on the lexical information contained in a German FrameNet.

Section 3 discusses the implications of these characteristics of German grammar to inform and influence the architecture of a German constructicon. To this end, we begin by reviewing insights from research in contrastive linguistics, which has demonstrated significant problems when analyzing grammatical phenomena from a contrastive perspective (James, 1980; Chesterman, 1998; Haspelmath, 2007). The second part of this section puts these insights into the context of cross-linguistic generalizations over constructions by comparing the approaches of Croft (2001) and Boas (2010a). While the former explicitly argues that categories and constructions are language-specific, the latter proposes that it is in fact possible to apply constructions as a *tertium comparationis* for the analysis of particular types of constructions. In this context, we also briefly point to parallel research on creating construction entries for Swedish and Portuguese in parallel to the Berkeley constructicon for English (see Bäckström, Lyngfelt & Sköldberg, 2014; Laviola, 2015; Lyngfelt, Bäckström et al., this volume).

In Section 4, we discuss how the empirical and theoretical insights about the syntax-lexicon continuum should drive the design of a constructicon for German. Building on prior research such as Boas (2014), Ziem (2014a), Ziem, Boas and Ruppenhofer (2014), Ziem and Ellsworth (2016), Boas, Dux and Ziem (2016) and the contributions in Boas and Ziem (in press a), we investigate what types of construction entries from the English constructicon (Fillmore, Lee-Goldman & Rhomieux, 2012) can be reused for creating parallel construction entries for a German constructicon (similar to proposals in Boas (2002) for reusing English semantic frames for other languages). Specifically, we discuss and compare three constructions in German and English, ranging from quasi synonymous and structurally homologous ones, such as the *just_because_doesn't_mean* construction, to constructions with significant language-specific characteristics, such as the *way* construction (Goldberg, 1995; Oya, 1999) and the family of exclamative

constructions (d'Avis, 2013; Michaelis, 2001; Ziem & Ellsworth, 2016).[1] The empirical evidence leads us to propose a "continuum of constructional correspondence" to argue that reusing English construction entries has only limited benefits.[2] We therefore propose a language-specific corpus-based methodology that focuses on the creation of German-specific construction entries by primarily relying on syntactic and semantic categories of German. This approach has the advantage of first providing detailed lexico-syntactic construction entries for German, linking these in larger networks of (families of) constructions. Only at a later point in time, is it feasible to link German construction entries to construction entries of other languages, similar to approaches in rule-based machine translation (Slocum, 1987) and preliminary results from research linking Swedish construction entries with their English counterparts (see Bäckström, Lyngfelt & Sköldberg, 2014).

Having this in mind, Section 5 finally documents the current status of the German Constructicon project hosted at the University of Düsseldorf (http://gsw.phil.uni-duesseldorf.de). Specifically, we introduce the annotation and analysis pipeline that has been created to cope with peculiarities of German constructions (also discussed in Section 2), while at the same time being principally compatible with both the lexicographic FrameNet database and the constructicons of other languages, most notably in English (Petruck & Lee Goldman, this volume), Swedish (cf. Lyngfelt, Bäckström et al., this volume), Brazilian Portuguese (cf. Torrent et al., this volume), Japanese (cf. Ohara, this volume), and Russian (cf. Janda et al., this volume).

2. Typological considerations

The goal of this section is to briefly discuss the particularities of a selected range of grammatical phenomena of German that have important consequences for the architecture of a German constructicon, with particular reference to its reliance on the lexical information contained in a German constructicon.[3]

---

1. Following the style sheet for this volume, frame and construction names are written in a sans serif font (in this case `Concolas`). Please note that typical FrameNet conventions use Courier (New) font for frame names and italicized Courier (New) font for construction names.

2. Our case study presented in the following sections suggests that creating parallel constructicons based on the Berkeley constructicon for English turns out to be much more complicated than finding translation equivalents in lexical FrameNets as described by Padó (2007) and Padó and Lapata (2009).

3. For more detailed descriptions of German grammar, see, e.g., Abraham (1995), Eisenberg and Thieroff (2013), Hentschel and Weydt (2013), and Zifonun, Hoffmann and Strecker (1997). This section is based in part on Boas and Ziem (in press b).

2.1   Word order

We begin with issues related to word order. In contrast to English, which is assumed to be an SVO language, German has often been characterized as an SOV language, i.e. the SOV order is considered to be "basic", while other word orders are derived from this word order (see Bierwisch, 1963; Haider, 1993). Consider the following examples, in which the order of the finite verb differs between SOV (1a), VSO (1b), and OVS (1c).

(1) a.   …dass Fritz den Wein austrinkt.   (SOV)
          …that Fritz the   wine out-drinks
          'that Fritz drinks the wine up.'
     b.   Trinkt Fritz den Wein aus?   (VSO)
          drink   Fritz the   wine out
          'Does Fritz drink the wine up?'
     c.   Den Wein trinkt Fritz aus.   (OVS)
          the   wine drinks Fritz up
          'Fritz drinks the wine up.'

Generative syntactic models such as Government and Binding/Minimalism (Reis, 1980; den Besten, 1983; Webelhuth, 1992), Generalized Phrase Structure Grammar (Jacobs, 1986; Uszkoreit, 1987), Lexical Functional Grammar (Berman, 2003), and Head-Driven Phrase Structure Grammar (Kathol, 2000; Meurers, 2000; Müller, 2005) assume that the "basic" German word order as in (1a), in which the finite verb occurs in the last position (V-L) in sentences introduced by complementizers, serves as the basis for deriving other word order configurations in which the verb occurs in second position (V-2) such as in (1c) (see Müller 2005 for details). Similarly, generative theories account for other differences in word order by assuming a basic underlying SOV word order in order to derive specific word orders such as those for infinitives (Haider, 1986; von Stechow & Sternefeld, 1988), left dislocation (Haider, 1990), topicalization (Fanselow, 1989; Haider, 1990), passives (Grewendorf, 1989), and relative clauses (Haider, 1985; Rimsdijk, 1985).

From the viewpoint of Construction Grammar, the assumption that one word order is more basic than others and should therefore serve as the basis for deriving other types of word orders is rather difficult to maintain, since there are no a priori empirical criteria for determining what types of constructions are more basic than others, or what types of constructions should be derived from "basic" constructions (see Fillmore & Kay, 1993; Croft, 2001; Goldberg, 2006). In other words, the constructional view holds that (1) there are no constructions which are necessarily more basic than other constructions (though prototype effects may yield

similar observations; see Lakoff, 1987; Goldberg, 1995), and (2) constructions are organized in networks with inheritance hierarchies in which related constructions inherit information from each other (Goldberg, 1995; Fillmore, 1999; Boas, 2011; Michaelis, 2012; Sag, 2012; Ziem & Lasch, 2013, pp. 95–102). We return to this point below when we discuss some basic procedures for identifying, classifying, and capturing different types of constructions in German.

More traditional approaches to German syntax employ the so-called topological fields model to classify the basic clause types of German based on the position of the finite verb, among other factors (for details, see Eisenberg, 2006, pp. 394–420; Eisenberg & Gallmann, 2016, pp. 871–899; Imo, 2016, pp. 199–226). We briefly review some of the key insights of this model before showing how some of them can be integrated into a constructional approach to German syntax. The topological fields model captures generalizations about the position of the finite verb by employing different sets of so-called fields and brackets, as the following figure illustrates:

Prefield	Left Sentence Bracket	Middle Field	Right Sentence Bracket	Final Field

Figure 1. Topological fields[4]

On this view, the clause is structured around a left bracket ("linke Satzklammer" = "LS"), which hosts the verb in either initial or second position and a right bracket ("rechte Satzklammer" = "RS"), which is the position taken by clause-final verbs (finite and non-finite) and verbal particles (Höhle, 1986). The left and right brackets are used to define structural positions, so-called fields: The position to the left of the LS is the so-called prefield ("Vorfeld"), which can host only one constituent with varying degrees of complexity. The prefield remains empty in a variety of sentences, such as in subordinate clauses, verb-first sentences, and yes-no questions. The left bracket contains either the finite verb or a subordinating conjunction and may only be left empty in a few select instances such as special cases of relative clauses, infinitival clauses, and an embedded constituent question (see Reis, 1985; Wöllstein-Leisten et al., 1997).

The position between the LS and the RS is the so-called middle field ("Mittelfeld") and the position to the right of the RS is the so-called final field ("Nachfeld"). According to Wöllstein-Leisten et al. (1997), the middle field can host a potentially unlimited number of constituents of various types, each of which have an internal

---

4. Some accounts also assume a so-called "pre-prefield" ('Vorvorfeld') and "final final field" ('Nachnachfeld'). For the sake of simplicity, we do not include these additional fields here and in the following explanations.

structure of their own (e.g. they can also be clauses). When dealing with complex predicates, the right sentence bracket hosts all non-finite verbal parts. In the case of subordinate sentences, the finite verb also appears in this position. The final field typically contains constituents of subject, object, adverbial, and relative clauses.[5]

According to the topological fields model, different types of elements (which themselves can have internal structure of their own) can occur in different fields, thereby covering the three types of sentence patterns, characterized in terms of the position of the finite verb, in German, as Figure 2 shows, in which items in italics are obligatory.

	Prefield	Left Bracket	Middle Field	Right Bracket	Final Field
V-1		*Finite verb*	Constituents	Inf. V	Constituents
V-2	*Constituent*	*Finite verb*	Constituents	Inf. V	Constituents
V-L		*Conjunction*	Constituents	Inf. V *finite Verb*	Constituents

Figure 2. Three sentence types according to position of the finite verb. Items in italics are obligatory (see Wöllstein-Leisten et al., 1997, p. 54)[6]

The information in Figure 2 is a generalization over a multitude of different sentence types (declarative, imperative, interrogative, etc.) defined by the position of the finite verb, i.e. V-1, V-2, and V-L sentences. In fact, Wöllstein-Leisten et al. (1997, p. 55) list a total of 28 different types of sentence templates depending on different combinations, configurations, and positions of the finite verb and other constituents in the topological fields model.[7] For the purpose of designing and building a German constructicon we propose to adopt the basic insights of the topological fields model. On this view, each of the 28 sentence templates can be regarded as part of the form of a construction (we leave aside other issues regarding the form of

---

5. For further details about the different types of constituents occurring in the various positions of German sentences in the topological fields model, see Lenerz (1977), Bech (1983), Höhle (1986), Reis (1987), Abraham (1995), and Wöllstein-Leisten et al. (1997).

6. Elements in italics are obligatory while other elements are optional. Depending on the verb, a subject and different types of objects may also be obligatory or optional, which directly influences the number and ordering of elements in the middle fields and final fields. See Wöllstein-Leisten et al. (1997) and Welke (2011) for more details. In a constructicon that adopts key insights from the topological fields model it will thus also be necessary to determine how lexical entries of words (specifically verbs) interact with different types of constructions, i.e. under what circumstances particular verbs may fuse with constructions (for details see Boas, 2008, 2011).

7. Space limitations prevent us here from going into any further detail about the 28 different configurations of constituents according to the topological fields model.

constructions such as intonation), and could thus serve as the basis for an inventory of German word order constructions that could eventually be organized in terms of a network of constructions with inheritance relations (see Ackerman & Webelhuth, 1998; Sag, 2012). Following the concept that constructions are pairings of form with meaning also requires addressing the meaning side of each of the 28 constructions (and others) in a systematic fashion. We return to related issues in Section 4 below, where we discuss some features of a constructicon of German.

So far, we addressed only syntactic ordering (focused primarily on the position of the finite verbs) as a particularly interesting phenomenon in German syntax. We now turn to pragmatic ordering, which orders sentence constituents not only based on syntactic ordering mechanisms, but also because of the role and function they play in communication. That is, the order of constituents in a sentence may depend on the specific circumstances in which the sentence is uttered, e.g. on the particular emphasis required, on what has been said before, and so on (Fox, 1990, p. 251). Consider, for example, the sentences in (2a)–(2d), which contain the same constituents, but ordered in different ways.

(2) a. Der Mann hat dem Jungen gestern den Ball gegeben.   (subject)
     b. Den Ball hat der Mann dem Jungen gestern gegeben.   (direct object)
     c. Dem Jungen hat der Mann gestern den Ball gegeben.   (indirect object)
     d. Gestern hat der Mann dem Jungen den Ball gegeben.   (adjunct)
         'Yesterday, the man gave the ball to the boy.'

The examples above show that the prefield position can host different elements: the subject, the direct object, the indirect object, and an adjunct. The ordering is based on the communicative function that the speaker intends to encode, depending on the context and depending on what is already known (and what is not known) by the hearer.[8] Typically, animate NPs tend to precede inanimate ones, short constituents (like pronouns) tend to occur before longer ones, and given information precedes new information (Behaghel, 1930).

For example, depending on the question that has been asked, such as *Who gave the boy the ball?, When did the man give the boy the ball?*, or *What was going on?*, the speaker will likely prefer one of the pragmatic orderings in (2) above over the others.[9] An additional factor complicating the choice and interpretation of different pragmatic orders is the nucleus of the intonation pattern that can be moved

---

8. Acceptability judgements may vary depending on a speaker's background.

9. Note that most likely, a speaker will reply to one of these questions in natural discourse by just answering *Der Mann.* ('the man') or *Gestern.* ('yesterday'), leaving out the rest of the information. We thank Bernhard Ost for pointing this out to us.

around in each of the sentences in (2), thereby achieving different interpretations depending on the communicative context (for details see Lenerz, 1977; Höhle, 1982; Eroms, 1986; Fox, 1990). In a constructional approach leading to the creation of a constructicon for German, these different intonation patterns will also require a careful analysis as a part of the form side of each individual construction. This entails that we will most likely have to identify and classify the full range of intonation patterns as a part of the form side of the construction entries for each of the 28 different constructional patterns pointed out above. With this short overview of German word order we now turn to another important issue, namely the German case system.

2.2   Case

Unlike most other Germanic languages, German has a relatively elaborate case system consisting of four cases (nominative, accusative, dative, and genitive), all of which may be used to inflect nouns, adjectives, pronouns, and determiners. Case is either assigned structurally (configurationally), i.e. to identify the grammatical functions such as subject (nominative), direct object (accusative), or indirect object (dative) in a sentence. The system of four cases allows German to encode a variety of grammatical functions in many different word order combinations, giving rise to a much more flexible (close to free) word order when compared with other languages such as English, which has a relatively fixed word order because of the almost complete absence of an overt case marking system (except for the pronouns) (for details see Kirkwood, 1969; Hawkins, 1986; Barðdal, 2013; Fischer, 2013). Case is also assigned lexically by verbs and prepositions (see Engel, 1988; Zifonun, Hoffmann & Strecker, 1997). As can be seen in (3), the paradigm of German case marking on NPs is quite extensive, involving number and gender. The NPs in (3) differ in number: those in (a) are singular, those in (b) are plural. The first row shows nominative marking, the second row accusative marking, the third row genitive marking, and the fourth row dative marking (the first column contains masculine nouns, the second column feminine nouns, and the third column neuter nouns).

(3) a.   der gute Mann   die gute Frau   das gute Kind
         den guten Mann   die gute Frau   das gute Kind
         des guten Mannes   der guten Frau   des guten Kindes
         dem guten Mann   der guten Frau   dem guten Kind
     b.   die guten Männer   die guten Frauen   die guten Kinder
         die guten Männer   die guten Frauen   die guten Kinder
         der guten Männer   der guten Frauen   der guten Kinder
         den guten Männern   den guten Frauen   den guten Kindern

Each of the case markers in (3) can be regarded as their own constructions, combining a specific form with a specific meaning. For example, the sequence [$_{<Nom-sing-masc>}$*der*, [Adj]-*e*, [N]-Ø] is the form side of a nominative singular masculine NP construction which specifies three elements: the determiner *der* ('the'), an adjective with an ending in −*e*, and a noun with no marker. The meaning-function side of the construction is typically that of Agent (subject) or some semantically more specific instantiation of Agent, depending on the semantic frames evoked by the noun (and verb in the same sentence) (see Van Valin and Wilkins, 1996; Boas, 2010c).[10] In contrast, the form side of the accusative case marking construction for singular masculine nouns is [$_{<Acc-sing-masc>}$*den*, [Adj]-*en*, [N]-Ø], while the meaning-function side is typically that of a Patient (direct object) or some specific semantic instantiation of it.

Of course, case in German has many more facets than what we discussed above (for more details, see Zifonun, Hoffmann & Strecker, 1997). At this point, however, we hope to have shown that a constructional approach to case in German requires a great number of case-marking constructions that apply to determiners, adjectives, and nouns, and that case is typically assigned structurally depending on the grammatical function of a NP in a sentence or it is assigned based on the (lexical) properties of particular verbs and prepositions that govern specific cases; for more details on how case can be analyzed in a constructional approach see Barðdal (2006, 2008, 2009). When designing a constructicon for German it is necessary to take account of all these constructions peculiar to German.

2.3 Constructions at different levels of abstraction

So far, we addressed only two types of constructions that are (almost) completely regular and that typically come without any significant restrictions. While word order constructions are fairly abstract – their meanings encode relatively high-level schematic meanings such as declarative, interrogative, or imperative semantics – they are also rather complex when it comes to the number of slots and constituents involved in each construction. Similarly, case marking constructions are regular and predictable because they attach to particular determiners, adjectives, and nouns only in specific contexts. They differ from word order constructions in that they encode relatively specific meanings such as Agent (nominative), Patient (accusative), and Beneficiary (dative). In terms of Goldberg's (2006) typology of constructions, we are dealing with a group of (word order) constructions that are

---

10. Note that the nominative in German has different types of functions, for details see Sommerfeldt and Starke (1992, pp. 103–104).

relatively schematic and a group of constructions (case marking) that are much more specific, of which each member consists of groups of morphemes attached to determiners, adjectives, and nouns.

We now turn to a brief discussion of some other types of constructions in German that differ from the two families of constructions above in terms of complexity and level of abstraction (for a more detailed discussion, see Boas, 2014). Our cursory discussion is intended to show that German has roughly the same types of constructions as those discussed by Goldberg (2006, p. 5) for English, including high-level abstract constructions, meaningful argument structure constructions, partially filled idioms, idioms, words, and morphemes.[11] What unifies all constructions is their common architecture, i.e. they are form-meaning pairings as Figure 3 from Croft (2001) shows.
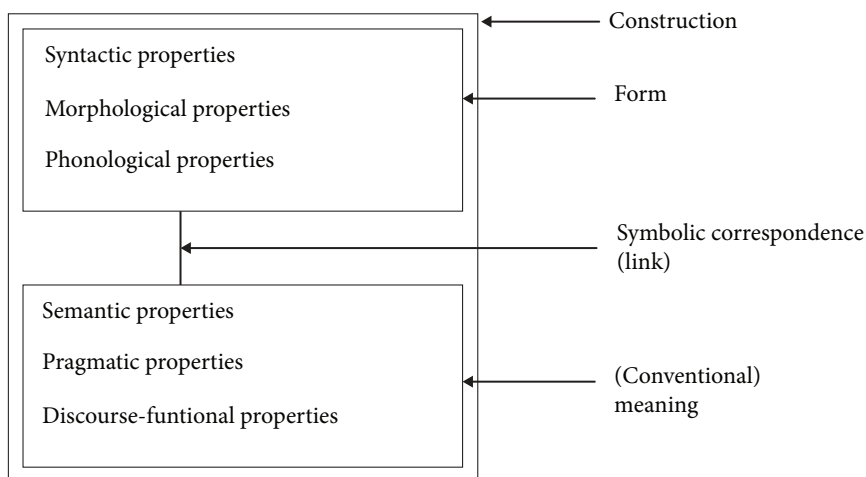
Figure 3. The symbolic structure of a construction according to Croft (2001, p. 18)

---

11. There is some disagreement among construction grammarians whether all constructions have meaning. For example, Goldberg (2006) proposes that the subject-auxiliary inversion construction in English is meaningful and motivated, whereas Fillmore (1999) argues for an abstract `auxiliary_inversion` construction that does not involve any significant meaning component(s). More explicitly, Fillmore, Lee-Goldman and Rhomieux (2012: Section 3) argue in a more recent paper that there are indeed "constructions without meaning", such as the so-called `gapping` construction or the `shared_completion` construction, among others.

Just like English, German has a `subject_predicate` construction which ensures that the subject and the predicate agree in number, as the following examples illustrate.

(4) a. Peter gibt seiner Tochter einen Kuss.
'Peter gives his daughter a kiss.'
b. Laura backt Bob einen Kuchen.
'Laura bakes Bob a cake.'

While the form side of the construction is straightforward (the two daughters of the construction, the NP and VP, need to agree in number), the meaning side of the `subject_predicate` construction is not that easy to identify because it is rather abstract. In terms of level of abstraction it is thus fair to say that the `subject_predicate` construction in German is more abstract than the different types of word-order constructions discussed in Section 2.1 above, which encode more concrete meanings such as declarative, interrogative, imperative, etc. For this reason, Fillmore, Lee-Goldman and Rhomieux (2012: Section 3) suggest that the `subject_predicate` construction falls under the category "constructions without meaning". However, even though its meaning is neither concrete nor non-transparent, we assume that the `subject_predicate` construction bears some type of (minimal) meaning.[12]

The `passive` in German constitutes another interesting family of constructions, because the different constructions used to express passive in German differ not only in their form aspects, but also in their meaning aspects since they all differ slightly from each other. This is one major aspect in which the `passive` in German differs from the `passive` in English. Ackerman and Webelhuth (1998) present an extensive account of 14 different passive and passive-like constructions in German, which are all related to each other in a constructional network (using HPSG-style inheritance hierarchies), and which differ from each other in their syntactic-semantic properties, as Figure 4 illustrates (do = direct object; io = indirect object). Each of the fourteen different passive constructions is a combination of specific features (e.g. P1, *german-short-pers-werden-pas-lci* combines the features "short" and "werden").

---

12. There are some combinatorial restrictions, however, these come from the predicate *sleep* (instantiated in the `subject_predicate` construction), which requires an animate subject, unless metaphorical extensions are intended.

Passive		do	io	imp	long	short	modal	werden	bek.	sein
P1	german-short-pers-werden-pas-lci					✓		✓		
P2	german-long-pers-werden-pas-lci	✓			✓			✓		
P3	german-short-imp-zuinf-pas-lci	✓	✓			✓		✓		
P4	german-long-imp-zuinf-pas-lci		✓	✓				✓		
P5	german-short-pers-zunif-pas-lci	✓				✓	✓			✓
P6	german-long-pers-zuinf-pas-lci	✓			✓		✓			✓
P7	german-short-imp-zuinf-pas-lci		✓			✓	✓			✓
P8	german-long-imp-zuinf-pas-lci		✓	✓			✓			✓
P9	german-short-pers-bekommen-pas-lci		✓			✓			✓	
P10	german-long-pers-bekommen-pas-lci		✓		✓				✓	
P11	german-short-attrpart-pas-lci	✓				✓				
P12	german-long-attrpart-pas-lci	✓			✓					
P13	german-short-attrzuinf-pas-lci	✓				✓	✓			
P14	german-short-attrzuinf-pas-lci	✓			✓		✓			

Figure 4. 14 German passive constructions and their properties
(Ackerman & Webelhuth, 1998, p. 238)[13]

Using inheritance hierarchies, Ackerman & Webelhuth (1998) show that even though the semantic and syntactic properties of the 14 different passive constructions in German differ from each other, it is nevertheless possible to systematically identify certain characteristics shared by all constructions. This allows them to state an inventory of Lexical Combinatorial Items (LCIs, similar to grammatical constructions that combine form with meaning) that they arrange in a hierarchical network of constructions that inherits properties from even more abstract types of constructions. Consider, for example, Figure 4, which in the box at the top contains a set of abstract LCIs (comparable to constructions consisting of form-meaning pairings) from which the highest-level passive LCIs inherit their information. Thus, the `german-pred-pas-lci` is the top-level `passive` construction from which other lower-level `passive` constructions such as `german-pred-zuinf-pas-lci,` `german-bekommen-pas-lci,` and `german-werden-pas-lci` inherit information, which in turn are the mother constructions from which the concrete passive constructions in Figure 4 above inherit information.

---

13. The labels in the top row represent the properties of each of the 14 passive constructions: direct object, indirect object, impersonal, long, short, modal, *werden*, *bekommen*, and *sein*.
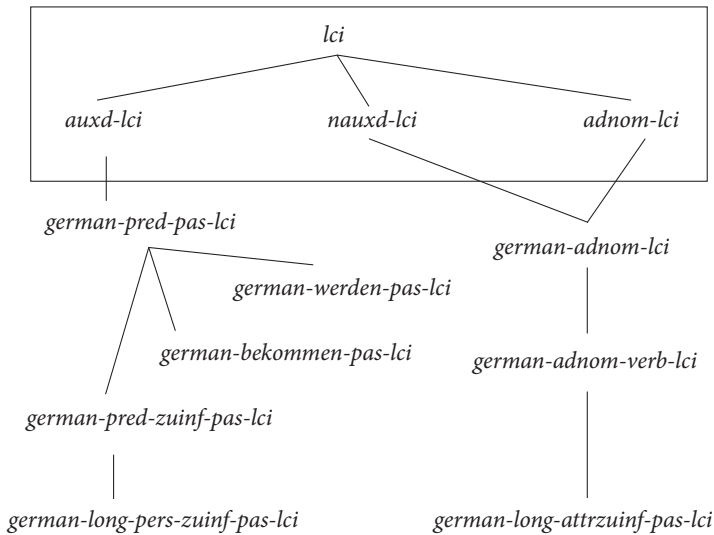
Figure 5. Network of passive constructions (lexical combinatorial items; LCIs) (Ackerman & Webelhuth, 1998, p. 248)

Ackerman and Webelhuth (1998, p. 248) characterize the advantages of such a network approach as follows: "By systematically extending these two type hierarchies in accordance with the demands of empirical data, it becomes possible to capture all the generalizations, sub-generalizations, and idiosyncrasies of the German passives." With this short overview of how the different passive constructions in German can be analyzed using a constructional network, we turn our attention to a different family of constructions, namely argument structure constructions (see also Lasch, 2016, for a constructional analysis of German constructions without agents, including passive constructions).

Goldberg's (1995) seminal work presents a number of similar analyses of a variety of so-called argument structure constructions (ASCs), which are independently existing meaningful constructions that are capable of fusing with lexical entries of verbs to provide them with extra meaning and hence with extra arguments at the syntactic level. One of the main motivations behind this approach is the wish to avoid implausible verb senses, such as in *Joe cooked Mary a meal* where *to cook* has an extra sense expressing a beneficiary receiving a theme from an agent. The solution, according to Goldberg (1995), is to propose an independently existing `ditransitive` construction that fuses with the lexical entry of *to bake*, thereby providing it with extra semantics and hence additional arguments. While the `ditransitive` construction has a fairly straightforward counterpart in German (cf. *Josef kochte Maria ein Essen*, 'Josef cooked Maria a meal') and appears to be fairly

productive, not all of the constructions discussed by Goldberg are equally productive. For example, Boas (2003, 2011) discusses the `resultative` construction (*Joe hammered the metal flat*; *Sue laughed herself silly*) in English and German, showing that an independently existing meaningful `resultative` construction is problematic because the restrictions placed on the fusion of the `resultative` construction with lexical entries are not sufficient. Based on several thousand corpus examples, Boas (2003, 2011) argues that the English `resultative` is in fact a network of so-called mini-constructions that are conventionalized form-meaning pairings at the level of verb senses (lexical unit, cf. Cruse, 1986). On this view, individual senses of verbs, which combine particular aspects of form (providing restrictions on phrase type and collocations) with particular aspects of meaning (such as discourse function, perspective, general pragmatic constraints), place their own restrictions on what types of resultative phrases and postverbal objects can combine with particular verbs and verb senses, respectively. What appears to look like an independent `resultative` construction turns out to be an epiphenomenon due to high type and token frequency (for a detailed discussion, see Boas, 2003).

When comparing English resultatives with their German counterparts, Boas (2003, 2011) demonstrates that the German counterparts exhibit some of the same properties, but also many other properties that are attributed to differences in the various polysemy networks[14] and conventionalized verb senses in the two languages. One example is the verb *to drive*, which has many different German counterparts depending on the context: *fahren*, *treiben*, and *befördern*, among others. While English *drive* appears with a variety of different resultatives depending on the context ('Joe drives Mary to town'; 'Joe drives Mary up the wall'; 'Joe drives the nail into the door'), German requires a different verb for each of the senses/contexts, where each (sense of the) verb has its own specific semantic, pragmatic, and syntactic restrictions. This means that resultatives in German are, in principle, very similar to their English counterparts, but the exact specifications on the postverbal elements are language-specific and conventionalized and as such they need to be accounted for in terms of mini-constructions organized in a hierarchical network.

The analysis in Boas (2011) suggests that the kind of abstract meaningful constructions postulated by Goldberg (1995) are in fact compatible with the types of mini-constructions proposed by Boas (2003); here they are conceptualized in terms of a constructional network in which the abstract construction is at the very top of the network, with intermediate levels of abstraction and specification, while the mini-constructions specifying the many idiosyncratic collocational restrictions are found at the bottom of the network. On this view, Goldberg's abstract `resultative`

---

14. See also Fillmore and Atkins (2000), discussing the polysemy of *to crawl*.

construction is thought to be activated when a resultative is interpreted (decoding idiom), while Boas' concrete mini-constructions are involved when producing a `resultative` construction (encoding idiom).

Another interesting ASC is the English *way* construction (e.g. *They laughed their way off the stage*), which comes with relatively few restrictions on the types of verbs that can fuse with it, as long as the resulting sentence can be interpreted as motion involving the main verb (means or manner) (Jackendoff, 1990; Goldberg, 1995). Unlike the `resultative` construction, however, the *way* construction does not have clear counterparts in German. Thus, as we will show in Section 4, it is not helpful to develop parallel entries for German. Our discussion of parallels and differences between English constructions and their German counterparts have important implications for the design of a constructicon of German, to which we turn in Section 4. Before that, we briefly discuss a few methodological issues regarding the re-usability of English constructions for the description of constructions in German.

3.   Contrastive issues

Given the differences and similarities between English constructions and their German counterparts discussed so far, what are the implications for the potential design of a German constructicon? This is an important issue because there are different views of how constructions can be compared across languages and whether insights about the nature of a particular construction in one language can also be applied to a similar type of construction in another language.

One major approach is Radical Construction Grammar by Croft (2001, 2013), who argues, contra many claims in the generative literature (Chomsky, 1981; 1995; Bresnan, 1982; Pollard & Sag, 1993; among others), that categories and constructions are language-specific and can therefore not be used to analyze the inventories of languages across the board. Croft shows that the distributional method applied by most generative accounts defines syntactic categories in terms of their possibility of filling certain roles in grammatical constructions. A comparison of a wide array of cross-linguistic data leads Croft (2001, p. 6) to propose that constructions are the basic units of syntactic representation, and that constructions are themselves language-specific. As such, Croft appears to be skeptical that cross-linguistic generalizations of the types proposed by generative frameworks are possible. Croft's non-reductionist concept of language regards categories as defined in terms of the constructions in which they occur, and as such "valid cross-linguistic generalizations are generalizations about how function is encoded in linguistic form" (2001, p. 363).

While Croft's (2001) Radical Construction Grammar seems to suggest that cross-linguistic generalizations are difficult to obtain unless they are generalizations about how function is encoded in linguistic form, there are other constructional approaches showing that some limited cross-linguistic generalizations are in fact possible when comparing pairs of languages with each other. This idea does not only rest on the insights from contrastive linguistics (James, 1980; Chesterman, 1998; Haspelmath, 2007), but it is also expressed by the founding fathers of Berkeley Construction Grammar (see Fillmore, 2013), who, when discussing a constructional analysis of English, make the following observation with respect to extending constructional insights from English to other languages: "We will be happy if we find that a framework that seemed to work for the first language we examine also performs well in representing grammatical knowledge in other languages" (Fillmore & Kay, 1993, pp. 4–5).

These ideas are developed further in a series of publications by Boas (2002, 2003, 2005, 2009b, 2010a, 2014) and Iwata (2008), which investigate how semantic frames and grammatical constructions from English can be used to analyze other languages such as German and Japanese.[15] The detailed contrastive analyses show that it is indeed possible to use semantic frames and grammatical constructions from English as a starting point for the description and analysis of semantic frames and grammatical constructions in other languages (see also Ziem, 2014b). We begin our discussion with semantic frames and will turn to constructions further down. Boas (2002, 2005) demonstrates that the `Motion` and `Communication` frames of English as described in FrameNet (https://framenet.icsi.berkeley.edu; Fillmore & Baker, 2010; Ruppenhofer et al., 2010; Ruppenhofer, Boas & Baker, 2013) can be adopted straightforwardly for the description and analysis of the vocabulary of the `Motion` and `Communication` frames in German. Subsequently, the papers in Boas (2009a) build on this insight and show that this approach is also in principle applicable for other languages such as French, Japanese, Hebrew, and Spanish, with some minor typological exceptions. More recent efforts by other research teams to use English FrameNet frames for constructing FrameNets for other languages include FrameNets for Swedish (Borin et al., 2010), Brazilian Portuguese (Salomão et al., 2013), and Russian (see Janda et al., this volume).

Extending this contrastive approach to grammatical constructions, however, appears to be more challenging. Building on earlier contrastive research on the `resultative` construction in English and German (Boas 2003), each of the papers in Boas (2010b) investigate how English constructions such as the `ditransitive`, the `resultative`, the `caused-motion`, the `comparative`, and the `conditional`

---

15. See also the contributions in Boas and Gonzalvez-Garcia (2014), which apply a similar methodology to the contrastive analysis of various Romance languages.

are realized in a variety of languages including Japanese, Swedish, Thai, Spanish, Finnish, and Russian. The papers show that there are typological differences in what types of English constructions can be used for the description and analysis of similar constructions in other languages. While sometimes there are straightforward counterparts in which the semantics of an English construction finds a direct equivalent in a different language (but with different specifications on the form side), including its restrictions, there are many cases of divergence in which the English construction requires a set of different constructions in another language.

Similar insights have emerged from more recent applied research on the architecture of so-called constructicons, which are constructional extensions to existing FrameNet projects for different languages. Building on original research by Fillmore (2008) and Fillmore, Lee-Goldman and Rhomieux (2012), who discuss the design and implementation of a constructicon for English, research groups have started building constructicons for Swedish (Lyngfelt et al., 2012), Japanese (Ohara, 2014), Brazilian Portuguese (Torrent et al., 2014), and Russian (Janda et al., this volume); additionally, there are also increasing efforts to create a German constructicon (for more details cf. Section 5). In contrast to the FrameNets for other languages than English, which reused the English frames, these constructicon projects do not primarily reuse the constructions from the English constructicon. Instead, they rely on language-internal resources to find, describe, and analyze the constructions found in their languages in order to then write construction entries. More recently, Bäckström, Lyngfelt, and Sköldberg (2014) explored how information from English construction entries can be used to create corresponding entries in a Swedish constructicon. Similarly, Lyngfelt, Torrent et al. (this volume) discuss interlingual relations between constructions, specifically between English, Swedish, and Brazilian Portuguese construction entries (see also Laviola, 2015). Their investigation shows how some constructions can be aligned with their equivalent constructions in other languages, and how at a practical level linking construction entries also involves linkability between resources (i.e. constructicons).

In the following section, we take an alternative approach by exploring how three different families of constructions from the Berkeley constructicon for English can be reused in one or the other way for the description and analysis of German constructions. In contrast to the language-internal strategy pursued by the constructicon projects for Swedish, Japanese, and Brazilian Portuguese, we are thus interested in exploring what types of information from English constructions can be directly reused for developing comparable construction entries for German. The results of our discussion form the basis for outlining a contrastive methodology that relies on both (1) a contrastive extension of English construction entries to German, and (2) language-internal analysis and writing of construction entries in cases in which the contrastive approach is not fruitful (see Boas, 2014).

4.   The 'continuum of constructional correspondences': Consequences for the design of a German constructicon

In the remainder of this paper, we demonstrate how the empirical and theoretical insights about the syntax-lexicon continuum should drive the design of a constructicon for German. Building on findings in previous works (e.g., Boas, 2014; Ziem, 2014a; Ziem, Boas & Ruppenhofer, 2014; Ziem & Ellsworth, 2016), we specifically investigate what types of construction entries from the English constructicon (Fillmore, Lee-Goldman & Rhomieux, 2012; Boas, 2017) can be reused for creating parallel construction entries for a German constructicon. This is similar to proposals in Boas (2002) for reusing English semantic frames for other languages (see also Lyngfelt, Torrent et al., this volume). Specifically, we discuss and compare three types of constructions in German and English, ranging from quasi synonymous and structurally homologous ones, such as the `just_because_doesn't_mean` construction, to constructions with more or less language-specific characteristics, such as the family of `exclamative` constructions (d'Avis, 2013; Michaelis, 2001; Ziem & Ellsworth, 2016) and the `way` construction (Goldberg, 1995, pp. 199–218; Oya, 1999).[16] The `way` construction, as we shall see, is interesting since there indeed is a counterpart in German which, however, substantially differs from the English `way` construction. Even further to the far end of the 'continuum of constructional correspondences' is the English `progressive be-present-participle` construction that entirely lacks a German counterpart. We do not discuss such constructions, however, simply because they only offer little, if anything, for cross-linguistic mappings.

The advantages of reusing existing resources for building a German constructicon are numerous. In the case of German, however, the possible resources are limited. So far, the most elaborated repository anchored in a constructionist framework is the frame-based database created by the SALSA project (Burchardt et al., 2009).[17] Just like the Berkeley FrameNet project, however, the resources concentrate on valence-bearing linguistic expressions. Thus, they are first and foremost lexical resources ignoring to a large extent constructional information beyond the word level (for an overview cf. Ziem, 2014a). To this end, Fillmore (2013, p. 17) observes

---

16. Note that to date only the `way` construction is included in the current Berkeley FrameNet constructicon. However, this database merely documents the results achieved in a two-year pilot project. The database will be supplemented by many more constructions in the future. Indeed, there are many English constructions that have not made their way into the database even though they are already analyzed in detail.

17. More recently, collaborators of the German Frame-based Online Lexicon (G-FOL) (http://coerll.utexas.edu/frames/) at the University of Texas at Austin have begun compiling lexical entries of German verbs, nouns, adjectives, and adverbs for learners of German (see Boas & Dux, 2013; Boas, Dux & Ziem, 2016).

that a full account of the linguistic structures a sentence instantiates requires not only information about the syntactic and semantic valency of each of the words constituting a sentence, but also information about the grammatical constructions that have meanings and functions on their own. Indeed, many syntactic and semantic structures, such as (semi-)idiomatic constructions (for example `just_because_does_not_mean`, see Section 4.1), constructions on the sentence level (e.g. `exclamative` constructions, see Section 4.2) and even argument structure constructions such as the `way` construction (see Section 4.3) "cannot be fully explained in terms of the kind of structures recognized in FN's [= FrameNet's] annotation database, or simple conjoinings or embeddings of these" (Fillmore, Lee-Goldman & Rhomieux 2012, p. 312).

As we pointed out above, it is worthwhile noting that there is a continuum between lexicon and grammar, allowing the use of the same formalisms and annotation criteria for both frame-bearing words and grammatical constructions. More specifically, Fillmore demonstrates how to integrate the latter into the FrameNet database (Fillmore, 2008; Fillmore, Lee-Goldman & Rhomieux, 2012). Since we use these formalisms in a slightly simplified way for creating a constructicon for German, we briefly introduce the most important annotation categories before turning to three types of constructions illustrating the continuum of (non-)correspondences between English constructions and their German counterparts.

We begin with the linguistic unit evoking a construction, which is called a 'Construction Evoking Element' (CEE).[18] To illustrate, consider (5), an instantiation of an `exclamative` construction (Ziem & Ellsworth, 2016), discussed in detail in Section 4.2.

(5) Was   für ein   spektakulärer Blick von der   Stadt!
     what for a   spectacular   view of   the   city
     'What a spectacular view of the city!'

In (5), the pronoun *was* ('what') serves as the CEE. The complete expression, the so-called construct licensed by the `exclamative` construction, comprises the scope of the surprise conveyed by the `exclamative` construction. Since the meaning of

---

18. Note that Fillmore, Lee-Goldman and Rhomieux (2012) argue that there does not always have to be a CEE that evokes a construction. We think that this is, to some extent, problematic. We see in this the possible danger of postulating empty elements in parallel to empty categories and invisible traces in generative grammar, an issue that constructionist approaches seek to avoid in the first place. Part of the problem seems to be that a CEE is usually conceptualized as a fixed lexical element, or a fixed string of words, bound to a construction. However, we are convinced that a CEE can also be structural in nature. This is true, for example, for most abstract constructions such as the `transitive` construction and the `subject_predicate` construction.

the construction is determined by the `Experience_obj` frame, its Constructional Elements (CEs) can also be annotated with recourse to the FEs constituting the `Experience_obj` frame. Specifically, the scope of the surprise equates with the frame element STIMULUS. Hence, CEs can be defined as those constituents of a syntagmatically complex linguistic structure that instantiate parts of a construction.

Constructional annotations help describe and define a construction appropriately. To this end, the CEE is identified in the first place. In contrast to frame annotations, a target LU providing a link to the construction is often missing. We then name those parts of sentences that form the constituents of the constructs licensed by the construction. Finally, these components are labeled as elements of the construction. Following this procedure, (6) exemplifies the annotation of (5) regarding (a) the CEE, (b) the CEs and their functions within the construction, and (c) the construct that is licensed by the construction. Following FrameNet annotation conventions, we tag CEs with square brackets and constructs with curly brackets, while labeling the meanings or functions of these elements with the help of subscripts.

(6)   {[$_{CEE}$<What>] a [$_{STIMULUS}$spectacular view of the city]!}.

(6) does not yet include annotations of the grammatical functions and phrase types of each of the CEs (if applicable). In line with the descriptions of the respective FEs in the `Experience_obj` frame, the CE STIMULUS realized in (5) can be defined as follows: STIMULUS is the event or entity which brings about the emotional or psychological state – that is, surprise in the case of a `exclamative` construction – of the EXPERIENCER.

Overall, there is a plethora of information that goes into a constructional entry in a German constructicon. Full descriptions of grammatical constructions should include, but are not limited to the following:

– lists of the construction-evoking elements (CEEs)
– descriptions of the construction's lexical head, if applicable,
– descriptions of constructional elements (CEs), including the function of each CE within a construction as well as the phrase types in which each CE may be realized,
– illustrations and descriptions of the realization patterns of a construction
– reports on pragmatic, semantic, and syntactic constraints (preemption)
– explanations of collostructional preferences for each CE, if applicable,
– explanations of covariational preferences of CEs, if applicable,
– annotated sample sentences illustrating the range of realization patterns
– definitions of both form- and meaning-related relations connecting a construction to other constructions in the constructicon.

Clearly, providing all information for each grammatical construction in German is a very ambitious endeavor. Describing and explaining collostructional preferences, for example, requires extensive corpus studies for each target construction. However, even though setting up a constructicon for German is undeniably a large-scale project, it helps to consider it a collaborative work in progress. In the first place, there may be missing pieces of information but these gaps can be filled once validated empirical data are available. As we will see in the next sections, even in the case of well-documented constructions, not all information required for a full construction entry is available. For the sake of usability, it is important to bear in mind that only a complete constructional entry meets the requirement for capturing what a language user needs to know in order to use and understand a grammatical construction appropriately.

4.1 The `just_because_doesn't_mean` construction: exemplifying one end of the 'continuum of constructional correspondences'

Unlike its German counterpart ('*nur_weil_heißt_das_[noch_lange]_nicht* '), the English `just_because_doesn't_mean` construction (henceforth: JBDM construction) has been discussed extensively in the literature (e.g., Hirose, 1991; Bender & Kathol, 2001; Hilpert, 2005, 2007; Kanetani, 2011). What makes this construction an interesting example in our discussion of a constructicon for German is that its German counterpart does not only seem to exhibit very similar idiosyncratic syntactic properties, but it also seems to have a very similar range of meanings. If this proves to be true, the JBDM construction represents a clear instance of a related set of English constructions with clear correspondences in German. This means that we can seriously consider reusing English constructional entries as a starting point for creating their counterparts in the German constructicon.

Let us first have a look at the semantic properties of the JBDM construction. Standard *because*-clauses in English are ambiguous, because they can be interpreted as carrying both causal and inferential meaning (Bender & Kathol, 2001, pp. 14–16). Once the main clause of such sentences is negated, another ambiguity occurs due to the scope of the negation and the common cause/inference ambiguity. Narrow scope negation (i.e., solely the main clause is negated) only allows for the causal reading, while wide scope negation licenses both cause denial as well as inference denial. The same holds for the *because* clause. However, once the *because* clause takes the sentence-initial position, the two types behave differently as (7) shows.

(7) Nur weil für den Aufbau eines Konstruktikons große Anstrengungen nötig sind, heißt das nicht, dass dies unmöglich ist.
'Just because the development of a constructicon requires a lot of effort does not mean that it is impossible.'

Simple *because*-clauses only allow for a narrow scope reading of the negation, whereas, as Bender and Kathol (2001, p. 15) observe, "preposed *just because*-clauses continue to allow for both a narrow and a wide scope construal of the negation. However, the wide scope negation only allows for the inference denial interpretation." In (7), for example, the JBDM construction is used to deny the inference that building up a constructicon is impossible due to the required effort. To conclude, initial *just because* clauses behave in a way that is not licensed by combinations of other existing constructions, and, accordingly, they have to be seen as being licensed by a specific construction, "which calls for a *just because* adjunct preceding a negated main clause, and specifies that the negation in the main clause should take scope over the adjunct." (Bender & Kathol, 2001, p. 15).

Note that this holds true just as well for the German counterpart. Undoubtedly, an important function of the German nur_weil_heißt_das_nicht construction (henceforth: NWHN construction) is also to indicate that the *heißt nicht*-part does not necessarily follow from the *because*-part. Even more, using the construction means to distance oneself from the proposition expressed in the second part.

Hilpert (2005, p. 88; 2007, p. 31) observes that inference denial is only one meaning of the JBDM construction. In some cases the JBDM construction gives also rise to a more general meaning, namely that of concessivity. In contrast to (7) discussed above, the meaning of instances such as (8) cannot be reduced to inference denial. In the case of (8), for example, there is no way to infer what can be done from what one considers desirable.

(8) Nun, nur weil es wünschenswert ist, heißt das nicht, dass es machbar ist.
'Now, simply because it's desirable doesn't mean it's doable.'

(9) Nun, obwohl wir das gerne machen möchten, steht nicht fest, dass wir das tatsächlich machen können.
'Now, although we might want to do it, it is not certain that we actually can do it.'

Hilpert argues that instances of the JBDM construction such as (8) can be translated straightforwardly into the concessive construction in (9). He concludes that historically the JBDM construction has evolved "into a general marker of concessivity in modern usage" (Hilpert, 2007, p. 31), displaying idiosyncratic semantic properties that do not derive from the meanings of the parts the construction is made of. Again, the same concessive meaning is at work in the German counterparts. Semantically there is thus no difference whatsoever between the English JBDM constructions and its German counterpart.

Turning to the syntactic properties of the JBDM construction, we would like to point out two properties that are worth looking at more closely. First, the syntactic status of the *because*-part is anything but clear. Hirose (1991, pp. 18–19) argues that

the *just because*-part has a nominal structure serving as a grammatical subject while the *doesn't mean*-part instantiates the verb phrase. In contrast, however, Bender and Kathol question the subject status of the *just because*-clause, arguing that in many cases we find indeed a realized pronominal subject like *that* or *it*.

(10)   Nur weil wir Beispiele finden, heißt das nicht, dass die Analyse richtig ist.
'Just because we find examples it does not mean that the analysis is correct.'

Instances like (10) provide clear counterevidence against the assumed subject status of the *because*-part. Even in cases where there is no pronominal subject realized, Bender and Kathol (2001, p. 18) stick to this view, claiming that such instances feature an unexpressed subject. As a result, they generally assign adjunct status to the *doesn't mean*-part. Overall, the issues regarding the grammatical status of constructional elements arise from the (missing) realization of a pronominal subject. Both in English JBDM constructions and in their German counterparts we observe the same syntactic variability.

     Second, the JBDM construction features some lexical variation. Hilpert points out that instead of *mean* a variety of other verbs, such as *be*, *assume*, *give*, *make*, *have to*, *imagine*, among others, may enter the verb slot in the second part of the construction. The verb *mean*, however, clearly remains the most common and most frequently used verb. Almost all verbs indicate some kind of inferencing process, supporting its dominant semantics of inference denial. Even though a thorough corpus study on the German JBDM construction is still missing, at this fine-grained level of analysis there may be language-specific differences in German. Surprisingly, variation is also found in the negation (*doesn't*) in the second part of the construction. Hilpert (2007, p. 29) quotes examples challenging the standard view (e.g. Hirose, 1991) that the negation is a fixed property of this construction, as Hilpert's (2007) examples below show.

(11)   Nur weil's schwierig ist, ist ein armseliger Grund, es nicht auszuprobieren.
'Just because it's difficult is a poor reason not to try.'

Again, this variation is rarely found, interestingly both in English and in German. Regardless of these potential differences, we can conclude that the JBDM construction exhibits the same range of meaning variation as its German counterpart. The grammatical properties of the JBDM and the NWHN construction are also very similar. In particular, both in English and in German (a) the order of the constructional elements are fixed, (b) the *because*-part is headed by comparable lexical items (*just because/nur weil*), although variation exists but is very rare (e.g. *simply because/ einfach weil*), (c) the second part of the construction regularly contains a negation, again, however, exceptions are possible both in English and German, and (d) in both languages, there are narrow restrictions on the verbs that may enter the

construction, albeit language-specific differences do exist. Based on these common properties we propose that the German NWHN and the English JBDM construction are almost identical both semantically and syntactically. Pairs of constructions such as NWHN and JBDM exemplify one side of the continuum of constructional correspondences. In such cases, it would be most effective to reuse existing English construction entries to compile their German counterparts.

In the remainder of this paper, we discuss two more constructions, illustrating the 'continuum of correspondences': one with some remarkable language-specific features (the family of exclamative constructions) and one with largely language-specific peculiarities (the *way* construction).

4.2 The family of exclamative constructions: Exemplifying constructions with partial commonalities in German and English

We take the family of exclamative constructions as a good sample for exemplifying constructions with partial commonalities in English and German, thus illustrating the middle part of the 'continuum of constructional correspondences'. While, as we will show below, exclamative constructions exhibit some striking language-specific peculiarities on the form-side, the range of semantic variation in German and English remains roughly the same.

What are "exclamatives"? Even though forms and functions of exclamatives are well-examined (Zanuttini & Portner, 2009; d'Avis, 2013; Rett, 2009), most studies do not advance a constructional approach (for an exception cf. Michaelis, 2001; Ziem & Ellsworth, 2016). Across the stances taken in these studies, it is common sense that exclamatives constitute a sentence type that allows to express a speaker's surprise with regard to any kind of perceived entity, including events, situations, and objects the speaker comes across (d'Avis, 2013, p. 171; Rett, 2009, p. 607), provided that what is evoking the surprise diverges significantly from an expected default. To illustrate, (12) conveys the speaker's evaluation that the car referred to appears to him or her particularly big.

(12) Was für ein riesiges Auto!
'What a huge car!'

As Michaelis puts it, the surprise expressed by exclamatives generally entails

> a judgment by the speaker that a given situation is noncanonical. A noncanonical situation is one whose absence a speaker would have predicted, based on a prior assumption or set of assumptions, e.g., a stereotype, a set of behavioral norms, or a model of the physical world. (Michaelis, 2001, p. 1039)

To be more precise, Zanuttini and Portner identify three distinguishing features of exclamatives, namely (1) factivity, (2) scalar implicature, and (3) the inability to function in question/answer pairs (Zanuttini & Portner, 2003, pp. 46–50). As to (1), exclamatives are taken to be only embeddable in so-called factive predicates.

(13) a. Tom weiß, was das für ein riesiges Auto ist.
          'Tom knows what a huge car that is.'
      b. *Tom denkt, was für ein riesiges Auto das ist.
          'Tom thinks what a huge car that is.'

In Standard English, (13b) is not licensed since either quotation marks or commas are required to label the subordinate clause as a quotation (respectively as Tom's thoughts). As to (2), the scalar implicature peculiar to exclamatives derives from the very nature of surprises. More specifically, the implicature results from an unexpected property, such as the car's seize in (13), or any other observation the speaker comes across. Finally, the third characteristic, namely the inability of exclamatives to function in question/answer pairs, help distinguish exclamatives from other sentence types, particularly interrogative and declarative sentences. In contrast to the latter, exclamatives are not licensed to be part of question/answer pairs, as demonstrated in (14). More precisely, the problem with (14a) and (14b) is that B was pragmatically only acceptable under the condition that it would serve as the second part of an adjacency pair; however, in both cases the exclamatives A clearly do not instantiate the first part of an adjacency pair since they do not impose conditionally relevance on B.

(14) a. A: Was für ein riesiges Auto! B: [??]Es hat eine Länge von sechs Metern.
          'A: What a huge car! B: [??]It is six meters in length.'
      b. A: Wie riesig ist das Auto? B: Es hat eine Länge von sechs Metern.
          'A: How huge is the car? B: It is six meters in length.'
      c. 'A: Das Auto ist groß. B: Es hat eine Länge von fünf Metern.
          'A: The car is huge. B: It is five meters long.'

Beyond this widely uncontroversial set of defining criteria, some linguists additionally assume that exclamatives require some kind of gradable element, either explicitly expressed, such as "huge" in (12), or implicitly entailed, and thus inferred, as in (15).

(15) Was für ein Auto!
      'What a car!'

Depending on the context, the surprise expressed in (15) could be evoked by very different properties of the car, be it its elegance, its size, or some other distinctive characteristics such as its huge tires. However, it is anything but clear whether

scalarity, that is, the existence of gradable elements, is indeed a necessary property of exclamatives (for an extensive discussion see Ziem & Ellsworth, 2016; also d'Avis, 2013). If scalarity were constitutive for exclamatives, it would be impossible to account for many instances usually included in the family of exclamative constructions. Among them are, to mention but a few, the `What's_x_doing-_y` construction (Kay & Fillmore, 1999), `What_NP` constructions (e.g. '*What a mess this is!* 'Was für eine Schweinerei das ist!', *What a mess!* 'Was für eine Schweinerei!'), `Bare_NP` constructions (e.g., *A mess!* 'Eine Schweinerei!'), and constructions surfacing as NP initiated sub-clauses (e.g. *No surprise they didn't win!* 'Kein Wunder, dass sie nicht gewonnen haben!'). In the remainder of this section, aiming at maximalist coverage, we will not exclude instances of exclamative constructions for purely theoretical reasons. For the sake of the argument made here, we refrain from a more extensive discussion of scalarity (but see Ziem & Ellsworth, 2016, pp. 163–184).

We first like to draw attention to form-side variations of exclamative constructions by comparing the range of constructions in German with those in English. As summarized in Table 1, at least four parameters of cross-linguistic variations are worth a closer examination.

Table 1. Cross-linguistic variations of exclamative constructions

Cross-linguistic variation	German example	English counterpart
Verb position	(i) **Ist** das ein schöner Tag!	*Is this a nice day! What a nice day!
	(ii) Was für ein schöner Tag (**das ist/ist das**)!	*What a nice day is this! What a nice day this is!
Sub-clause initiated by complementizer	(iii) **Dass** das Wetter **so** herrlich ist!	??That the weather is so beautiful! This is such a beautiful weather!
	(iv) **Wenn** das kein voller Erfolg ist!	??If this is not a great success! This is such a great success!
Obligatory particle	(v) Das ist **aber** ein schnelles Auto!	*But this a fast car! What a fast car this is!
Lexical gap	(vi) Was **für** ein schöner Tag (das ist)!	*What for a nice day (this is)! What a nice day (this is)!
	(vii) **Als** wäre es Frühling!	*As it were spring! As if it were spring!

To begin with, verb-first exclamative constructions, such as (i), are well possible in German but not in English. This is also true for exclamatives surfacing as sub-clauses initiated by complementizers such as *dass* ('that') or *wenn* ('if') exemplified in (iii) and (iv). Furthermore, an interesting case are German `exclamative`

constructions with an obligatory particle such as *vielleicht* ('perhaps') or *aber* ('but'), as exemplified in (v). The particle is obligatory since its omission turns the exclamative into a declarative sentence. Finally, cross-linguistic variation concerns lexical gaps, that is, lexical categories missing in a construction in one language but turning up in the other, such as in (vi) and (vii). For example, only in German the `What_NP` exclamative, often considered the prototype of exclamative constructions, entails the preposition *für* 'for' (cf. (vi)).

Overall, these mismatches seem to indicate language specific peculiarities constituting at least partially distinctive families of exclamative constructions in German and English. On the other hand, however, there are also numerous homologous constructions including no deviations whatsoever. Among them are the following:

- `NP_as_exclamative` construction, e.g. *Ein schöner Tag!* ('A beautiful day')
- `Wh_NP` construction with *what* 'welch', e.g. *What a beautiful day!* ('Welch ein schöner Tag!')
- `AP_as_exlamative` constructions, e.g. *Schön!* ('Beautiful!')
- `comparative_as_exclamative` constructions initiated by *such* ('so'), e.g. *So ein schöner Tag!* ('Such a beautiful day!*')
- `WXDY` constructions with nominal focus elements, e.g. *Was macht mein Tagebuch auf deinem Schreibtisch?* ('What's my diary doing on your desk?')

Turning to the semantics of `exclamative` constructions the question arises to what extent the meanings of English and German `exclamative` constructions differ from one another. Presuming that generally, with the exception of synonymy, different forms trigger different meanings, we expect semantic deviation between German and English `exclamative` constructions where we detected cross-linguistic syntactic variation (see Table 1). Vice versa, full constructional correspondence is expected in cases of cross-linguistically homologous constructions.

The prototypical meaning of `exclamative` constructions can be summarized as follows: (a) `exclamative` constructions presuppose some kind of situation, in (15), for example, that the car referred to does exist; (b) `exclamative` constructions convey the speaker's surprise regarding a specific facet of the situation, for instance the car's seize in (12); (c) they cannot figure as part of a question/answer pair; (d) by means of scalar implicatures `exclamative` constructions implicitly or explicitly express the speaker's evaluation of what she or he is surprised about. In addition to that, `Bare_NP` and `Bare_AP` constructions (*A kangaroo!* 'Ein Kängeruh!'; *Wonderful!* 'Wunderbar!') require some pragmatic adjustment either regarding the property being surprised about in the case of `Bare_NPs` or regarding the entity referred to in the case of `Bare_APs`.

Also, most instances of the English WXDY construction (*What is my diary doing on your desk?* 'Was macht mein Tagebuch auf deinem Schreibtisch?') have equivalents in German. Note, however, that the WXDY construction also comprises instances with missing German counterparts. This holds, for example, for WXDY constructions with verbs as focus elements (e.g. *What are you doing looking at my diary?* 'Was in aller Welt schaust du in mein Tagebuch?').

Interestingly, we even find full semantic correspondences in some cases of syntactic mismatches, including English constructions with lexical gaps (e.g. *What a nice day!* 'What [for] a nice day!'), and constructions with/without word order constraints (*What a beautiful weather this is!* 'Was für ein schöner Tag [das ist]/[ist das]!'. Yet, in all other cases listed in Table 1, syntactic mismatches yield semantic mismatches, since English grammar neither licenses exclamatives with verb-first position (*Ist das ein schöner Tag*! '*Is this a nice day!') nor exclamatives surfacing as sub-clauses initiated by complementizers (*Dass das Wetter so herrlich ist!* *??That the weather is so beautiful). Furthermore, there are no equivalents to the particles constitutive for German exclamatives such as *Das ist aber/vielleicht ein schönes Auto!* ('What a nice car!'). When an exclamative is translated from one language into another, the semantic mismatches forces one to draw on related exclamative constructions. For example, both German exclamative constructions with an obligatory particle ('Das ist aber ein schnelles Auto!') and German verb_first_exclamatives are translated as what_NP construction.

In sum, we can conclude that the range of meanings expressed by exclamatives in German does not substantially differ from the range of meanings expressed by their English counterparts. At the same time, however, it is important to highlight that the family of German exclamative constructions encompasses a bunch of constructions with only partial correspondences to English counterparts. Compared with English, German offers a broader range of syntactic patterns to encode exclamative meaning. Most significantly, particles, such as *aber* and *vielleicht* (*Das ist aber/vielleicht ein schöner Tag!*) as well as exclamatives initiated by complementizers (*Dass der Tag so schön ist!*) and verb_first_exclamatives ('*Ist das ein schöner Tag!*') allow for encoding surprise without having one-to-one English equivalents. Hence, the family of exclamative constructions exhibit partial correspondences between English and German. They are located in the middle field of the 'continuum of construction correspondences'.

4.3 The `way` construction: Towards the other end of the 'continuum of constructional correspondences'

Finally, we turn to the *way* construction to illustrate what types of English constructions have no clear constructional correspondences in German and can thus be found towards the other end of the 'continuum of constructional correspondences'. In contemporary English, the *way*-construction, as illustrated in (16) and (17), is not only an interesting and well-documented type of argument structure construction (Marantz, 1992; Goldberg, 1995, 1997; Oya, 1999; van Egmond, 2009; Christie, 2011; among others), its distribution is also constantly expanding over time, as Israel (1996) observes.

(16)   They laughed their way off the stage.

(17)   The rat chewed his way through the wall.

While the English *way* construction comes with relatively few restrictions on the types of verbs that can fuse with it, as long as the resulting sentence can be interpreted as motion involving the main verb (means or manner) (Jackendoff, 1990; Goldberg, 1995), its German counterpart is different. In line with previous work by Maienborn (1994) and Kunze (1995), Oya (1999) shows that the German `reflexive_motion` construction can express similar types of scenarios. Following this analysis, the best candidates for German counterparts of the *way* construction as exemplified in (16) and (17) are the `reflexive` constructions (18) and (19).

(18)   Sie lachten sich von der Bühne.
       'They laughed themselves off the stage.'

(19)   Die Ratte kaute sich durch die Wand.
       'The rat chewed itself through the wall.'

Note, first, that the types of semantic restrictions licensing (18) and (19) are very similar to those of the their English equivalents: (1) only activity verbs (not unaccusative verbs) can fuse with the *way* construction and the German `reflexive` construction, (2) the motion expressed by the *way* construction is often metaphorical rather than literal (cf., for example, *She drank her way through a case of vodka*; Goldberg, 1995, p. 204), and (3) there is often an implication that the agent is overcoming some type of obstacle when moving in order to reach a specific goal.

However, Oya also points out some systematic differences between the English *way* construction and its German counterpart. For example, while English allows a non-causal interpretation as illustrated in (20) (Goldberg, 1995, p. 206), in which the motion and the sound emission run parallel, this is not possible in German, according to Oya (1999, p. 363).

(20) They were clanging their way up and down the narrow streets.
'*Sie klapperten sich die schmalen Straßen rauf und runter.'

Moreover, the *way* construction and the German `reflexive` construction differ in terms of the types of meanings they convey. Jackendoff (1990: Section 10.1) already emphasized that the *way* construction triggers two meanings; the main verb can either be considered the means for moving along a path, as in (21), or the manner of moving along a path, as shown in (22).

(21) Paul elbowed his way through the crowd.
'Paul benutzte seine Ellbogen, um durch die Menschenmenge zu kommen.'

(22) Peter danced his way through the hall in an hour.
'Peter tanzte sich in einer Stunde durch den Saal.'

Thus, (21) is interpreted in such a way that Paul uses his elbows (as a means) to get through the crowd, whereas in (22) the verb *to dance* describes a manner of moving rather than a means to move along. Interestingly, just like the English `reflexive` construction (Goldberg, 1997; Egmond, 2009; for an overview: Christie, 2011, pp. 3–4), its German equivalent does not allow for a means interpretation.

Since the *way* construction is, in that sense, polysemous, the constructicon needs to provide as many entries as constructional meanings exist.[19] The current prototype FrameNet constructicon for English thus comprises three entries for the *way* construction, namely *way_manner*, *way_means*, and *way_neutral*,[20] which all incorporate the same set of CEs, most importantly THEME, GOAL, and DIRECTION. They differ in specifying either the MANNER or MEANS of moving, or in being NEUTRAL in this respect. (23) exemplifies annotations for example (21).

(23) {[$_{\text{THEME}}$Paul] [$_{\text{MEANS}}$elbowed] [$_{\text{CEE}}$<his way>] [$_{\text{PATH}}$through the crowd]}.
'Paul benutzte seine Ellbogen, um durch die Menschenmenge zu kommen.'

Here, interestingly, the denominal verb incorporates a specification of MEANS. The instantiation of this CE qualifies (23) as a *way_means* construction.

It is worth having a closer look at the English `reflexive` construction, assuming that it is a true constructional equivalent of the German `reflexive` construction discussed above. Since, at least at first sight, the German `reflexive` construction

---

19. In this context it is interesting to note that Swedish basically patterns with German in that it, too, has a similar type of *way* construction (see Bäckström, Lyngfelt & Sköldberg, 2014; Lyngfelt, Torrent et al., this volume)), while Brazilian Portuguese does not (which means it is towards the far end of the constructional correspondence continuum).

20. See http://www1.icsi.berkeley.edu/~hsato/cxn00/21colorTag/index.html, accessed on August 5, 2016.

resembles much more the English `reflexive` construction than the *way* construc-
tion, it seems reasonable to suppose that the English `reflexive` construction may
differ from the *way* construction in the same way the latter distinguishes itself from
the German `reflexive` construction. But is this really the case? Summarizing and
compiling previous findings, Christie (2011) identifies four distinguishing features.
She argues that a `reflexive` construction is distinct from the *way* construction in
that (a) it does not allow for a manner interpretation, as shown in (22) above, (b)
it does not give rise to an atelic reading, (c) its PP does not denote a path, and (d)
it does not entail an event.

To be more precise, we would like to elaborate on each difference individually
while also looking at the commonalities of the German and English `reflexive` con-
structions. As already shown above, neither the English `reflexive` construction nor
its German counterpart permits a means interpretation of the event described. With
respect to telicity, (24) provides an illustrative example, reusing the instantiation of
the *way* construction given in (22).

(24) a. Peter danced himself through the hall in an hour.[21]
      b. Peter tanzte sich in einer Stunde durch den Saal.

Supplemented by a PP (*for an hour*/*eine Stunde lang*) supporting an atelic interpre-
tation, (25a) illustrates that the instantiated `reflexive` construction indeed cannot
express the event of dancing as being uncompleted. Instead, it is interpreted as fully
completed, while the time it takes for completion might be specified by an addi-
tional PP as exemplified by *in an hour* in (25a). This is also true for their German
equivalents provided in (24b) and (25b).

(25) a. ??Peter danced himself through the hall for an hour (and still hasn't finished).
      b. ??Peter tanzte sich eine Stunde lang durch den Saal (und ist noch nicht fertig).

Thus, in contrast to the *way_manner* construction, both the English and the German
`reflexive` constructions describe an activity with a terminal point. Another impor-
tant point is that the PP in the English `reflexive` construction (*through the hall*)
as well as its German equivalent (*durch den Saal*) do not encode a path because in
both languages the constructions do not permit indirect anaphoric references to
the respective path argument (for English see Christie, 2011, Section 3.3).

---

21. One reviewer noted that the sentences in (24) might not be found acceptable by all speakers
of English. We are aware of this issue as it has been noted before in the literature on `resultative`
and `caused_motion` constructions. Sentences such as those in (24) involve non-conventional
`resultative` and `caused_motion` uses of these verbs and have been analyzed as one-shot ex-
tensions based on analogical association, also known as pattern of coining (see, e.g. Boas, 2003;
Hanks, 2013; Kay, 2013).

(26) a.   Peter danced himself through the hall.??It led straight to the exit door.
      b.   Peter tanzte sich durch den Saal.??Er/Es führte direkt zur Ausgangstür.

In contrast to (26), indirect anaphoric references to the path are well possible in the *way_manner* construction. Presumably, this is so due to the grammaticalization of *way*. Finally, Christie maintains that the English `reflexive` construction does not entail an event. According to her, only the *way_manner* construction has "the ability for anaphoric reference to an entailed event" (Christie, 2011, p. 10). This point is arguably very similar to the last one, namely that the PP in the `reflexive` construction does not encode a path. However, encoding a path and depicting an event are two different things that should be kept separate. While indirect anaphoric reference to the path denoted in the PP of a `reflexive` construction fails, anaphoric reference to the event addressed in a `reflexive` construction is well possible, as demonstrated in (27).

(27) a.   Peter danced himself through the hall. It was beautiful.
      b.   Peter tanzte sich durch den Saal. Es war wundervoll.

In contrast to Christie (2011), we therefore doubt that the *way_manner* construction is distinct from the `reflexive` construction in that it entails an event. Indeed, English and German reflexive constructions do encompass an event more or less explicitly. Based on these observations, we conclude that the English `reflexive` construction and its German counterpart feature very similar characteristics, both syntactically and semantically. Most importantly, both trigger a manner interpretation and do not allow a means interpretation. While, however, in English there is also a `way_manner` construction that shares most features with the reflexive construction, there is no such alternative in German.

In addition, Oya's (1999) comparison of the English *way* construction with its German counterpart provides evidence that these two constructions are also relatively similar with respect to the types of verbs with which they can fuse. Importantly, however, they are very different when it comes to the specific constraints regulating the fusion of verbs and constructions, and these constraints are construction- and language-specific.

Overall, this makes the German `reflexive` construction a construction in its own right. It shares many, if not all, features with its English equivalent, it is functionally equivalent to the `way_means` construction, but it greatly differs from the *way_means* construction. Hence, in this case, we see no good reason to reuse the constructional entry of the English *way* construction for compiling an entry for the German `reflexive_motion` construction.

5.   Towards a German constructicon

The ultimate goal of the German Constructicon project (GCon, cf. http://gsw.phil. uni-duesseldorf.de), hosted at the University of Düsseldorf, is to identify and describe all constructions constituting the grammar of German in such a way that everything language users have to know in order to appropriately use and understand a construction is captured. Besides semantic, pragmatic, discourse-functional, and syntactic specifications, a full-fledged description of a construction also comprises information about relations to other constructions. Capturing the entire network of constructions in German constituting the constructicon is, to say the least, an ambitious long-term project that necessitates not only immense intellectual efforts, but also massive amounts of funding. However, it is worth getting started (for an overview cf. Boas & Ziem, in press a). More specifically, the project began by identifying possible German counterparts of English construction entries in the Berkeley FrameNet Constructicon (cf. Section 4.1). These include the family of so-called `negation_induced_connector` constructions, including the somewhat famous *let_alone* construction, and the family of `exclamative` constructions.

In the GCon project, we follow the directions of the Berkeley FrameNet constructicon approach. Particularly, we also aim at integrating constructions into a lexical frame-type database (Fillmore, 2008; Fillmore, Lee-Goldman & Rhomieux, 2012; for an overview Ziem, 2014a) by drawing on annotation categories and formalisms developed by the Berkeley pilot project (Fillmore, Lee-Goldman & Rhomieux, 2012). The constructional annotations are used to appropriately describe and define a construction. The most important constructicographic annotation categories include the Construction Evoking Element (CEE), the Construction Elements (CE), and the construct. In addition, we introduce a new annotation category called Correlated Element (CorE).

Just like a frame-evoking element in FrameNet, a CEE provides an explicit link to the respective target structure (here: the construction). Generally, a CEE is defined as the linguistic unit evoking a construction (Fillmore, Lee-Goldman & Rhomieux 2012: Section 2.2). Consider the following example.

(28)   Was für ein schöner Tag ist das!
       'What a beautiful day this is!'

In (28) the string of words *was für* evokes the `exclamative` construction. A CEE thus constitutes the lexical 'anchor' of a construction. Note, however, that such lexical elements are often missing. This is because not all constructions entail fixed lexical constituents. The more schematic a construction gets, that is, the more a construction is located towards the grammar pole in the lexicon-grammar continuum,

the more likely it is that it does not include one or more fixed lexical items. The `dis-transitive` construction, for example, is defined by its structural properties alone. The construct equates with the linguistic expression licensed by the `exclamative` construction. A construct is therefore an instance resp. an individual realization of the construction. The span of the construct is annotated by curly brackets. In (28) the construct encompasses *was für ein schöner Tag* ('what a beautiful day'). CEs can be defined as those constituents, or slots, of a grammatical construction that are instantiated by the respective parts of constructs. Following FrameNet annotation conventions, CEs are tagged with square brackets, while subscripts are used for labeling meaning or function. An `exclamative` construction, for example, essentially entails the CE Stimulus denoting the event or entity triggering the emotion of surprise (see Section 4.2).

Finally, a CorE is a word, or a string of words, that co-occurs with a construction in such a way that it enhances, or supplements, a (semantic, pragmatic, discourse-functional, syntactic) property of a construction. In the case of `exclamative` constructions, modal particles, such as *aber*, *denn*, *doch*, *vielleicht*, among others, used to function as CorEs in that they reinforce the speaker's surprise conveyed by the exclamative construction.

Next, consider (29), which exemplifies the annotations regarding the CEE, CEs, and the construct licensed by the construction. In addition to such semantic annotations, each CE and CEE is annotated syntactically (in terms of phrase type, part of speech, and grammatical function; cf. Section 4.2).

(29)　{$_{Was\_für}$[$_{CEE}$ Was für] [$_{Stimulus}$ ein schöner Tag]} ist das!
　　　'What a beautiful day this is!'

Collecting and analyzing all relevant information for each grammatical construction is a challenging and very time-intensive empirical task. In order to proceed both efficiently and consistently, it is necessary to have precise annotation guidelines and a uniform workflow guiding all construction analyses in the same way. To reduce efforts, the workflow should benefit from computational resources wherever possible. We therefore developed a partly computational workflow consisting of the five consecutive steps below (http://gsw.phil.uni-duesseldorf.de/).

– *Subcorporation and preliminary analysis*: Using existing corpora (particularly the DWDS corpus and the German Reference Corpus[22]), the first step aims at setting up a corpus of typical instances of the constructions under

---

22. Cf.http://www.dwds.de and http://www1.ids-mannheim.de/kl/projekte/korpora.html, last access: August 18, 2017.

investigation.[23] Once a set of corpus examples are identified and extracted, a preliminary analysis is performed to determine semantic, pragmatic, discourse-functional, and syntactic properties peculiar to the respective constructions. At this stage, CEs are also identified and tentatively defined to prepare semantic annotations.

– *Parsing Pipeline*: The parsing pipeline includes automatic annotations of part of speech, using the TreeTagger (Schmid, 1995) as well as phrase type and grammatical function by means of the BerkeleyParser (Petrov et al., 2006).

– *WebAnno*: For semantic annotations, we use the web-based annotation software WebAnno,[24] which supports the annotation of a wide range of project-specifically defined linguistic categories (cf. Castilho et al., 2016).

– *Construction Analyzer (CA)*: The Construction Analyzer is a web-based program that we developed for two purposes. First, it helps to automatically transform annotations into the annotation style used in the Berkeley FrameNet Constructicon. Second, and more importantly, it facilitates analyses of the annotations in several ways. For example, it identifies syntactic realization patterns of constructions as well as possible realizations of CEs and CEEs. Currently, CorE is implemented as an additional annotation layer.

– *Compilation of Construction Entries*: Finally, the results obtained from these analyses are carefully evaluated and interpreted with respect to their relevance for compiling a construction entry. Ideally, a construction entry should contain all information licensing a construct.

With this approach, both our methodology and our annotation procedure differ to some extent from the Swedish Constructicon Project (cf. Lyngfelt, Bäckström et al., this volume). In parallel with the Berkeley pilot project,[25] we provide a set of fully annotated sample instances for each construction along with a full-fledged definition of the construction, its CEs and CorEs (if any). As a consequence, in contrast to the Swedish Constructicon, each construction entry provides detailed information about in-depth analyses of the construction's usage conditions, including pragmatic, semantic, and syntactic constraints. So far, due to the elaborate analytical process, GCon covers only a few families of constructions, most notably

---

23. In FrameNet, this preliminary work is called "subcorporation"; this step subsumes "the automatic processes used to extract example sentences for annotation from the corpus" (https://framenet.icsi.berkeley.edu/fndrupal/glossary, last access: January 5, 2018; for more details cf. also Fillmore et al., 2003).

24. https://webanno.github.io, last access: August 18, 2017.

25. Cf. http://www1.icsi.berkeley.edu/~hsato/cxn00/21colorTag/index.html

the family of `exclamative` constructions and the `reduplication` construction (Ziem & Ellsworth, 2016; Ziem, in press). Overall, however, GCon does not differ from the Swedish constructicon project in that it uses corpora; both GCon and the Swedish constructicon are corpus-based. Rather, the major difference is that analyses in GCon are essentially driven by semantic annotations.

At present (December 2017), GCon is still in an early stage, during which the constructicographic workflow is being optimized for requirements specific to German constructions. However, in the near future we aim at a broader coverage by means of (a) including constructions that have already been extensively analyzed (for an overview cf. Ziem & Lasch, 2013, pp. 143–165), (b) addressing conceptually less complex constructions that do not require such fine-grained investigations, and (c) providing the opportunity to collaboratively compile constructional entries using a web-based platform. As for the latter, we plan to make the hitherto password protected repository of German constructions freely accessible in order to allow the scientific community to suggest new construction entries and preliminary analyses. This way, we ultimately intend to make the construction of GCon a joint collaborative project, open for everyone who would like to contribute to the constructionist enterprise.

6. Conclusions and outlook

As we have shown above, it is a hard and winding road from an English to a German constructicon. Even though there are some one-to-one constructional correspondences between English and German constructions, such as the *just_ because_doesn't_mean* construction and its German counterpart, many English constructions do not have clear-cut German equivalents. The *way* construction and the German `reflexive_motion` construction fall into this category. The fact that numerous basic German constructions (like those discussed in Section 2) do not have a straightforward English counterpart at all makes the situation even more complicated.

We take these findings as empirical support for doubting the usefulness of the Berkeley FrameNet constructicon, or any other constructicon, for directly creating parallel construction entries without questioning the annotation schema developed there. To be as comprehensive and precise as possible, we need a language-specific constructicon that meets the most fundamental grammatical requirements peculiar to German. In this view, the empirical evidence discussed so far suggests that reusing English construction entries is not always helpful (see also Lyngfelt, Bäckström et al., this volume). We therefore propose to start with parallel construction entries,

focusing solely on language-internal evidence from German as the basis for construction entries. This will ensure that the German constructicon will evolve in the style of the FrameNet constructicon while remaining at the same time conceptually independent of it. The corpus-based methodology we have in mind first focuses on the creation of German-specific construction entries by primarily relying on syntactic and semantic categories of German. This approach has the advantage of first providing detailed lexico-syntactic construction entries for German, linking these in larger networks of (families of) constructions. At a later point in time it may then be feasible to link German construction entries to construction entries of other languages, similar to approaches using transfer rules in rule-based machine translation (Slocum, 1987).

In line with the FrameNet Constructicon project (and in contrast to, for example, the Swedish constructicon, cf. Lyngfelt, Bäckström et al., this volume), such a non-contrastive German constructicon primarily builds on thoroughly annotated corpus examples illustrating (a) the set of CEEs evoking the construction, (b) the range of CEs specifying the construction, and (c) the syntactic variation of these CEs. Proceeding this way, we are currently developing a constructicon for German (http://gsw.phil.uni-duesseldorf.de) that is in part interlinked with G-FOL (the "German Frame-based Online Lexicon") (Boas & Dux, 2013; Boas, Dux & Ziem, 2016). Beyond the constructicon building efforts illustrated in Section 5, we are currently involved in a pilot project that uses the first-year German textbook "Deutsch im Blick" (http://coerll.utexas.edu/dib/) for full-text annotation of both lexical items (frame-based) and grammatical structures (construction-based). The project is designed as a long-term collaboration between UT Austin and HHU Düsseldorf, linking resources for both the manual annotation work and the web-based storing of constructions and frames in a FrameNet-like database.

References

Abraham, W. (1995). *Deutsche Syntax im Sprachenvergleich*. Tübingen: Narr.

Ackerman, F., & Webelhuth, G. (1998). *A Theory of Predicates*. Stanford: CSLI Publications.

Bäckström, L. Lyngfelt, B., & Sköldberg, E. (2014). Towards interlingual constructicography. On correspondence between constructicon resources for English and Swedish. *Constructions and Frames*, 6(1), 9–32. doi:10.1075/cf.6.1.02bak

Barðdal, J. (2006). Construction-Specific Properties of Syntactic Subjects in Icelandic and German. *Cognitive Linguistics*, 17(1), 39–106. doi:10.1515/COG.2006.002

Barðdal, J. (2008). *Productivity. Evidence from Case and Argument Structure in Icelandic*. Amsterdam & Philadelphia: John Benjamins. doi:10.1075/cal.8

Barðdal, J. (2009). The Development of Case in Germanic. In J. Barðdal, & S. Chelliah (Eds.), *The Role of Semantic, Pragmatic and Discourse Factors in the Development of Case* (pp. 123–159). Amsterdam & Philadelphia: John Benjamins. doi:10.1075/slcs.108.09bar

Barðdal, J. (2013). Construction-Based Historical-Comparative Reconstruction. In T. Hoffmann, & G. Trousdale (Eds.), *Oxford Handbook of Construction Grammar* (pp. 438–457). Oxford: Oxford University Press.

Bech, G. (1983). *Studien über das deutsche verbum infinitum*. 2. Auflage. Tübingen: Narr.

Behaghel, O. (1930). Von deutscher Wortstellung. *Zeitschrift für Deutschkunde*, 44, 81–89.

Bender, E. M., & Kathol, A. (2001). Constructional effects of *just because … doesn't mean*. *Berkeley Linguistics Society*, 27, 13–25.

Berman, C. (2003). *Clausal Syntax of German*. Stanford: CSLI Publications.

Bertoldi, A., & Chishman, R. L. de Oliveira. (2011). Developing a frame-based lexicon for the Brazilian legal language: The case of the Criminal_Process frame. *AICOL* 2011, 256–270.

Bierwisch, M. (1963). *Grammatik des deutschen Verbs*. Studia Grammatica No. 2. Berlin: Akademie Verlag.

Boas, H. C. (2002). Bilingual FrameNet dictionaries for machine translation. In M. González Rodríguez, & C. Paz Suárez Araujo (Eds.), *Proceedings of the Third International Conference on Language Resources and Evaluation*, Vol. IV, 1364–1371. Las Palmas, Spain.

Boas, H. C. (2003). *A constructional approach to resultatives*. Stanford: CSLI Publications.

Boas, H. C. (2005). Semantic frames as interlingual representations for multilingual lexical databases. *International Journal of Lexicography*, 18(4), 445–478. doi:10.1093/ijl/eci043

Boas, H. C. (2008). Determining the structure of lexical entries and grammatical constructions in Construction Grammar. *Annual Review of Cognitive Linguistics*, 6, 113–144. doi:10.1075/arcl.6.06boa

Boas, H. C. (Ed.) (2009a). *Multilingual FrameNets in Computational Lexicography*. Berlin & New York: Mouton de Gruyter. doi:10.1515/9783110212976

Boas, H. C. (2009b). Recent trends in multilingual computational lexicography. In H. C. Boas (Ed.), *Multilingual FrameNets in Computational Lexicography* (pp. 1–36). Berlin & New York: Mouton de Gruyter. doi:10.1515/9783110212976.0.1

Boas, H. C. (2010a). Comparing constructions across languages. In H. C. Boas (Ed.), *Contrastive Studies in Construction Grammar* (pp. 1–20). Amsterdam & Philadelphia: Benjamins. doi:10.1075/cal.10.02boa

Boas, H. C. (Ed.) (2010b). *Contrastive Studies in Construction Grammar*. Amsterdam & Philadelphia: Benjamins. doi:10.1075/cal.10

Boas, H. C. (2010c). Linguistically relevant meaning elements of English communication verbs. *Belgian Journal of Linguistics*, 24, 54–82. doi:10.1075/bjl.24.03boa

Boas, H. C. (2011). Zum Abstraktionsgrad von Resultativkonstruktionen. In S. Engelberg, K. Proost, & A. Holler (Eds.), *Sprachliches Wissen zwischen Lexikon und Grammatik* (pp. 37–69). Berlin & New York: Mouton de Gruyter. doi:10.1515/9783110262339.37

Boas, H. C. (2013a). Cognitive Construction Grammar. In T. Hoffmann, & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 233–254). Oxford: Oxford University Press.

Boas, H. C. (2013b). Wie viel Wissen steckt in Wörterbüchern? Eine frame-semantische Perspektive. *Zeitschrift für Angewandte Linguistik*, 57, 75–97.

Boas, H. C. (2013c). Frame Semantics and translation. In A. Rojo, & I. Ibarretxte-Antunano (Eds.), *Cognitive Linguistics and Translation* (pp. 125–158). Berlin & New York: Mouton de Gruyter. doi:10.1515/9783110302943.125

Boas, H. C. (2014). Zur Architektur einer konstruktionsbasierten Grammatik des Deutschen. In A. Ziem, & A. Lasch (Eds.), *Grammatik als Inventar von Konstruktionen? Sprachliches Wissen im Fokus der Konstruktionsgrammatik* (pp. 37–63). Berlin & New York: de Gruyter. doi:10.1515/9783110353693.37

Boas, H. C. (2017). Computational Resources: FrameNet and Constructicon. In B. Dancygier (Ed.), *The Cambridge Handbook of Cognitive Linguistics* (pp. 549–573). Cambridge: Cambridge University Press.

Boas, H. C. & Dux, R. (2013). Semantic frames for foreign language education: Towards a German frame-based dictionary. *Veridas Online. Special Issue on Frame Semantics and its Technological Applications* (pp. 82–100). http://www.ufjf.br/revistaveredas/edicoes-2013

Boas, H. C., Dux, R., & Ziem, A. (2016). Frames and constructions in an online learner's dictionary of German. In S. de Knop, & G. Gilquin (Eds.), *Applied Construction Grammar* (pp. 303–326). Berlin: De Gruyter.

Boas, H. C., & Gonzalvez-Garcia, F. (2014). *Romance perspectives on Construction Grammar*. Amsterdam & Philadelphia: John Benjamins.

Boas, H. C., & Ziem, A. (Eds.) (In press a). *Constructional Approaches to Argument Structure in German*. Berlin & Boston: De Gruyter Mouton.

Boas, H. C., & Ziem, A. (In press b). Approaching German syntax from a constructionist perspective. In H. C. Boas, & A. Ziem (Eds.), *Constructional Approaches to Argument Structure in German*. Berlin & Boston: De Gruyter Mouton.

Borin, L., Dannells, D., Forsberg, M., Toporowska Gronostaj, M., & Kokkinakis, D. (2010). The past meets the present in the Swedish FrameNet++. https://svn.spraakdata.gu.se/sb/fnplusplus/pub/SweFN_Euralex_extended.pdf.

Bresnan, J. (1982). Control and complementation. *Linguistic Inquiry*, 13(3), 343–434.

Burchardt, A., Erk, K., Frank, A., Kowalski, A., Padó, S., & Pinkal, M. (2009). Using FrameNet for the semantic analysis of German: annotation, representation, and automation. In H. C. Boas (Ed.), *Multilingual FrameNets: Methods and Applications* (pp. 209–244). Berlin & New York: Mouton de Gruyter.

Busse, D. (2012). *Frame-Semantik. Ein Kompendium*. Berlin & New York: Mouton de Gruyter. doi:10.1515/9783110269451

Castilho, R. E. de, Mújdricza-Maydt, É., Yimam, S. M., Hartmann, S., Gurevych, I., Frank, A., & Biemann, C. (2016). A Web-based Tool for the Integrated Annotation of Semantic and Syntactic Structures. In *Proceedings of the LT4DH workshop at COLING 2016*. Osaka, Japan.

Chesterman, A. (1998). *Contrastive Functional Analysis*. Amsterdam & Philadelphia: John Benjamins. doi:10.1075/pbns.47

Christie, E. (2011). Investigating the differences between the English way-construction and the fake reflexive resultative construction. In L. Armstrong (Ed.): CLA Conference Proceedings, 1–14. [Online publication, URL: http://homes.chass.utoronto.ca/~cla-acl/actes2011/Christie_2011.pdf, accessed on August 24, 2016]

Croft, W. (2001). *Radical Construction Grammar*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780198299554.001.0001

Croft, W. (2012). *Verbs*. Oxford: Oxford University Press.

Croft, W. (2013). Radical Construction Grammar. In T. Hoffmann, & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 211–232). Oxford: Oxford University Press.

Cruse, A. (1986). *Lexical Semantics*. Cambridge: University Press.

Čulo, O. (2003). Constructions-and-Frames Analysis of Translations: The Interplay of Syntax and Semantics in Translations between English and German. *Constructions and Frames*, 5(2), 143–167. doi:10.1075/cf.5.2.02cul

D'Avis, F. (2013). Exklamativsatz. In J. Meibauer, M. Steinbach, & H. Altmann (Eds.), *Satztypen des Deutschen* (pp. 171–201). Berlin & Boston: de Gruyter. doi:10.1515/9783110224832.171

Den Besten, H. (1983). On the interaction of root transformations and lexical deletive rules. In W. Abraham (Ed.), *On the formal syntax of the Westgermania: Papers from the 3rd Groningen Grammar Talks* (pp. 47–131). Amsterdam & Philadelphia: John Benjamins. doi:10.1075/la.3.03bes

Eisenberg, P. (2006). *Grundriss der deutschen Grammatik: Band 2: Der Satz*. Stuttgart: Metzler.

Eisenberg, P. & Gallmann, P. (2016). *Der Duden in 12 Bänden: 4 – Die Grammatik*. Bibliographisches Institut & FA Brockhaus AG.

Eisenberg, P. & Thieroff, R. (2013). *Grundriss der deutschen Grammatik: Band 2: Der Satz*. Stuttgart: Metzler.

Engel, U. (1988). *Deutsche Grammatik*. Heidelberg: Julius Groos.

Engelberg, S., König, S., Proost, K., & Winkler, E. (2011). Argumentstrukturmuster als Konstruktionen? In S. Engelberg, A. Holler, & K. Proost (Eds.), *Sprachliches Wissen zwischen Lexikon und Grammatik* (pp. 71–112). Berlin & New York: De Gruyter. doi:10.1515/9783110262339.71

Ellsworth, M., Ohara, K., Subirats, C., & Schmidt, T. (2006). Frame-semantic analysis of motion scenarios in English, German, Spanish, and Japanese. Paper presented at the Fourth International Conference on Construction Grammar, Tokyo, Japan. Available at http://jfn. st.hc.keio.ac.jp/publications/HoundICCG4.pdf.

Eroms, W. (1986). *Funktionale Satzperspektive*. Berlin & New York: Mouton de Gruyter. doi:10.1515/9783111336046

Fanselow, G. (1989). Coherent infinitives in German: Restructuring vs. IP-Complementation. In C. Bhatt, E. Loebel, & C. Schmidt (Eds.), *Syntactic phrase structure phenomena in noun phrase and sentences* (pp. 1–16). Amsterdam & Philadelphia: John Benjamins. doi:10.1075/la.6.02fan

Faulhaber, S. (2011). *Verb valency patterns*. Berlin & New York: Mouton de Gruyter. doi:10.1515/9783110240788

Fillmore, C. J. (1975). An alternative to checklist theories of meaning. *Proceedings of the First Annual Meeting of the Berkeley Linguistic Society (BLS)*, 123–131.

Fillmore, C. J. (1982). Frame Semantics. In Linguistic Society of Korea (Ed.), *Linguistics in the Morning Calm* (pp. 111–138). Seoul: Hanshin.

Fillmore, C. J. (1985). Frames and the Semantics of Understanding. *Quaderni di Semantica*, 6, 222–254.

Fillmore, C. J. (1999). Inversion and constructional inheritance. In G. Webelhuth, J. -P. Koenig, & A. Kathol (Eds.), *Lexical and Constructional Aspects of Linguistic Explanation* (pp. 113–128). Stanford: CSLI Publications.

Fillmore, C. J. (2007). Valency issues in FrameNet. In T. Herbst & K. Götz-Vetteler (Eds.), *Valency: theoretical, descriptive, and cognitive issues* (pp. 129–160). Berlin & New York: Mouton de Gruyter.

Fillmore, C. J. (2008). Border Conflicts: FrameNet meets Construction Grammar. In *Proceedings of the XIII EURALEX International Congress* (Barcelona, 15–19 July 2008), 49–68.

Fillmore, C. J. (2013). Berkeley Construction Grammar. In T. Hoffmann, & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 111–132). Oxford: Oxford University Press.

Fillmore, C. J. & Atkins, B. T. S. (2000). Describing polysemy: The case of *crawl*. In Y. Ravin, & C. Leacock (Eds.), *Polysemy: Linguistic and computational approaches* (pp. 91–110). Oxford: Oxford University Press.

Fillmore, C. J. & Baker, C. (2010). A frames approach to semantic analysis. In B. Heine, & H. Narrog (Eds.), *The Oxford Handbook of Linguistic Analysis* (pp. 313–340). Oxford: Oxford University Press.

Fillmore, C. J. & Kay, P. (1993). *Construction Grammar Course Book*. UC Berkeley: Department of Linguistics.

Fillmore, C. J., Lee-Goldman, R., & Rhomieux, R. (2012). The FrameNet Constructicon. In H. C. Boas, & I. A. Sag (Eds.), *Sign-based Construction Grammar* (pp. 309–372). Stanford: CSLI Publications.

Fillmore, C. J., Petruck, M. R. L., Ruppenhofer, J., & Wright, A. (2003). FrameNet in Action: The Case of Attaching. In *International Journal of Lexicography* 16, 297–333.

Fischer, K. (2013). *Satzstrukturen im Deutschen und Englischen*. Berlin: Akademie Verlag.

Forsbert, M., Johansson, R., Bäckström, L., Borin, L., Lyngfelt, B., Olofsson, J., & Prentice, J. (2014). From construction candidates to constructicon entries. *Constructions and Frames*, 6(1), 114–135. doi:10.1075/cf.6.1.07for

Fox, A. (1990). *The structure of German*. Oxford: Clarendon Press.

Goldberg, A. (1995). *Constructions: A Construction Grammar Approach to Argument Structure*. Chicago: University of Chicago Press.

Goldberg, A. (1997). Making one's way through the data. In A. Alsina, J. Bresnan, & P. Sells (Eds.), *Complex Predicates*. Stanford: CSLI Publications.

Goldberg, A. (2006). *Constructions at work*. Oxford: Oxford University Press.

Grewendorf, G. (1989). *Ergativity in German*. Dordrecht: Foris. doi:10.1515/9783110859256

Haider, H. (1985). A unified account of case and theta-marking: The case of German. *Papiere zur Linguistik*, 32, 3–36.

Haider, H. (1986). Fehlende Argumente: Vom Passiv zu kohärenten Infinitiven. *Linguistische Berichte*, 101, 3–33.

Haider, H. (1990). Topicalization and other puzzles of German syntax. In G. Grewendorf, & W. Sternefeld (Eds.), *Scrambling and Barriers*, (pp. 93–112). Amsterdam & Philadelphia: John Benjamins doi:10.1075/la.5.06hai

Haider, H. (1993). *Deutsche Syntax – Generativ*. Tübingen: Narr.

Hanks, P. (2013). *Lexical Analysis*. Cambridge, Mass.: MIT Press. doi:10.7551/mitpress/9780262018579.001.0001

Hasegawa, Y., Lee-Goldman, R., Ohara, K. H., Fuji, S., & Fillmore, C. J. (2010). On expressing measurement and comparison in English and Japanese. In H. C. Boas (Ed.), *Contrastive Studies in Construction Grammar* (pp. 169–200). Amsterdam & Philadelphia: John Benjamins. doi:10.1075/cal.10.08has

Haspelmath, M. (2007). Pre-established categories don't exist: Consequences for language description and typology. *Linguistic Typology*, 11(1), 119–132. doi:10.1515/LINGTY.2007.011

Hawkins, J. (1986). *A comparative typology of English and German: Unifying the contrasts.* London: Routledge.

Heid, U. (1996). Creating a multilingual data collection for bilingual lexicography from parallel monolingual lexicons. In *Proceedings of the VIIth EURALEX International Congress*, Gothenburg 1996, 573–559.

Hentschel, E., & Weydt, H. (2013). *Handbuch der Deutschen Grammatik*. 4th edition. Berlin & Boston: Walter de Gruyter. doi:10.1515/9783110312973

Herbst, T. (2014). The valency approach to argument structure constructions. In T. Herbst, H. -J. Schmid, & S. Faulhaber (Eds.), *Constructions – Collocations – Patterns* (pp. 167–216). Berlin: de Gruyter Mouton.

Hilpert, M. (2005). A diachronic perspective on concessive constructions with *just because*. In A. Makkai, W. J. Sullivan, & A. R. Lommel (Eds.), *Lacus Forum XXXI: Interconnections* (pp. 67–80). Houston: LACUS.

Hilpert, M. (2007). Just because it's new does not mean people will notice it: The idiosyncratic properties of a recent usage. *English Today*, 91–92(23), 29–33.

Hirose, Y. (1991). On a certain nominal use of because-clauses: Just because because-clauses can substitute for that-clauses does not mean that this is always possible. *English Linguistics*, 8, 16–33. doi:10.9793/elsj1984.8.16

Höhle, T. N. (1982). Über Komposition und Derivation: zur Konstituentenstruktur von Wortbildungsprodukten im Deutschen. *Zeitschrift für Sprachwissenschaft*, 1(1), 76–112.

Höhle, T. (1986). Der Begriff "Mittelfeld", Anmerkungen über die Theorie der topologischen Felder. In W. Weiss, H. E. Wiegand, & M. Reis (Eds.), *Akten des VII. Kongresses der Internationalen Vereinigung für germanische Sprach- und Literaturwissenschaften*. Göttingen 1985. 3rd Vol. (pp. 329–340). Tübingen: Max Niemeyer Verlag.

Hoffmann, T. & Trousdale, G. (Eds.) (2013). *The Oxford Handbook of Construction Grammar*. Oxford: Oxford University Press. doi:10.1093/oxfordhb/9780195396683.001.0001

Imo, W. (2016). *Grammatik: Eine Einführung*. Stuttgart: Metzler.

Israel, M. (1996). The *way* constructions grow. In A. Goldberg (Ed.), *Conceptual structure, discourse and language* (pp. 217–230). Stanford, CA: CSLI Publications.

Iwata, S. (2008). *Locative alternation: a lexical-constructional approach*. Amsterdam & Philadelphia: John Benjamins. doi:10.1075/cal.6

Jacobs, J. (1986). The syntax of focus and adverbials in German. In W. Abraham, & S. De Meij (Eds.), *Topic, Focus, and Configurationality. Papers from the 6th Groningen Grammar Talks*, Groningen 1984 (pp. 103–127). Amsterdam & Philadelphia: John Benjamins.

Jackendoff, R. (1990). *Semantic Structures*. Boston: MIT Press.

James, C. (1980). *Contrastive Analysis*. London: Longman.

Kanetani, M. (2011). Analogy in Construction Grammar: The case of *Just Because of X Doesn't Mean Y'. Tsukuba English Studies*, 29, 77–94.

Kathol, A. (2000). *Linear Syntax*. Oxford: Oxford University Press.

Kay, P. (2013). The limits of (Construction) Grammar. In T. Hoffmann, & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 32–48). Oxford: Oxford University Press.

Kay, P. & Fillmore, C. J. (1999). Grammatical constructions and linguistic generalizations: The 'What's X doing Y?' Construction. *Language*, 75, 1–33. doi:10.2307/417472

Kirkwood, H. W. (1969). Aspects of word order and its communicative function in English and German. *Journal of Linguistics*, 5, 85–107. doi:10.1017/S0022226700002073

Kunze, J. (1995). Reflexive Konstruktionen im Deutschen. *Zeitschrift für Sprachwissenschaft*, 14, 3–53.

Lakoff, G. (1987). *Women, fire, and dangerous things. What categories reveal about the mind.* Chicago: University of Chicago Press.

Lasch, A. (2016). *Nonagentive Konstruktionen des Deutschen*. Berlin & Boston: de Gruyter.

Laviola, A. (2015). *Frames e Construções em Contraste: uma análise comparativa português-inglês no tangente à implementação de constructicons* ['Frames and Constructions in Contrast: a Portuguese-English comparative analysis in regards to the implementation of constructicons']. MA Thesis, Federal University of Juiz de Fora.

Leino, J. (2010). Results, cases, and constructions: Argument structure constructions in English and Finnish, In H. C. Boas (Ed.), *Contrastive Studies in Construction Grammar* (pp. 103–136). Amsterdam & Philadelphia: Benjamins. doi:10.1075/cal.10.06lei

Lenerz, J. (1977). *Zur Abfolge nominaler Satzglieder im Deutschen*. Tübingen: Stauffenburg.

Lyngfelt, B., Borin, L., Forsberg, M., Prentice, J., Rydstedt, R., Sköldberg, E., & Tingsell, S. (2012). Adding a constructicon to the Swedish resource network of Språkbanken. In *Proceedings of KONVENS 2012, Vienna*, 452–461.

Marantz, A. (1992). The *way*-construction and the semantics of direct arguments in English. In E. Wehrli, & T. Stowell (Eds.), *Syntax and the lexicon* (pp. 179–188). New York: Academic Press.

Meurers, W. D. (2000). *Lexical generalizations in the syntax of German non-finite constructions*. Arbeitspapiere des SFB 340, No. 145. Eberhard-Karls-Universität Tübingen.

Michaelis, L. (2001). Exclamative Constructions. In M. Haspelmath, E. Wulf Österreicher, & W. Raible (Eds.), *Language Universals and Language Typology: An International Handbook* (pp. 1038–1050). Berlin: de Gruyter.

Michaelis, L. (2012). Making the case for Construction Grammar. In H. C. Boas, & I. Sag (Eds.), *Sign-based Construction Grammar* (pp. 30–68). Stanford: CSLI Publications.

Michaelis, L. & Ruppenhofer, J. (2001). *Beyond Alternations*. Stanford: CSLI Publications.

Müller, S. (2005). German: A grammatical sketch. In A. Alexiadou, & T. Kiss (Eds.), *Syntax – ein internationales Handbuch zeitgenössischer Forschung*, 2nd Edition. Berlin: de Gruyter.

Ohara, K. H. (2009). Frame-based contrastive lexical semantics in Japanese FrameNet: The case of *risk* and *kakeru*. In H. C. Boas (Ed.), *Multilingual FrameNets: Methods and Applications* (pp. 163–182). Berlin & New York: Mouton de Gruyter.

Ohara, K. H. (2014). Relating Frames and Constructions in Japanese FrameNet. *Proceedings of the Ninth International Conference on Language Resources and Evaluation (LREC '14)*, 2474–2477.

Oya, T. (1999). Er bettelt sich durchs Land: Eine one's way Konstruktion im Deutschen? *Deutsche Sprache*, 27, 356–369.

Padó, S. (2007). Translational equivalence and cross-lingual parallelism: The case of FrameNet frames. *Proceedings of the NODALIDA workshop on building frame semantics resources for Scandinavian and Baltic languages*, 39–46.

Padó, S., & Lapata, M. (2009). Cross-lingual annotation projection of semantic roles. *Journal of Artificial Intelligence Research*, 36, 307–340.

Petrov, S., Barrett, L., Thibaux, R., & Klein, D. (2006). Learning Accurate, Compact, and Interpretable Tree Annotation. In *Proceedings of the 21st International Conference on Computational Linguistics and 44th Annual Meeting of the ACL* (pp. 433–440). Sydney, Australia: Association for Computational Linguistics.

Petruck, M. R. L. (2009). Typological considerations in constructing a Hebrew FrameNet. In H. C. Boas (Ed.), *Multilingual FrameNets: Methods and Applications* (pp. 183–208). Berlin & New York: Mouton de Gruyter.

Pollard, C. & Sag, I. A. (1993). *Head-driven Phrase Structure Grammar*. Stanford: CSLI Publications.

Reis, M. (1980). On justifying topological frames: 'positional field' and the order of non-verbal constituents in Berman. *Documentation et Recherche en Linguistique Allemande Contemporaire*, 22/23, 59–85.

Reis, M. (1985). Wer glaubst Du hat recht? On the so-called extractions from verb-second clauses and verb-first parenthetical constructions in German. *Sprache und Pragmatik*, 36, 27–83.

Reis, M. (1987). Die Stellung der Verbargumente im Deutschen. Stilübungen zum Grammatik-Pragmatik-Verhältnis. In I. Rosengren (Ed.), *Sprache und Pragmatik* (pp. 139–177). Stockholm: Almqvist & Wiksell.

Rett, J. (2009). *A degree account of exclamatives*. Semantics and Linguistic Theory, XVIII, 601–618.

Rimsdijk, H. van. (1985). Zum Rattenfängereffekt bei Infinitiven in deutschen Relativsätzen. In W. Abraham (Ed.), *Erklärende Syntax des Deutschen*. Tübingen: Narr.

Roja, A. (2002). Frame semantics and the translation of humor. *Babel: International Journal of Translation*, 48(1), 34–77. doi: 10.1075/babel.48.1.03lop

Rojo, A., & Valenzuela, J. (1998). Frame semantics and lexical translation: The risk frame and its translation. *Babel: International Journal of Translation*, 44(2), 128–138. doi: 10.1075/babel.44.2.04lop

Ruppenhofer, J., Boas, H. C., & Baker, C. (2013). The FrameNet approach to relating syntax and semantics. In R. H. Gouws, U. Heid, W. Schweickhard, & H. E. Wiegand (Eds.), *Dictionaries. An International Encyclopedia of Lexicography* (pp. 1320–1329). Berlin & New York: Mouton. doi: 10.1515/9783110238136.1320

Ruppenhofer, J., Ellsworth, M., Petruck, M. R. L., Johnson, C., & Scheffczyk, J. (2010). *FrameNet II: Extended theory and practice*. Available at [http://framenet.icsi.berkeley.edu].

Sag, I. A. (2012). Sign-based Construction Grammar: An informal synopsis. In H. C. Boas, & I. A. Sag (Eds.), *Sign-based Construction Grammar* (pp. 69–202). Stanford: CSLI Publications.

Salomão, M. M. M., Torrent, T. T., & Sampaio, T. F. (2013). A Linguística de Corpus Encontra a Linguística Computacional: Notícias do Projeto FrameNet Brasil. *Cadernos de Estudos Linguísticos*, 55(1), 7–34.

Schmid, H. (1995). Improvements in Part-of-Speech Tagging with an Application to German. In *Proceedings of the ACL SIGDAT-Workshop*. Dublin, Ireland.

Schmidt, T. (2009). The Kicktionary – A multilingual lexical resource of football language. In H. C. Boas (Ed.), *Multilingual FrameNets: Methods and Applications* (pp. 101–134). Berlin & New York: Mouton de Gruyter.

Slocum, J. (1987). Concept-lexeme-syntax triangles: A gateway to interlingual translation. *Computers and Translation*, 2, 243–262. doi: 10.1007/BF01682183

Sommerfeldt, K. -W., & Starke, G. (1992). *Einführung in die Grammatik der deutschen Gegenwartssprache*. Tübingen: Niemeyer.

von Stechow, A., & Sternefeld, W. (1988). *Bausteine Syntaktischen Wissens*. Opladen: Westdeutscher Verlag.

Steels, L. (2013). Fluid Construction Grammar. In T. Hoffmann, & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 153–167). Oxford: Oxford University Press.

Subirats, C. (2009). Spanish FrameNet: A frame-semantic analysis of the Spanish lexicon. In H. C. Boas (Ed.), *Multilingual FrameNets: Methods and Applications* (pp. 135–162). Berlin & New York: Mouton de Gruyter.

Subirats, C. & Petruck, M. R. L. (2003). Surprise: Spanish FrameNet! *Proceedings of CIL 17*. CD-ROM. Prague: Matfyzpress.

Timyan, N., & Bergen, B. (2010). A contrastive study of the caused-motion and ditransitive construction in English and Thai: Semantic and pragmatic constraints. In H. C. Boas (Ed.), *Contrastive Studies in Construction Grammar* (pp. 137–168). Amsterdam & Philadelphia: John Benjamins. doi:10.1075/cal.10.07tim

Torrent, T. T., Lage, L. M., Sampaio, T. F., da Silva Tavares, T., & da Silva Matos, E. E. (2014). Revisiting border conflicts between FrameNet and Construction Grammar: Annotation policies for the Brazilian Portuguese Constructicon. *Constructions and Frames*, 6(1), 33–50.

Uszkoreit, H. (1987). *Word order and constituent structure in German*. Stanford: CSLI Publications.

Van Egmond, M. -E. (2009). *Two way-constructions in Dutch: motion along a path and transition to a location*. VDM Verlag.

Van Valin, R. & Wilkins, D. P. (1996). The case for 'Effector': Case Roles, agents, and agency revisited. In M. Shibatani, & S. Thompson (Eds.), *Grammatical Constructions* (pp. 289–322). Oxford: Oxford University Press.

Webelhuth, G. (1992). *Principles and Parameters of Syntactic Saturation*. Oxford: Oxford University Press.

Welke, K. (2011). *Valenzgrammatik des Deutschen*. Berlin: de Gruyter.

Willems, D. (2012). Verb typology: Between construction and lexicon. In M. Bouveret, & D. Legallois (Eds.), *Constructions in French* (pp. 23–48). Amsterdam & Philadelphia: John Benjamins. doi:10.1075/cal.13.03wil

Wöllstein, A. (2010). *Topologisches Satzmodell*. Heidelberg: Winter.

Wöllstein-Leisten, A., Heilmann, A., Stepan, P., & Vikner, S. (1997). *Deutsche Satzstruktur*. Tübingen: Stauffenburg Verlag.

Ziem, A. (2014a). Von der Kasusgrammatik zum FrameNet: Frames, Konstruktionen und die Idee eines Konstruktikons. In A. Ziem, & A. Lasch (Eds.), *Grammatik als Inventar von Konstruktionen? Sprachwissen im Fokus in der Konstruktionsgrammatik* (pp. 263–290). Berlin & New York: de Gruyter. doi:10.1515/9783110353693.263

Ziem, A. (2014b). *Frames of Understanding in Text and Discourse: Theoretical Foundations and Descriptive Applications*. Amsterdam & Philadelphia: Benjamins.

Ziem, A. (in press). *Tag für Tag Arbeit über Arbeit*: konstruktionsgrammatische Zugänge zu Reduplikationsstrukturen im Deutschen. In K. Steyer (Ed.): *Sprachliche Verfestigung. Wortverbindungen, Muster, Phrasem-Konstruktionen*. Tübingen: Narr.

Ziem, A., Boas, H. C., & Ruppenhofer, J. (2014). Grammatische Konstruktionen und semantische Frames für die Textanalyse. In J. Hagemann, & S. Staffeldt (Eds.), *Syntaxtheorien. Analysen im Vergleich* (pp. 297–333). Tübingen: Stauffenburg.

Ziem, A., & Ellsworth, M. (2016). Exklamativsätze im FrameNet-Konstruktikon. In R. Finkbeiner, & J. Meibauer (Eds.), *Satztypen und Konstruktionen im Deutschen* (pp. 146–191). Berlin & Boston: de Gruyter.

Ziem, A., & Lasch, A. (2013). *Konstruktionsgrammatik. Konzepte und Grundlagen gebrauchsbasierter Ansätze*. Berlin: de Gruyter. doi:10.1515/9783110295641

Ziem, A., & Staffeldt, S. (2011). Compositionality and embodied meanings of phraseologisms: the case of somatisms. In D. Schönefeld (Ed.): *Converging evidence: Methodological and Theoretical Issues for Linguistic Research* (pp. 195–219). Amsterdam & Philadelphia: Benjamins. doi:10.1075/hcp.33.13zie

Zifonun, G., Hoffmann, L., & Strecker, B. (1997). *Grammatik der deutschen Sprache.* 3 volumes. Berlin & New York: Mouton de Gruyter.

Zanuttini, R., & Portner, P. (2003). Exclamative Clauses: At the Syntax-Semantics Interface. *Language*, 79(1), 39–81. doi:10.1353/lan.2003.0105

Linguistics vs. language technology in constructicon building and use

Lars Borin, Dana Dannélls and Normunds Grūzītis

In this chapter, we describe the close interaction of linguists and language technologists in the Swedish constructicon project. This kind of collaboration is not so common today, because of the way that language technology has developed in recent decades, but in our case the collaboration has been very successful, and constituted a genuine instance of cross-fertilization, where an evolving language technology infrastructure and a computational lexical macroresource described in the chapter has formed an integral part of the Swedish constructicon development environment, while at the same time the structured linguistic knowledge described in the constructicon has informed the language technology making up the infrastructure.

Keywords: language technology, natural language processing, computational linguistics, machine learning, grammatical framework, computational lexical resource, corpus, Swedish, research infrastructure

1. Introduction

It may come as a surprise to most linguists – who we assume form the primary readership of this book – that, despite its name, present-day computational linguistics has a quite tenuous and in many cases nonexistent relationship to the academic discipline of linguistics. According to at least one author, "computational linguistics is not a specialization of linguistics at all; it is a branch of computer science" (Abney, 2011, p. 1).

While it is true that computational linguistics (or language technology/natural language processing/language engineering)[1] encompasses many kinds of activities

---

1. In this chapter, we will prefer the term "language technology" (LT), as that being most frequently encountered today, and also understood to subsume both text and speech technologies. We will also talk about NLP (natural language processing), as the application of LT systems to empirical language data, typically large volumes of text.

all dealing with language,[2] the fact of the matter is still that it rarely overlaps in its most fundamental concerns with those of linguistics. Rather than striving to describe human language or explain human linguistic competence, current language technology has as its main aim to build systems capable of processing unrestricted natural-language text (or speech), typically for some particular purpose or purposes (e.g., information access, translation, grammar correction, spoken dialogue). Even though such systems could be thought of as being crucially dependent on human-style language understanding, no claim is made on the part of their designers that these systems actually mimic how humans process language. And of course identical external behavior in a system can be caused by an infinitude of internal configurations. The designers of language technology systems are only interested in how to produce the desired behavior, not whether the computational machinery whereby this is achieved has anything in common with the cognitive or neurological mechanisms underlying human linguistic abilities.

For many years now, language technology has been completely dominated by statistical systems based on machine learning, so-called *data-driven* systems; for every grammar-based or rule-based LKB (Copestake, 2002) or Giellatekno (Trosterud, 2006) type system we find at least a dozen data-driven RASP (Briscoe, Carroll, & Watson, 2006) or MALT (Nivre et al., 2007) type systems in the literature.

As part of this development, even though the two fields started out close, research in language technology has become increasingly disassociated from the concerns of linguistics (Reiter, 2007). According to Wintner (2009, p. 642), there are mainly three reasons for this: (1) "applications that were based on explicit linguistic knowledge didn't scale up well"; (2) "[f]unding agencies (mainly in the U.S.) are motivated by short-term practical goals"; and (3) "[linguistics] focused mainly on syntax (and predominantly on English): and its theory became so obscure, so baroque, and so self-centered, that it became virtually impenetrable to researchers from other disciplines". This characterization primarily refers to the kind of linguistics that informed much early research in language technology, much of it being pursued under the label of generative grammar. On the other hand, language technology remained strangely untouched by other vigorous strands of linguistic research, such as typological and contact linguistics, sociolinguistics, psycholinguistics, or lexicography (including lexical semantics), to mention a few that all have generated a substantial body of work simultaneously with – sometimes even well before – the kind of linguistics that Wintner refers to.

Be that as it may, the fact of the matter is that data-driven approaches have made huge progress in recent times, in the wake of increasingly powerful computer

---

2.   See <https://aclanthology.coli.uni-saarland.de/>

hardware and the availability of enormous amounts of training data in the form of billions of words of digital text.[3] The most recent incarnation of data-driven language technology is the so-called "deep learning" paradigm, which – being capable of "learning language" automatically from large amounts of raw textual data – according to some predictions will put language technologists out of business completely. However, others point out that machine learning will not make the domain problems go away, and that instead of focusing on numbers, "[m]ore of the field's effort should go into problems, approaches, and architectures" (Manning, 2015, p. 702).

In a discussion of the role of linguistics versus language technology in constructicon building and use, this particular strand of language technology research must loom large, for at least two reasons: (1) This is arguably the kind of language technology research which at present is most distant from contemporary linguistics, which makes the contrast to the constructicon work described elsewhere in this volume maximal (and the corresponding comparison maximally methodologically and theoretically fruitful); (2) since the data-driven systems seem to be here to stay and in reality today dominate the leading language technology conferences completely, they cannot be ignored – any serious attempt to include the kind of linguistic knowledge encoded in constructicons into state-of-the-art language processing systems must somehow accommodate this fact – which is why in the language technology research community, arguments are now made for the desirability of a new "linguistic turn" in language technology, and for allowing explicit linguistic knowledge to inform the statistical language processing paradigm (Reiter, 2007; Wintner, 2009; Bender, 2011; Manning, 2015). Except at its very beginning, the interaction of computational and other linguistics has been almost exclusively with the kind of linguistics that Wintner characterizes as syntax-focused and English-centered. What is rarely noted is that, even here, the traffic has been largely one-way, since, with some rare exceptions, work in computational linguistics has had very little influence on the development of theory or methodology in general linguistics (Pullum, 2009). However, Reiter (2007), Wintner (2009), Abney (2011), Bender (2011) and Manning (2015) see many opportunities for closer interaction between LT and linguistics, which hopefully would not be one-way

---

3.   But note that the availability of training data is extremely unevenly distributed over the world's languages and further that many of the most popular data-driven methods contain a good deal of hidden language dependence (Bender, 2011). Because of this, grammar-driven language technology maintains a more modest but thriving existence in "ecological niches" which are still out of reach to data-driven techniques: Low-resource languages (Trosterud, 2006; Tyers & Pirinen, 2016), historical language varieties (Bollmann, 2013), controlled languages (Angelov & Ranta, 2009), and some others.

anymore, but a genuinely synergistic undertaking. In our view, too, these two fields stand to benefit considerably from a rapprochement. On the one hand, data-driven LT offers a well-developed and rigorous methodology for conducting large-scale hypothesis-driven empirical linguistic investigations (see, e.g., Abney, 2011), even though it is normally not referred to in these terms, but rather presented under the narrower heading "evaluation". This methodology, together with the analysis tools being developed and refined through it, has the potential to both scale up and widen the scope of corpus-based linguistics (see, e.g., Pullum 2007), which will benefit endeavors such as constructicon building (see Section 3 below), empirical investigation of constructs such as "entrenchment" (Schmid 2010), and others.

On the other hand, because of the Zipfian distribution of linguistic phenomena, even very large corpora, while revealing many things that linguists did not know about a language or language in general, will also almost certainly fail to provide evidence of some things that linguists do know. For this reason, it would be useful if these two sources of knowledge could somehow be combined.

In this connection, highly refined linguistic knowledge such as that encoded in framenets or constructicons is especially interesting, but at the same time raises many methodological challenges. In part this is because of the sometimes very different theoretical assumptions and methodological decisions characteristic of linguistics and language technology, especially those assumed to be self-evident and consequently unexpressed (Borin et al., 2010; Borin, Forsberg, & Lyngfelt, 2013).

2.   Some theoretical and methodological observations

There is a more trivial and less interesting form of interaction between linguistics and data-driven language technology. Most data-driven methods learn a classification or labeling, e.g., of text word occurrences for their parts of speech, their syntactic functions, or their lexicon senses. Data-driven methods are conventionally subclassified into *supervised* and *unsupervised* methods. The unsupervised methods learn a classification from being exposed only to the data itself, i.e., the unannotated text in our case,[4] while supervised methods require preclassified training data, i.e., text where each word occurrence has been assigned its correct part of speech, syntactic function, lexicon sense, etc.

Since supervised methods tend to achieve considerably better classification accuracy for most kinds of linguistic annotation, language technology relies on

---

4.   Although note that even "raw" text may come "pre-cooked" with some linguistic analysis, e.g., in the form of (orthographic) word and sentence segmentation, orthographic marking of proper nouns or nouns in general, etc.

linguistics for producing training data labelled with the relevant linguistic analyses. This is a very time-consuming and consequently expensive undertaking, and it furthermore requires that the language variety has been described linguistically to sufficient detail and also that there are trained annotators available, so that sufficient amounts of training data can be produced. In practice, for many of the world's languages we do not have good linguistic descriptions, and even for languages with many speakers and a long history of linguistic description, annotators with suitable training may be hard to come by (Liberman, 2009; Peldszus & Stede, 2013).

Consequently, a more promising avenue would consist in somehow making explicit linguistic descriptions, such as those found in a constructicon, inform data-driven language analysis, e.g., by introducing biases or/and guiding feature selection in the learning algorithms. In the case of grammar-driven systems, constructicon entries can of course be incorporated into the formalism itself, either directly or in modified form (see Section 4 below).

It was mentioned above that language technology has influenced linguistics (at least theoretical linguistics) even less than the opposite. This unfortunate state of affairs may in fact be rectified through a natural development, stemming from the availability of enormous corpora with increasingly sophisticated linguistic annotations and search interfaces, which linguists are more or less forced to use because of the sheer volumes of text available today, which cannot be investigated except with the help of computational tools. As empirical linguistics is becoming increasingly corpus-driven, linguists are exposed to annotations produced by the analysis tools of language technology, which arguably must influence their own thinking about linguistic matters. In the work on the Swedish constructicon, the analysis tools of language technology have been brought to bear on the task of extending the Swedish constructicon – and also the Swedish FrameNet (Dannélls, Friberg Heppin, & Ehrlemark, 2014; Johansson, 2014) – with new entries. This work is described in the next section.

The "turtles-all-the-way-down" notion that language has nothing but constructions may be intellectually satisfying and theoretically enlightening, but not obviously immediately compatible with the mechanics of practical language processing systems, which minimally assume a set of word-level units combining into phrase-level (and/or clause/sentence-level) units, and where the mechanisms for dealing with the two levels normally are quite different. This contradiction may be more imagined than real, however. At the level of their genetic machinery, all organisms on Earth are of the same kind (except maybe prions), but for many (most?) practical purposes we do not want to treat amoebas as being the same as elephants or chihuahuas. Against this background and also because of the many far-from successful attempts to build practical language processing systems based exclusively on linguistic formalisms, we may think of more pragmatically feasible

modes of drawing on construction grammar in such systems. One such mode could be as a way of reducing ambiguity in analyses. Especially computational syntactical analysis is plagued by immense ambiguity. In pure grammar-based systems this manifests itself as many analyses – hundreds of syntax trees even from fairly short sentences – whereas data-driven systems – which typically return only the most probable analysis – will suffer a general drop in performance as the number of theoretically possible alternatives grows.[5] The Swedish analysis system described below in Section 4.1 introduces information from the Swedish constructicon in its grammar in order not to have to distribute (seeming) noncompositionality among the components of constructions, as it were.

3.   The role of language technology in constructicon population

There is also an undeniable sociological component to the interaction of language technology and linguistics, in the sense that this presupposes interaction among researchers in the respective fields, or rather interaction between researchers trained in computer science and those trained in linguistics. This seems to be happening increasingly rarely in language technology, as the observations cited above reveal. A happy exception to this trend, the work on the Swedish constructicon has benefitted from a long history of linguistically informed language technology research conducted in Gothenburg in close collaborations involving Swedish and general linguists and computer scientists.

Thus, the Swedish constructicon project was conceived and executed in a context where work on a lexical macroresource for Swedish language technology was well underway and crucially included a framenet for Swedish as a central component. This meant that the new project could draw on the same kind of language technology support for building the constructicon, that was being developed for the other resources, including the Swedish FrameNet. This context is described in the next section, and the general infrastructure developed in the collaboration for building the constructicon and editing constructicon entries is discussed in Section 3.2.

The most central contribution of language technology in the Swedish constructicon project has been the application of tools for automatic language analysis to the problem of finding construction candidates in large text corpora. This work is described in Section 3.3.

---

5.   Although supervised data-driven systems will have the advantage that they will not be exposed to the full theoretical range of analyses, since a large share of these will not occur in the training data at all.

3.1   Towards a lexical macroresource for Swedish language technology

The Swedish FrameNet++ (SweFN++) project (Borin et al., 2010) was many things simultaneously. Its main goal was the creation of an integrated lexical macroresource for Swedish to be used as a basic infrastructural component in Swedish language technology research and in the development of natural language processing (NLP) applications for Swedish. A significant result of the project is the Swedish FrameNet, containing almost 40,000 lexical units, making it the world's most extensive framenet on this measure.

The macroresource is topologically a hub-and-spokes structure. There is one primary, central lexical resource, a pivot, to which all other resources are linked. This is SALDO (Borin & Forsberg, 2009; Borin, Forsberg, & Lönngren, 2013), a large (ca. 147K entries and 2M wordforms), freely available (under a Creative Commons Attribution license) morphological and lexical-semantic lexicon for modern Swedish. It has been selected as the pivot partly because of its size and quality, but also because its form and sense units are identified by carefully designed unique *persistent identifiers* (PIDs) to which the lexical information in other resources are linked.

As a semantic lexicon, SALDO is a kind of lexical-semantic network, superficially similar to Princeton WordNet (Fellbaum, 1998), but quite different from it in the principles by which it is structured. The basic organizational principle of SALDO is hierarchical. Every entry in SALDO – representing a word sense – is supplied with one or more semantic descriptors, which are themselves also entries in the dictionary. All entries in SALDO are actually occurring words or conventionalized or lexicalized multi-word expressions (MWEs) of the language.[6] No attempt is made to fill perceived gaps in the lexical network using definition-like paraphrases, as is sometimes done in WordNet (Fellbaum, 1998, 5f). One of the descriptors, called primary, is obligatory. The primary descriptor is the entry which better than any other entry fulfills two requirements: (1) it is a semantic neighbor of the entry to be described; and (2) it is more central than it. Both these aspects need some clarification.

---

6.   In order to make SALDO into a single hierarchy, an artificial entry, called PRIM, is used as the primary descriptor of 45 semantically unrelated entries at the top of the hierarchy, making all of SALDO into a single rooted tree. Here we may also add that SALDO contains words of *all* parts of speech, again distinct from WordNet, which only includes the four open word classes nouns (including proper nouns), verbs, adjectives, and adverbs. About 5% of the entries in SALDO are MWEs, representing all parts of speech. An interesting methodological question in this connection – to which no definitive answer has been found – is how Swedish constructicon entries should relate to SALDO MWE entries.

That two entries are semantic neighbors means that there is a direct semantic relationship between them, for instance synonymy, hyponymy, argument–predicate relationship, etc. Centrality is determined by means of several criteria, the most important being frequency: a frequent entry is more central than an infrequent entry.[7] The basic linguistic idea underlying SALDO is in effect that, semantically speaking, the whole vocabulary of a language can be described as having a center – or core – and (consequently) a periphery. The notion of core vocabulary is familiar from several linguistic subdisciplines (Borin, 2012). In SALDO, the higher levels in the hierarchy contain simpler and more basic entries. Contrast this with WordNet, where the higher nodes in the hierarchy contain very abstract vocabulary (e.g. 'entity').

Below, we give a few examples of entries with their primary and secondary descriptor senses:

balkong – hus	'balcony (n)' – 'house (n)'
Berlin – stad + Tyskland	'Berlin (prop)' – 'city (n)' + 'Germany (prop)'
berusa – yr	'intoxicate (v)' – 'dizzy (adj)'
bete – fånga	'bait (n)' – 'catch (v)'
bete² – beta	'grazing land (n)' – 'graze (v)'
bete³ – tand + djur	'tusk (n)' – 'tooth (n)' + 'animal (n)'
bröd – mat + mjöl	'bread (n)' – 'food (n)' + 'flour (n)'
brödföda – uppehälle	'daily bread (n)' – 'subsistence (n)'
bröllop – gifta sig	'wedding (n)' – 'get married (v)'

The standard scenario for a lexical resource to be integrated into the macroresource is to (partially) link its entries to the sense PIDs of SALDO, via their citation form and part of speech. This typically leads to ambiguity, since most of the resources associate lexical information to part-of-speech-tagged lemmas, and not to word senses. Some of these ambiguities can be resolved automatically – especially if information from several resources is combined – but in the end, manual work is required for complete disambiguation. However, like many other linguistic phenomena, the distribution of senses over citation forms in lexical resources is roughly Zipfian (Moon, 2000; Borin, 2010). Thus, the vast majority of the lemmas

---

7. The actual work on SALDO relies mainly on the lexicographical experience and linguistic intuition of the compilers, who use clues such as stylistic value, word-formation complexity, the type of semantic relation holding between an entry and its primary descriptor, acquisition order in first-language acquisition, etc. Frequency correlates highly with these, however: It turns out that about 90% of the SALDO entries have primary descriptors which are at least as frequent as the entries themselves in a corpus of more than one billion words of Swedish. A more detailed description and discussion of the semantic organization of SALDO can be found in Borin, Forsberg, and Lönngren (2013, pp. 1196–1200).

are monosemous, reducing the sense mapping problem to the much simpler problem of pairing up forms between lexical resources.

The Swedish constructicon project uses the lexical infrastructure developed in the SweFN++ project and maintained as part of Språkbanken's (the Swedish Language Bank) LT infrastructure (see the next section), including the use of SALDO word-sense identifiers. This means that the constructicon shares all the linguistic information available in the component resources, which opens for exciting opportunities both in NLP and linguistics. A concrete example: The macroresource also includes historical lexical resources (Borin, Forsberg, & Kokkinakis, 2010; Borin & Forsberg, 2011), the oldest describing Old Swedish (13th–16th c.). The hope is that a successful (but possibly partial) linking of SALDO to the historical lexicons will make it possible to project various linguistic information from the modern resources onto the historical resources, allowing, e.g., identification of possible counterparts of modern constructions in historical varieties of Swedish, making it possible to study the diachronic development of selected constructions in empirical corpus data.

3.2   A general lexical infrastructure and a language-aware lexicon editor

The activities described above also have a more technical side, a generalized lexical infrastructure, called Karp (Borin et al., 2012). The heart of the lexical infrastructure is the lexical macroresource described in the previous section, a large network of interconnected lexicons (Borin et al., 2010; Borin, Forsberg, & Lyngfelt, 2013), all encoded in the ISO Lexical Markup Framework (LMF) format (ISO, 2008; Francopoulo, 2013).

Even though the lexical macroresource is primarily intended for use in LT applications, its component lexicons are still very much lexicographical entities. Thus, from a linguistic point of view, the work on individual resources as well as on their integration is at heart a genuinely lexicographical activity, to boot one with considerable potential to make significant theoretical contributions to lexicology, lexical semantics and lexical typology because of the large-scale empirical nature of our endeavor and the diversity of the lexical resources involved. In general, working with large amounts of data as we do, requires good tools. The Karp lexical infrastructure has been designed with this in mind. It supports the work on creating, curating, and integrating the lexical resources, and the maintenance in parallel of online-browsable and downloadable versions of the resources. An important feature of the lexical infrastructure is that we maintain a strong bidirectional connection to our corpus infrastructure Korp (Borin, Forsberg, & Roxendal, 2012). For example, the corpora are annotated with the lexical information available through

Karp, and the language examples for the lexical resources in Karp are retrieved from Korp. Similarly, corpus frequencies for lemmas and wordforms are provided by Korp.

In the Swedish constructicon project a dedicated constructicon editor was developed in Karp, which has contributed greatly to achieving an efficient workflow in the project (see Lyngfelt, Bäckström et al., this volume).

3.3 Mining corpora for construction candidates with language tools

As discussed by Lyngfelt, Bäckström, et al. (this volume), the work on the Swedish constructicon has focused on *partly schematic constructions*, i.e., constructions containing some fixed lexical element or elements, such as the construction `i_adjek-tivaste_laget` illustrated in their Example (1), repeated here slightly modified as (1). In (1), the construction is set in boldface, and the fixed parts of the construction are underlined.

(1) Jag ska erkänn-a att det här är   en kladdkaka   **i**
    I   shall admit-INF that it here be.PRS a sticky.chocolate.cake **in**
    söt-<u>ast-e</u>   <u>**lag-et**</u>
    sweet-SUV-DEF **measure-DEF**
    'I must admit that this is a sticky chocolate cake **a bit too** **sweet**'

There has been a considerable amount of recent work on MWEs in language technology and in corpus linguistics. In the latter field MWEs are sometimes referred to under other terms such as *formulaic language* (Wray 2002) or *lexical bundles* (Biber & Conrad 1999). While linguists have devoted much energy to classifying and characterizing MWEs in general (but as a rule not cross-linguistically informed) terms, much of the language technology research has focused on developing data-driven methods for finding certain subtypes of MWEs in text (see Pecina, 2010, for a good overview). Since some MWE types recognized in the literature are structurally similar to the schematic constructions making up the Swedish constructicon, the Swedish constructicon team has experimented with methods inspired and informed by MWE research in language technology in order to generate construction candidates from linguistically annotated corpora (Bäckström et al., 2013; Forsberg et al., 2014), mentioned by Lyngfelt, Bäckström, et al. (this volume) and explained and motivated in more detail in what follows.

One of the goals of the Swedish constructicon project has been to develop tools for automatic identification of constructions in authentic texts. This is a highly desirable research objective in itself, with potential uses in a number of NLP applications. In addition, the same methods provide the project with a heuristic tool.

By automatically extracting various kinds of regularities in texts, we may discover patterns that might otherwise have been overlooked. This especially concerns seemingly insignificant constructions that do not stand out against the context the way spectacular idioms do. The resulting findings are treated as construction candidates, a subset of which may be considered actual constructions after manual evaluation.

Language technology is both a means and a goal for the Swedish constructicon project. The experiments presented by Bäckström et al. (2013) and Forsberg et al. (2014), turned out to be valuable means for identifying potential constructions; the method used in the experiments provides construction candidates, statistically identified recurring linguistic structures, which are then manually evaluated to filter out material for actual constructicon entries. In a longer perspective, we are also working towards developing methods for automatic identification of particular constructions, which is a desirable research objective in itself, with potential for improving automatic language analysis systems.

The experiments were set up and executed using the resource infrastructure of Språkbanken, a modular and interoperable set of resources and tools in the form of web services for accessing, browsing, editing and automatically annotating resources.

Concentrating on the partly schematic constructions targeted by the Swedish constructicon as well as on purely schematic constructions, the method chosen for suggesting potential constructions consists of two steps: (1) extracting and counting syntactic patterns occurring in a large corpus; (2) ranking those patterns by a relevance measure based on a hypothesis about what characterizes a construction from a statistical point of view.

The aim was to capture constructions which can be described as a sequential pattern – a so-called *n-gram* – of *n* adjacent units: specific words, words with a particular part-of-speech (POS) tag, or phrases. This heuristic covers many important grammatical constructions, in particular the most common basic phrases.

The method works by going through all text in a large linguistically annotated corpus (1 million words in Bäckström et al., 2013 and 19 million words in Forsberg et al., 2014), taking each successive word as the first word in a number of sequences up to a maximum length. For each such sequence, all patterns of which that sequence is an instance are generated and counted. For the two-word sequence *in London*, there are five such patterns. First, there are three ways to form a pattern by replacing words by POS tags:

> in [PROPER NOUN]
> [PREPOSITION] London
> [PREPOSITION] [PROPER NOUN]

Furthermore, since *London* is also an NP, there are two more generalizations in terms of that phrase:

> in [NP]
> [PREPOSITION] [NP]

These patterns are similar to the "hybrid n-grams" used by Wible & Tsao (2010) but more general since they can involve phrases.

The NLP aspect of this work was provided by the NLP tools available through Språkbanken's infrastructure. POS tags were assigned using HunPos (Halácsy, Kornai, & Oravecz, 2007) trained on the Stockholm–Umeå corpus (Gustafson-Čapková & Hartmann, 2006), and phrases were assigned on the basis of automatic syntactic analyses provided by MaltParser (Nivre et al., 2007) trained on the Swedish Treebank (Nivre, et al., 2008). MaltParser outputs dependency structures, which were converted into phrase structure representations using heuristics, e.g., a dependency subtree dominated by a common or proper noun or a pronoun becomes an NP.

Finally, the extracted patterns were ranked based on a statistical collocation measure, with due consideration of shorter patterns included in longer ones, and also only including patterns which had the same start and end point as some phrase. In the experiment described by Forsberg et al. (2014), the 1,200 most highly ranked patterns were then presented to three linguists for evaluation as to their suitability as construction candidates, together with the (up to) five most frequent instances of each pattern. In total, about 200 promising construction candidates were found, a lot considering that the automated part of the experiment took very little time to set up and execute, thanks to the well-developed LT infrastructure available, and that consequently the manual effort of the linguists was limited to going through a maximum of 7,200 examples (1,200 patterns plus a maximum of five instances of each), rather than 19 million words of text.[8]

There are many directions in which these experiments could be extended, such as:

– Using larger corpora to fight the sparse-data problem
– Using different text types in order to capture domain-specific constructions
– Working with dependency subtrees instead of word sequences

---

8.   Of course, we cannot know what has been missed by this search and filtering procedure, but as far as we can see, there is no realistic alternative way of finding this out, when the input data comprises millions or even billions of words of text.

4. Using constructicons in language technology systems

Whereas frame-semantic representations have seen a fair amount of interest from the language technology community, as a practical, light-weight semantic representation suitable for deployment in NLP systems, the construction grammar formalism sprung from the same intellectual source has received much less attention.

The main computational implementation of construction grammar, Embodied Construction Grammar (e.g., Chang & Maia, 2001; Bergen & Chang, 2005, 2013; Bryant, 2008; Chang, 2008; Schneider, 2010), does not in fact represent an effort to build a practical language-processing system based on construction grammar. Instead, it reflects an older tradition in computational linguistics, where computer simulation is seen as a productive way of investigating human linguistic behavior, by devising "computational implementations of cognitively motivated theories of morphology, metonymy, metaphor, generation, and mental spaces" (Bryant 2008, p. 214).

Just as in the case of frame semantics, in order for it to be considered at all for inclusion in an NLP system, a constructicon must have a high coverage in the domain where the system is to be applied. This, together with a general lack of awareness in the language technology community about construction grammar, makes it hard to believe that we will see construction-based practical NLP systems anytime soon. However, we are already seeing construction information being incorporated in more "conventional" NLP systems, a very exciting and promising development. The next section provides a concrete example of how the Swedish constructicon has been used in exactly this way.

4.1 Using the Swedish constructicon for language analysis

Although the Swedish constructicon (SweCcn) is primarily created for linguistic research, with the help of language technology, language technology itself can benefit considerably from formalized, computational constructions that make the non-compositional parts of the grammar-lexicon interface compositional again. Figure 1 illustrates this with a simple example where the Swedish construction `behöva_något_till_något` 'need *something* for *something*' is recognized as a part of the computed compositional abstract syntax tree of the given clause. Moreover, the construction is fully embedded in the grammatical analysis: it produces a regular verb phrase (VP), and it takes regular noun phrases (NPs) – the variable parts of the construction – as arguments. The fixed parts of the construction (the verb *behöva* 'need' and the preposition *till* 'to/for') are not included in the abstract interlingual analysis; their inclusion and surface realization are left to the language-specific grammar (the Swedish grammar in our case).
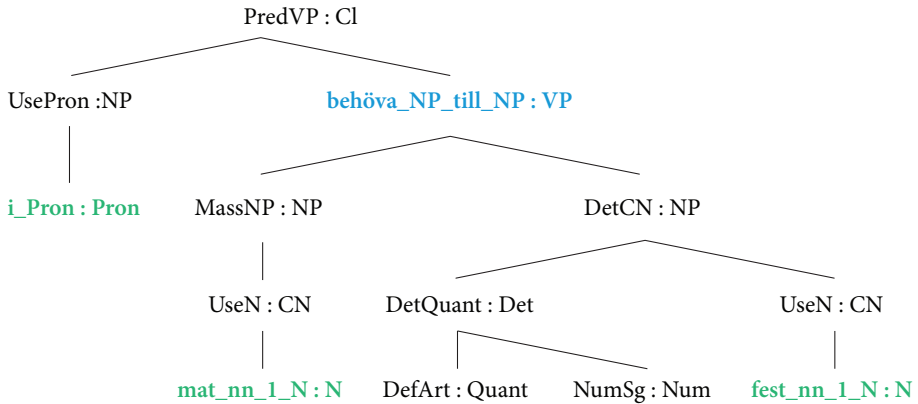
Figure 1. An abstract syntax tree for a clause containing an embedded construction: *jag behöver mat till festen* 'I need food for the party'

The role of computational constructions and potentially multilingual construc- ticons becomes more apparent in machine translation as illustrated in Figure 2. The preposition *till* in `behöva_något_till_något` literally translates as 'to', but its meaning in this construction is 'for'. In contrast to fixed multi-word expressions, it is impossible to list all instances of this construction in a translation lexicon. Statistical machine translation can theoretically handle such simple constructions, but it often fails because of insufficient training data.
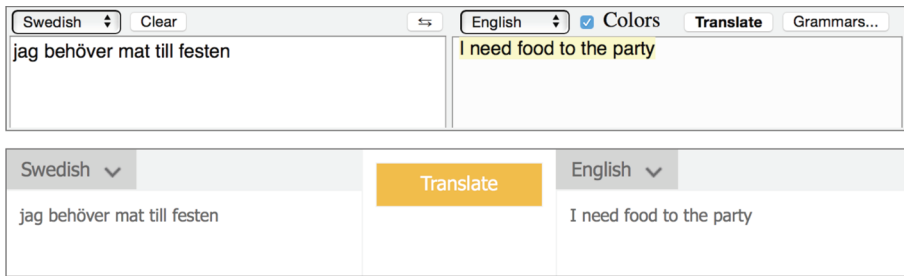
Figure 2. A simple example that illustrates the need for a multilingual constructicon in both rule-based machine translation (the upper screenshot of GF Wide Coverage Translator) and statistical machine translation (the lower screenshot of Bing Translator)

Since the current number of annotated examples per construction is relatively low in the Swedish constructicon, while the level of abstraction in the construction descriptions is relatively high, we have begun by converting the linguistically ori- ented descriptions into Grammatical Framework (GF), a computational grammar formalism (Ranta, 2004). The aim of formalizing the Swedish constructicon in GF

is twofold: (i) to obtain a more precise and consistent insight into the types and descriptions of Swedish constructions, (ii) to implement and test an extension to an existing grammar-based parser for Swedish, so that the parser would be able to recognize constructions as part of the sentence analysis and thereby reduce ambiguities. This will also facilitate the future development of construction-aware data-driven NLP systems.

In the rest of this section, we outline our methodology on how to systematically formalize the semi-formal representation of the Swedish constructicon in GF, showing that a computational GF construction grammar can be, to a large extent, acquired automatically. A side result of our approach is that it has been helpful for improving the consistency of the Swedish constructicon and for characterizing and annotating construction entries in the database of Swedish constructions.

4.2 The database of Swedish constructions

The Swedish constructicon follows the view that a construction is the basic linguistic unit in language (Goldberg, 1995; Croft, 2001). According to this view there is no strict separation between the lexicon and grammar. As mentioned by Lyngfelt, Bäckström et al., (this volume) and also adhered to here, constructions in the Swedish constructicon are partially schematic multi-word expressions with fixed and variable slots. Information about construction entries is encoded in an ISO LMF database that is integrated into the lexical infrastructure of Språkbanken described in Section 3.1 (see also Lyngfelt et al., 2012). This information includes: a descriptive name of the unit, a grammatical category, a free-text definition, a set of annotated example sentences extracted from Korp, a structure sketch (i.e., a formal description) of both the morphosyntactic structure, and the internal and external construction elements (CEs). Internal CEs are part of the construction proper, and external CEs are valency bound elements of the construction.

In the structure sketches, internal CEs are delimited by brackets, where alternative values and lexical units (LUs) are separated by a bar, for example the structure sketch for the SweCcn entry `behöva_något_till_något` is: [behöva[1] NP_1 till[1] NP_2 | VP]. This sketch combines four internal CEs, two lexically fixed elements and two variable slots of which one contains two alternative values: an NP and a VP. Each CE is further described with a feature matrix containing several attributes, including: (a) the lexical unit's identifier from SALDO (see Section 3.1) for the lexically fixed elements; (b) the semantic role of each element (except preposition elements); (c) the name of the element; and (d) the grammatical category (either part-of-speech tag or phrase type). An example of the feature matrix for the internal CEs of `be-höva_något_till_något` taken from the SweCcn database is:

{role = "State" name = "State" cat = "V" lu = "behöva..1"}
{role = "Requirement" name = "Requirement" cat = "NP"}
{name = "P" cat = "P" lu = "till..1"}
{role = "Purpose" name = "Purpose" cat = "NP|VP"

As the above example shows, the set of feature matrices is rather detailed but there is no explicit indication of the corresponding elements which are given in the structure sketch. In many matrices the order of the specifications for each element tends to be diverse; it does not necessarily follow the element specification given in the structure sketch. When we started analyzing the database for language technology analysis, systematic recording of the different specifications and their corresponding elements became very relevant for improving the automatic analysis.

Each construction bears a grammatical category. As of August 2016 there are nine categories specified in the database, viz. AdvP (Adverb Phrase), AP (Adjective Phrase), AP|AdvP (Adverb Phrase or Adjective Phrase), Intj (Interjection), NP (Noun Phrase), PP (Prepositional Phrase), S (Sentence), VP (Verb Phrase), XP (any phrase type). In our work we chose to begin with constructions of category VP mainly because this category predominates in the SweCcn database, with more than 100 constructions available, but also because VP constructions are expressed by complex internal structures.

4.3   Grammatical framework

Grammatical framework (GF; Ranta, 2004) is a categorial grammar formalism and a framework for implementing computational grammars. It provides built-in support for multilingual grammars, which holds great potential for implementing, unifying and interlinking constructions of different languages.

GF is characterized by its two-level approach to natural language representation: abstract syntax which defines the language-independent structure, and concrete syntax which defines the language-specific syntactic and lexical realization of the abstract syntax. The same abstract syntax can correspond to many (multilingual) concrete syntaxes – mappings from abstract syntax trees to feature structures and strings. GF grammars are bi-directional – they can be used for both parsing and language generation. The framework is suitable for implementing general-purpose syntactic grammars as well as domain-specific semantic grammars.

Notably, GF provides a general purpose resource grammar library, RGL (Ranta, 2009), for currently 30 languages that implement the same abstract syntax. The RGL has a high-level interface that provides constructors like `mkVP : V → NP → VP` for building a verb phrase from a verb and a noun phrase without the need of specifying low-level details like inflectional paradigms, syntactic agreement and word

order. These details are handled by the language-specific resource grammars. The coverage of the general-purpose Swedish grammar is one of the largest in the GF RGL comprising a lexicon with over 100,000 lexical entries from SALDO.

4.4 Constructing a computational constructicon

Our method of converting the lexicographic SweCcn entries into a computational GF construction grammar comprises several steps:

1. preprocessing: automatic normalization, consistency checking and rewriting of the structure sketches;
2. automatic generation of the abstract and concrete syntaxes of a GF grammar;
3. semi-automatic verification of the acquired grammar.

Constructions may have optional CEs, alternative types of CEs or alternative LUs, and even alternative word order. In the structure sketches, optional CEs are delimited by parentheses, and alternative types/LUs are separated by a vertical bar, e.g.:

(2) a. `behöva_något_till_något`: [behöva^1 NP$_1$ till^1 NP$_2$|VP]
   e.g. <u>behöva</u> kvällen <u>till</u> att plugga '<u>need</u> the evening to study'

(3) a. `verba_av_sig.transitiv`: [V av^1 Pn$_{refl}$ (NP)]
   e.g. ta <u>av mig</u> skorna 'take <u>off myself</u> the shoes'

(4) a. `snacka_NP`: [snacka^1|prata^1|tala^1 NP$_{indef}$]
   e.g. <u>prata</u> skolminnen '<u>talk</u> school memories'

(5) a. `få_resultativ.agentiv`: [få^1 NP PcP]
   e.g. <u>få</u> gräsmattan klippt '<u>get</u> the lawn trimmed'

(6) a. `x-städa`: [N|Adj+städa^1]
   e.g. stor<u>städa</u> 'big<u>clean</u> (V)'

The variable CEs may have indices denoting difference, formal identity (repetition), co-reference, etc. In the case of a lexical construction that is realized by a compound word, its internal CEs are delimited by the plus sign indicating concatenation. Suffixing (as in (1) in Section 3) is indicated by a hyphen.

The automatic preprocessing of constructicon entries comprises:

1. Normalization of the structure sketches and attribute values in the feature matrices, fixing a lot of various inconsistencies due to the manual annotation.
2. In case of optional CEs and alternative types of CEs, there are formally several constructions compressed into one. The original structures are rewritten so that for each combination there is a separate alternative structure. This however does not apply to alternative LUs. If a CE is represented by a fixed set of LUs, we

assume that they are interchangeable (synonymous). Otherwise they should be either split into alternative constructions (separate entries), or the CE should be made more general (variable).

3. The rewritten structure sketches are enriched with additional morphosyntactic information from the feature matrices, so that a complete description is at hand.

4. The grammatical categories used in the Swedish constructicon are converted into GF categories.[9] In specific cases, the conversion may lead into a more general or more specific description as well as it may include morphosyntactic tags and may depend on the contextual CEs. This requires a subsequent rewriting of the whole construction. A few categories, however, are not converted at this step; their conversion is postponed to the generation of the GF grammar. For instance, Pc (participle) and PcP (participle phrase) are not converted to V (verb) and VP respectively, as they have to be treated differently in the concrete syntax: PcP is a VP that is eventually converted to AP (adjectival phrase) or Adv (VP-modifying adverb) as illustrated by `få_resultativ.agentiv` (5).

Below in (2b)–(6b) are given rewritten structural descriptions of the sample constructions introduced above in (2a)–(6a). Note that we ignore the SALDO sense identifiers; the word sense information is not used so far in generating the grammar.

(2) b. behöva_V NP_1 till_{Prep} NP_2 | behöva_V NP till_{Prep} VP

(3) b.   V av_{Prep} Pron_{refl} NP | V av_{Prep} Pron_{refl}

(4) b. $\text{snacka|prata|tala}_V$ aSg_Det CN | $\text{snacka|prata|tala}_V$ aPl_Det CN | $\text{snacka|prata|tala}_V$ CN

(5) b. få_V NP PcP_{perf}

(6) b.   N + städa_V | A + städa_V

The rewritten structural descriptions of constructions provide sufficient information to generate both the abstract and the concrete syntax of a constructicon-based grammar, an extension to the existing wide-coverage Swedish GF resource grammar.

The generation of the abstract syntax is rather straightforward. Each construction is represented by one or more grammar rules (functions) depending on how many alternative structure descriptions are produced (rewritten) in the preprocessing phase. For VP constructions, the average number of alternative functions per construction is 1.4 while the maximum number is 6, produced by `snacka_NP.emfas`: [$\text{snacka}^1|\text{prata}^1$ (AP) NP_{indef}].

Each function takes one or more arguments that correspond to the variable CEs of the respective alternative construction description. The fixed CEs are not

---

represented by the abstract syntax. The variable CEs are represented only by their grammatical categories; other morphosyntactic constraints (if any) are handled by the language-specific concrete syntax.

The rewritten structure descriptions shown above in (2b)–(6b) are represented by the following abstract functions in the GF construction grammar (2c)–(6c):

(2) c. behöva_något_till_något$_1$: NP → NP → VP
        behöva_något_till_något$_2$: NP → VP → VP

(3) c. verba_av_sig_transitiv$_1$: V → NP → VP
        verba_av_sig_transitiv$_2$: V → VP

(4) c. snacka_NP$_1$: CN → VP
        snacka_NP$_2$: CN → VP
        snacka_NP$_3$: CN → VP

(5) c. få_resultativ_agentiv: NP → VP → VP

(6) c. x_städa$_1$: N → VP
        x_städa$_2$: A → VP

As for the concrete syntax, many constructions can be implemented in GF by systematically applying the high-level constructors provided by the GF RGL:[10]

mkVP:	VP → Adv → VP
mkVP:	V2 → NP → VP
mkV2:	V → V2
mkV:	Str → V
mkAdv:	Prep → NP → Adv
mkPerp:	Str → Prep
etc.	

For instance, behöva_något_till_något$_1$ can be implemented by first making a two-place verb (V2) from V, and then combining it with the first NP into a VP; the preposition can be combined with the second NP into a prepositional phrase (Adv) which can then be attached to the VP:

(2) d. behöva_något_till_något$_1$ np$_1$ np$_2$ =
            mkVP
                (mkVP (mkV2 (mkV "behöver")) np$_1$)
                (mkAdv (mkPrep "till") np$_2$)

The question is how to make such constructor applications systematically and, thus, automatically, given the various construction descriptions. Essentially, this is a parsing problem in itself: we can look at CEs as words in the formal construction description language for which we need a grammar to combine the lists of CEs into the parse trees of RGL constructors and their arguments. We have defined such an auxiliary GF grammar to generate the implementation of the abstract functions in a GF construction grammar. This approach has emerged from our work on creating a multilingual computational grammar for FrameNet (Grūzītis & Dannélls, 2017). The proposed approach to GF construction grammars can also be reused for other languages, and it is explained in more detail by Grūzītis et al. (2015).

The resulting module of the computational construction grammar yielded 127 GF functions: 1.4 alternative realizations of 93 VP constructions in SweCcn. In most cases, the automatically generated concrete (Swedish) implementation of the abstract GF functions is adequate. In general, however, the implementation requires manual validation and minor or major revisions. An example of a construction that was not generated successfully is disjunktiv_samordning.korr[varken[1]|vare_sig[1] XP$_1$ eller[1] XP$_2$]. Following the definition of the construction, the scope of negation extends over the coordinated sentences, either two coordinated sentences or more. In GF this is a special case of conjunction and therefore requires defining new operations to uncover dimensions of deep linguistic information, something that is very limited in an automatic process such as the one described here.

4.5 Preliminary analysis of the automatically generated computational constructicon

One possible evaluation procedure which we considered in this work was to parse some corpus data with the generated grammar and examine the results.

In the first step we compiled a benchmark collection of 337 example sentences from the annotated sentences of the VP constructions in SweCcn. The second step was to parse each sentence using the generated grammar. Several heuristics were applied before parsing to overcome problems such as: (i) lack of the subject which is necessary for generating a proper clause in GF, and (ii) missing lexical entries for proper nouns, verbs and compounds not covered in the GF lexicon for Swedish. Heuristics were formulated for inserting subjects, replacing compounds, proper nouns and verbs, and for changing the tense or the verb string.

Out of the 337 annotated example sentences, 281 turned out to have a corresponding concrete function in the construction grammar. The parser successfully parsed 80% of these sentences. Cases where the parser could not return any parse

trees were mainly due to annotation inconsistencies or errors in the SweCcn database, for instance, a feature matrix requires the singular form of an NP although the plural form exists among the annotated examples. This turned out to be helpful feedback to the linguists compiling the structural descriptions of the SweCcn constructions.

Most of the words the GF parser failed with were proper nouns and compounds. With respect to the grammar, the parser failed in cases where there existed ill-formed or syntactically complex sentences, often containing coordination and conjunctions and long sentences, containing irrelevant phrases and punctuation that fall outside the construction. In the future, the grammar analysis can be complemented with "inverse testing", i.e., we can use the GF built-in support for automatically generating a treebank and test whether our grammar is able to provide correct syntactic analyses for each construction. The SweCcn developers (the linguists) could in turn further inspect and validate the generated syntactic trees.

5. Conclusion and outlook

In this chapter, we have described the close interaction of linguists and language technologists in the Swedish constructicon project. The collaboration has been very successful, and – importantly – constituted a genuine instance of cross-fertilization, where an evolving LT infrastructure has formed an integral part of the constructicon development environment, while at the same time the structured linguistic knowledge described in the constructicon has informed the language technology making up the infrastructure.

From the point of view of the infrastructure, we have seen an overall quality improvement resulting from corrections of lots of bare syntactic categories incorrectly labeled in the constructicon database. From the point of view of the grammar formalism, we have seen improvements emerging as a result of reduced ambiguity in the parsing analyses.

We believe that the fact that the closely interlinked Swedish lexical macroresource now also includes a constructicon will allow linguists to conduct empirical research into the synchronic and diachronic properties of Swedish constructions on the basis of Språkbanken's vast and varied text corpora.

References

Abney, S. (2011). Data-intensive experimental linguistics. *Linguistic Issues in Language Technology* 6.

Angelov, K., & Ranta, A. (2009). Implementing controlled languages in GF. In N. E. Fuchs (Ed.), *Controlled natural language* (pp. 82–101). Berlin: Springer.

Bäckström, L., Borin, L., Forsberg, M., Lyngfelt, B., Prentice, J., & Sköldberg, E. (2013). Automatic identification of construction candidates for a Swedish constructicon. In *Proceedings of the workshop on lexical semantic resources for NLP at NODALIDA 2013* (pp. 2–11). Oslo: NEALT.

Bender, E. (2011). On achieving and evaluating language-independence in NLP. *Linguistic Issues in Language Technology* 6.

Bergen, B. K., & Chang, N. (2005). Embodied Construction Grammar in simulation-based language understanding. In J. -O. Östman & M. Fried (Eds.), *Construction grammars: Cognitive grounding and thoretical extentions* (pp. 147–190). Amsterdam: John Benjamins. doi:10.1075/cal.3.08ber

Bergen, B. K., & Chang, N. (2013). Embodied Construction Grammar. In T. Hoffman & G. Trousdale (Eds.), *The Oxford handbook of construction grammar* (pp. 168–190). Oxford: Oxford University Press.

Biber, D., & Conrad, S. (1999). Lexical bundles in conversation and academic prose. In H. Hasselgard & S. Oksefjell (Eds.), *Out of corpora: Studies in honor of Stig Johansson* (pp. 181–189). Amsterdam: Rodopi.

Bollmann, M. (2013). POS Tagging for historical texts with sparse training data. In *Proceedings of the 7th Linguistic Annotation Workshop and Interoperability with Discourse* (pp. 11–18). Sofia: ACL.

Borin, L. (2010). Med Zipf mot framtiden – en integrerad lexikonresurs för svensk språkteknologi [With Zipf into the future – an integrated lexical resource for Swedish language technology]. *LexicoNordica* 17, 35–54.

Borin, L. (2012). Core vocabulary: A useful but mystical concept in some kinds of linguistics. In D. Santos, K. Lindén, & W. Ng'ang'a (Eds.), *Shall we play the Festschrift game? Essays on the occasion of Lauri Carlson's 60th birthday* (pp. 53–65). Berlin: Springer. doi:10.1007/978-3-642-30773-7_6

Borin, L., Dannélls, D., Forsberg, M., Gronostaj, M. T., & Kokkinakis, D. (2010a). The past meets the present in Swedish FrameNet ++. In *14th EURALEX International Congress* (pp. 269–281). Leeuwarden: EURALEX.

Borin, L., & Forsberg, M. (2009). All in the family: A comparison of SALDO and WordNet. In *Proceedings of the Nodalida 2009 Workshop on WordNets and other Lexical Semantic Resources – between Lexical Semantics, Lexicography, Terminology and Formal Ontologies* (pp. 7–12). Odense. NEALT.

Borin, L., & Forsberg, M. (2011). A diachronic computational lexical resource for 800 years of Swedish. In C. Sporleder, A. Van Den Bosch, & K. Zervanou (Eds.), *Language technology for cultural heritage* (pp. 41–61). Berlin: Springer. doi:10.1007/978-3-642-20227-8_3

Borin, L., Forsberg, M., & Kokkinakis, D. (2010b). Diabase: Towards a diachronic BLARK in support of historical studies. In *Proceedings of LREC 2010* (pp. 35–42). Valletta: LREC.

Borin, L., Forsberg, M., & Lönngren, L. (2013). SALDO: a touch of yin to WordNet's yang. *Language Resources and Evaluation* 47(4), 1191–1211. doi:10.1007/s10579-013-9233-4

Borin, L., Forsberg, M., & Lyngfelt, B. (2013). Close encounters of the fifth kind: Some linguistic and computational aspects of the Swedish FrameNet ++ project. *Veredas* 17(1), 28–43. doi:10.1007/s10579-013-9233-4

Borin, L., Forsberg, M., Olsson, L. -J., & Uppström, J. (2012). The open lexical infrastructure of Språkbanken. In *Proceedings of LREC 2012* (pp. 3598–3602). Istanbul. ELRA.

Borin, L., Forsberg, M., & Roxendal, J. (2012). Korp – the corpus infrastructure of Språkbanken. In *Proceedings of LREC 2012* (pp. 474–478). Istanbul. ELRA.

Briscoe, T., Carroll, J., & Watson, R. (2006). The second release of the RASP system. In *Proceedings of the COLING/ACL 2006 interactive presentation sessions* (pp. 77–80). Sydney: ACL. doi:10.3115/1225403.1225423

Bryant, J. E. (2008). Best-fit constructional analysis. UC Berkeley PhD dissertation.

Chang, N. (2008). Constructing grammar: A computational model of the emergence of early constructions. UC Berkeley PhD dissertation.

Chang, N., & Maia, T. (2001). Learning grammatical constructions. In J. D. Moore & K. Stenning (Eds.), *Proceedings of the twenty-third annual conference of the Cognitive Science Society* (pp. 176–181). Mahwah: Lawrence Erlbaum.

Copestake, A. (2002). *Implementing typed feature structure grammars*. Stanford: CSLI Publications.

Croft, W. A. (2001). *Radical construction grammar: Syntactic theory in typological perspective*. Oxford: Oxford University Press. doi:10.1093/acprof:oso/9780198299554.001.0001

Dannélls, D., Friberg Heppin, K., & Ehrlemark, A. (2014). Using language technology resources and tools to construct Swedish FrameNet. In *Proceedings of the workshop on lexical and grammatical resources for language processing* (pp. 8–17). Dublin: ACL. doi:10.3115/v1/W14-5802

Fellbaum, C. (Ed.). (1998). *WordNet: An electronic lexical database*. Cambridge, Mass.: MIT Press.

Forsberg, M., Johansson, R., Bäckström, L., Borin, L., Lyngfelt, B., Olofsson, J., & Prentice, J. (2014). From construction candidates to constructicon entries: An experiment using semi-automatic methods for identifying constructions in corpora. *Constructions and Frames* 6(1), 114–135. doi:10.1075/cf.6.1.07for

Francopoulo, G. (Ed.). (2013). *LMF: Lexical markup framework*. London/Hoboken, NJ: ISTE/Wiley. doi:10.1002/9781118712696

Goldberg, A. (1995). *Constructions: A construction grammar approach to argument structure*. Chicago: Chicago University Press.

Grūzītis, N., & Dannélls, D. (2017). A multilingual FrameNet-based grammar and lexicon for controlled natural language. *Language Resources and Evaluation* 51(1), 37–66.

Grūzītis, N., Dannélls, D., Lyngfelt, B., & Ranta, A. (2015). Formalising the Swedish Constructicon in Grammatical Framework. In *Proceedings of the grammar engineering across frameworks (GEAF) workshop* (pp. 49–56). Beijing: ACL. doi:10.18653/v1/W15-3307

Gustafson-Čapková, S., & Hartmann, B. (2006). Manual of the Stockholm Umeå Corpus version 2.0. Stockholm University.

Halácsy, P., Kornai, A., & Oravecz, C. (2007). HunPos – an open source trigram tagger. In *Proceedings of the 45th annual meeting of the Association for Computational Linguistics companion volume: Proceedings of the demo and poster sessions* (pp. 209–212). Prague: ACL.

ISO (2008). Language resource management–Lexical markup framework (LMF). *International Standard ISO* 24613.

Johansson, R. (2014). Automatic expansion of the Swedish FrameNet lexicon: Comparing and combining lexicon-based and corpus-based methods. *Constructions and Frames* 6(1), 92–113. doi:10.1075/cf.6.1.06joh

Liberman, M. (2009). The annotation conundrum. In *Proceedings of the EACL 2009 workshop on the interaction between linguistics and computational linguistics: Virtuous, vicious or vacuous?* (p. 2). Athens: ACL. doi:10.3115/1642038.1642040

Lyngfelt, B., Bäckström, L., Borin, L., Ehrlemark, A., & Rydstedt R. (this volume). *Constructicography at work: Theory meets practice in the Swedish Construction*.

Lyngfelt, B., Borin, L., Forsberg, M., Prentice, J., Rydstedt, R., Sköldberg, E., & Tingsell, S. (2012). Adding a constructicon to the Swedish resource network of Språkbanken. In *Proceedings of KONVENS 2012* (LexSem 2012 workshop) (pp. 452–461). Vienna: ÖGAI.

Manning, C. D. (2015). Last words: Computational linguistics and deep learning. *Computational Linguistics* 41(4), 701–707. doi:10.1162/COLI_a_00239

Moon, R. (2000). Lexicography and disambiguation: The size of the problem. *Computers and the Humanities* 34(1–2), 99–102. doi:10.1023/A:1002605522910

Nivre, J., Hall, J., Nilsson, J., Chanev, A., Eryiğit, G., Kübler, S., Marinov, S., & Marsi, E. (2007). MaltParser: A language-independent system for data-driven dependency parsing. *Natural Language Engineering* 13(2), 95–135.

Nivre, J., Megyesi, B., Gustafson-Čapková, S., Salomonsson, F., & Dahlqvist, B. (2008). Cultivating a Swedish treebank. In J. Nivre, M. Dahllöf & B. Megyesi (Eds.), *Resourceful language technology: Festschrift in honor of Anna Sågvall Hein* (pp. 111–120). Acta Universitatis Upsaliensis: Studia Linguistica Upsaliensia 7.

Pecina, P. (2010). Lexical association measures and collocation extraction. *Language Resources and Evaluation* 44(1–2), 137–158. doi:10.1007/s10579-009-9101-4

Peldszus, A., & Stede, M. (2013). Ranking the annotators: An agreement study on argumentation structure. In *Proceedings of the 7th linguistic annotation workshop and interoperability with discourse* (pp. 196–204). Sofia: ACL.

Pullum, G. (2007). Ungrammaticality, rarity, and corpus use. *Corpus Linguistics and Linguistic Theory* 3(1), 33–47. doi:10.1515/CLLT.2007.002

Pullum, G. (2009). Computational linguistics and generative linguistics: The triumph of hope over experience. In *Proceedings of the EACL 2009 workshop on the interaction between linguistics and computational linguistics: Virtuous, vicious or vacuous?* (pp. 12–21). Athens: ACL. doi:10.3115/1642038.1642042

Ranta, A. (2004). Grammatical Framework, a type-theoretical grammar formalism. *Journal of Functional Programming* 14(2), 145–189. doi:10.1017/S0956796803004738

Ranta, A. (2009). The GF Resource Grammar Library. *Linguistic Issues in Language Technology (LiLT)* 2(2).

Reiter, E. (2007). The shrinking horizons of computational linguistics. *Computational Linguistics* 33(2), 283–287. doi:10.1162/coli.2007.33.2.283

Schmid, H. -J. (2010). Entrenchment, salience, and basic levels. In D. Geeraerts & H. Cuyckens (Eds.), *The Oxford handbook of cognitive linguistics* (pp. 117–138). Oxford: Oxford University Press.

Schneider, N. (2010). Computational cognitive morphosemantics: Modeling morphological compositionality in Hebrew verbs with embodied construction grammar. In N. Rolle, J. Steman, & J. Sylak-Glassman (Eds.), *Proceedings of the thirty-sixth annual meeting of the Berkeley Linguistics Society* (pp. 353–368). Berkeley: BLS.

Trosterud, T. (2006). Grammatically based language technology for minority languages. In A. Saxena & L. Borin (Eds.), *Lesser-known languages of South Asia: Status and policies, case studies and applications of information technology* (pp. 293–315). Berlin: Mouton de Gruyter.

Tyers, F., & Pirinen, T. A. (2016). Intermediate representation in rule-based machine translation for the Uralic languages. In *Proceedings of the second international workshop on computational linguistics for the Uralic languages* (pp. 92–104). Szeged: University of Szeged.

Wible, D., & Tsao, N. -L. (2010). StringNet as a computational resource for discovering and investigating linguistic constructions. In *Proceedings of the NAACL HLT workshop on extracting and using constructions in computational linguistics* (pp. 25–31). Los Angeles: ACL.

Wintner, S. (2009). What science underlies natural language engineering? *Computational Linguistics* 35(4), 641–644. doi:10.1162/coli.2009.35.4.35409

Wray, A. (2002). *Formulaic language and the lexicon*. Cambridge: Cambridge University Press. doi:10.1017/CBO9780511519772

Aligning constructicons across languages

A trilingual comparison between English, Swedish, and Brazilian Portuguese

Benjamin Lyngfelt, Tiago Timponi Torrent, Adrieli Laviola, Linnéa Bäckström, Anna Helga Hannesdóttir and Ely Edison da Silva Matos

This chapter addresses interlingual relations between constructions. The perspective is contrastive rather than typological, with an aim towards multilingual constructicon development. Building on previous work on the alignment of frame-based multilingual lexical databases, we explore possibilities and problems for multilingual constructicography. By comparing the dataset of Berkeley's English constructicon to Brazilian Portuguese and Swedish, we discuss the alignment of constructicons vis á vis the existence of more or less equivalent constructions or the deployment of different linguistic strategies in different languages.

1. Introduction

In this chapter we turn to the prospects for interlingual constructicography. Given the development of closely related constructicon resources for several languages, it is a natural next step to look into possibilities for connecting them. Such an endeavor requires contrastive construction analyses on the one hand, and the development of linking tools and a cross-linguistically useful representation format on the other.

To this end, we present an explorative trilingual comparison, where the full set of entries in the FrameNet English constructicon has been compared to Swedish and Brazilian Portuguese, respectively. This dataset is both intended as a startup material for a multilingual constructicon and as a testing ground for the practical and theoretical problems involved in such a development. The comparison

builds on previous bilingual studies of English-Swedish (Bäckström, Lyngfelt, & Sköldberg, 2014) and English-Brazilian Portuguese (Laviola, 2015), which approach the same set of English constructions with slightly different methodology. In the present paper we have harmonized and operationalized the approaches in a way intended to also be applicable to other languages (see Section 2). The results of the comparison are presented in Section 3.

We also tentatively address the question of how the interlingual relations between constructions can be accounted for in a multilingual online resource (Section 4). As constructicography may be perceived as a cross between construction grammar and lexicography (see Lyngfelt, this volume), we treat interlingual constructicography as a combination of contrastive construction grammar (Section 1.1) and interlingual lexicography (Sections 1.2–3). Regarding lexicography, Section 1.2 provides a general account of the bilingual lexicography tradition, in particular its notion(s) of equivalence and the role of directionality, as this perspective is central to the understanding of the present approach. In Section 1.3, we focus on frame-based computational lexicography, due to the similarities, both practical and theoretical, between contemporary framenet- and constructicon resources, and also the more or less close collaboration between ongoing framenet- and constructicon developments.

1.1   Contrastive construction grammar

Most – if not all – work in construction grammar (CxG) is based on the default assumption that constructions (cxns) are language-specific. This follows in part from the view of cxns as conventionalized form-function pairings (since conventions are at least to some extent dependent of historical circumstances), in part from the essentially data-driven approach. Thus, constructionists adopt the traditional structuralist practice, developed further in modern language typology, of addressing each language on its own terms, not presuming that distinctions and categories valid for one language are *a priori* applicable to others (F. Boas, 1911; cf. Haspelmath, 2007).

In CxG, this position is strongly advocated by Croft (2001), who states that "constructions as cross-linguistically valid configurations of morphosyntactic properties do not exist" (Croft, 2001, p. 283). A somewhat less categorical approach is not to rule out the possibility of at least some cross-linguistically valid cxns – as universals or as generalizations that hold for groups of languages – but hold off any strong claims in either direction until substantiated by detailed analysis of solid empirical evidence (Goldberg, 2013). Nonetheless, the working hypothesis remains that cxns are considered language-specific until convincingly demonstrated otherwise (e.g. Boas, 2010a).

Accordingly, the vast majority of work in construction grammar consists of case studies of particular cxns, or groups of cxns, in single languages. This does in no way, however, preclude cross-linguistic constructionist approaches. On the contrary, differences between languages are an excellent reason for comparative studies of constructions. While broad typological studies in construction grammar are rare (an exception being Croft, 2001, Chapters 8–9), contrastive studies are more common, typically comparing similar cxns in two or more closely related languages. Examples include Barðdal (2004) on impersonal cxns in German, Icelandic and Faroese, Hilpert (2008) on future cxns in Germanic languages, Kuzar (2012) on sentence patterns in English and Hebrew, the papers in Boas (2010b), and many more.

Contrastive CxG does not necessarily consist of interlingual comparisons between corresponding cxns *per se*. From a more onomasiological perspective, the starting point can be a functional space, or domain, which may be realized by different sets of cxns in different languages (e.g. Croft, 2001; Fried, 2006). The base of comparison could also be semantic frames (e.g. Hasegawa, Lee-Goldman, & Fillmore, 2014).[1] Furthermore, it is worth noting that even primarily monolingual studies are often contrastive to varying degrees, by contrasting properties of the cxn in question to those of similar cxns in other languages, chiefly English. A good example is Dooley's (2014) account of the Swedish comparative correlative `ju_desto` cxn (cf. `the_X-er_the_Y-er` in English).[2]

Another, somewhat less explored field for contrastive CxG concerns language contact in multilingual settings, including L2 varieties, translations, etc. (see, for example, the papers in Hilpert & Östman, 2014). In such contexts the question arises to what extent we are dealing with distinct varieties or more or less integrated linguistic systems. For example, Höder (2012; 2014) proposes that "multilingual speakers, psycholinguistically speaking, are not multiple monolinguals" (2014, p. 216) but rather possess multilingual grammars,[3] for which he introduces the model of Diasystematic CxG.

While constructional equivalence across languages is rarely assumed explicitly in contrastive studies, some notion of equivalence is often presupposed. By asking "What are the properties of construction C in languages X, Y and Z?" it is not presumed that C^X, C^Y and C^Z are identical, but a perception of C as a cross-linguistically applicable phenomenon is clearly adopted. Likewise, a statement such as "Function F is expressed by an R construction (e.g., reflexive) in languages X and Y", while not presupposing identical R cxns in X and Y, is based on the notion of R^X and R^Y

---

1.   The role of frames in cross-linguistic studies will be discussed in Section 1.3.

2.   Throughout, names of particular constructions are marked by a sans serif font (`Consolas`).

3.   For a different view, see e.g. Wassenscheidt (2014).

as corresponding cxns in some sense (cf. also the relation between F, F^X and F^Y). To what extent such presumptions imply constructional equivalence is usually not a major issue, however, since their purpose in such studies is to serve as a base for comparison, not to establish correspondence relations.

In constructicography, on the other hand, as in lexicography (see Section 1.2), establishing equivalence relations is a central goal of cross-linguistic comparisons. Consequently, the stance towards equivalence is a key feature of whatever approach adopted. How equivalence is operationalized – and differences accounted for – in the present study is detailed in Section 2.

As a final remark, cross-linguistic studies of cxns are typically detailed accounts of a small set of closely related cxns, and large-scale comparisons are rare. Two approaches that account for a larger number and more diverse types of cxns – although by less detailed and thorough analyses – are Bäckström, Lyngfelt & Sköldberg (2014) and Laviola (2015), who compare the full set of construction entries in the FrameNet English Constructicon (Fillmore, Lee-Goldman, & Rhomieux, 2012) to Swedish and Brazilian Portuguese, respectively. They also differ from most other work in contrastive CxG in being oriented towards constructicography and therefore focusing more on (approximate) equivalence relations. As mentioned above, the present work is a continuation of these two studies.[4]

1.2   Bilingual lexicography and equivalence

As the present approach is a combination of CxG and lexicography, in a predominantly CxG oriented setting, the perspective of lexicography warrants some introduction. In this section, we present bilingual lexicography in general, whereas Section 1.3 reviews the perhaps somewhat more familiar tradition of frame-based lexicography.

Bilingual lexicography has a long history, dating back to ancient Mesopotamia some 4.000 years ago. The oldest bilingual word list we know consists of Sumerian words provided with Akkadian equivalents or explanations (Snell-Hornby, 1986). This text witnesses the earliest attempts to lexicographic activities – presenting the meaning of a word in one language in the form of an equivalent in another language. Monolingual lexicography, on the other hand, is a more recent development that did not appear until millennia later. Thus, monolingual lexicography grew out of its bilingual counterpart and not, as one might sometimes get the impression, the other way around.

---

4.   Boas and Ziem (this volume) compare entries in the FrameNet English Constructicon with German. Unlike the present study, however, they present a more detailed comparison of a small number of cxns.

During the centuries, lexicographic trends have come and gone. Right up to the Middle Ages and the early modern time, most vocabularies were thematically organized in conventionalized topics. Due to the parallel thematic organization, each and every one of the present languages could provide access to the other languages. In course of time, this indeed interlingual lexicographic genre has become increasingly rare, replaced by the alphabetically organized semasiological bilingual dictionary.

In the alphabetically organized lexicographic description, bilingual as well as monolingual, the focus lies on the lexical item, the headword or lemma leading into the dictionary entry. Most lemmas are made up of a single word but also multiword expressions, recognized as "significant units of meaning" (Atkins & Rundell, 2008, p. 167), should be considered as lexical items on equal terms. The essence of the general dictionary entry is the description of the semantic properties of the lemma. In the monolingual dictionary these are presented in the form of a definition, near synonyms or a periphrastic explanation and in the bilingual dictionary as one or more equivalents or in the form of explanations, periphrastic or encyclopaedic. In addition to the semantic properties, also the pragmatic and to some degree discourse or functional properties are accounted for. However, the lexicographic perspective is as a rule strictly focused on the properties of the isolated lemma. Comments on the semantic, pragmatic or constructive properties of the lemma compared with these properties of its synonyms and near synonyms is not a recurring feature in general dictionaries.

The printed bilingual dictionary is necessarily monoscopal, viz. one of the languages is the source language (SL) and the other one the target language (TL). Further, the printed dictionary is monofunctional as it is conceived for a certain user group, either native speakers of the SL or native speakers of the TL. The description of each of the languages is adjusted to the intended users' skills and needs. Thus dictionaries for SL speakers aim at assisting the user in encoding text in the foreign TL while dictionaries conceived for TL speakers are intended to support the users comprehension when decoding text in the foreign source language. Since each activity demands quite different language skills – and thus essentially different lexicographic assistance – the intended function of the dictionary, active for encoding and passive for decoding respectively, directs the information provided in the entries.

1.2.1   *Lexicographic equivalence*

The objective of the bilingual dictionary has been summarized as to "help translating from one language into another, or in producing texts in language other [than] the user's native one, or both" (Zgusta, 1971, p. 213). Thus, the two disciplines, bilingual lexicography and translation science, share some of the central concepts – not only the different roles of the two languages involved, one being

the source language (SL) and the other one the target language (TL), but foremost the concept of equivalence: "Equivalence is the axis about which the activity of translation turns" (Kromann, Riiber, & Rosbach, 1991, p. 2717). These concepts are however far from identical in the two disciplines. In lexicography the SL unit is the isolated lexical item, completed with examples of usage and of the different senses that are identified, rendered in the TL in equivalent lexical items and perhaps translations of the examples of usage. In translation on the other hand it is the text that is the SL unit.

This is of course crucial also for the concept of equivalence in the two disciplines. In translation science different types of equivalence dimensions are distinguished, e.g. denotative, connotative or pragmatic equivalence, and the scope of the equivalence is the text (Koller, 2011, p. 219). The lexicographic equivalence on the other hand, albeit based on a meticulous contrastive analysis, merely focuses on one of these equivalence dimensions, viz. the relationship between the denotative meaning of the lemma and preferably one or more single word units in the target language. The denotative equivalence relation between the lemma and the equivalent can in turn be of various degrees. The equivalence can be total, i.e. there is a conventionalized TL item that matches the semantic properties of the SL expression, or it can be partial, i.e. the relationship between a lemma and an equivalent is such that one of the items is an interlingual hyperonym to the other one (Svensén, 2009, p. 255ff.).

Lately, however, the conventional concept of lexicographic equivalence has been questioned by metalexicographers and also by some lexicographers. The questions raised concerns e.g. whether equivalence is 'discovered', "(does it exist prior to being established by the lexicographer) or is it 'created' by the lexicographers act?" and also whether there is one kind of equivalence or many (Adamska-Sałaciak, 2010, p. 387 ff.). Another issue of present interest is the degree of conventionalization and lexicalization of concepts and expressions. A word can be accepted and frequently used even if the concept it denotes is vague and poorly conventionalized. It is then up to the lexicographer to establish some kind of equivalence between such units in the SL and units in the TL. In cases like that the types of adequate equivalence relations can be referential, denotative or nominal equivalence rather than the default types total or partial.

1.2.2   *Bilingual e-lexicography*

The description of bilingual lexicography above reflects the reality during the centuries, i.e. as long as the lexicographic efforts resulted in codices or, later on, in printed dictionaries. Now, in the age of the electronic dictionary with a variety of publication platforms, all the preconditions hitherto valid for the discipline have been thrown over.

Contrary to the printed dictionary the electronic dictionary is potentially biscopal as it – at least in theory – can offer equal access to both languages. Genuine electronic bilingual dictionaries are however still rare as most of the dictionaries published on electronic platforms are digitalized versions of already published, printed editions. The information provided is however still bound by the constraints of the printed form with one of the languages being the SL and the other one the TL. The possibilities and the challenges implied by e-lexicography have not yet been seriously adopted within the discipline but many of the central concepts must now be reconsidered and redefined. This goes for the lexicographic theory and methods alike (Hannesdóttir, 2015).

The core concept of bilingual lexicography, i.e. the distinction between source language and target language, is not consistent with e-dictionaries where the lexical units of both languages can be made equally accessible. Nor are the lexicographic functions in their absolute form, i.e. encoding v. decoding, valid in the e-dictionary. In contrast to the printed dictionary where the TL led a fairly anonymous existence, being accessible only through a specific SL unit, the lexicographic TL items can now be accessed just as easily as the ditto SL – and thus regarded as a SL by the user.

This means that the indirect lexicographic description of the TL as subordinate to the SL is no longer sufficient from the user's point of view. The arbitrary subset of the TL, motivated only by the equivalence relations to the carefully selected SL unit, is in the e-dictionary exposed as an equally representative subset of the actual language. Equivalent lacunae, unnoticeable in the printed dictionary, now appear as inauspicious lemma lacunae. Further, each of the two languages can be the L2 of the user. For bilingual dictionaries, conceived for electronic platforms, this entails a contrastive analysis and description of each of the languages as a foreign language, and for decoding and producing text alike. This in turn calls for adjustments of the information displayed. The bilingual e-dictionary is therefore not only a dictionary in the conventionalized meaning but rather an electronic resource in form of a parallel corpus completed with semantic, pragmatic, syntactic, phraseological etc. information based on contrastive analysis.

1.2.3   *Lexicography v. lexicology*

The monolingual lexicographic description is based on a lexical analysis. While lexicology involves studies of multiple linguistic aspects of the word, lexicography mainly focuses on the lexical semantics. The advances in that field, as e.g. within frame-based lexicography, has hitherto first and foremost gained monolingual lexicography. Contrastive lexicology concerning other features than semantics has not become a distinguished component in the bilingual lexicography.

The lexicographic description is a rather simplified presentation of lexical semantics. The division of word meaning into sub-senses is a way of structuring

dictionary entries rather than reflecting linguistic reality and "the lexicographic sense may bear, at best, a tenuous relationship to linguistic notions" (Lew, 2013, p. 285). Thus lexicographers and corpus linguists adopting a lexicological perspective on the lexicographic description now tend to speak of *meaning potential* rather than of *word senses* (Kilgariff, 1997; Hanks, 2000). The dichotomy of form and meaning is not irrefutable; based on massive corpus evidence it can be claimed that there is a strong co-occurrence of these two entities: "like meanings tend to be expressed through like structures" (Lew, 2013, p. 286). Further, due to the dichotomy between the lexical and the grammatical perspective in linguistic studies a wide range of phenomena tends to be neglected in both disciplines, phenomena such as discrepancies in selection and collocational restrictions of near synonyms as well as similarities and discrepancies in construction patterns in a language specific as well as a contrastive perspective (e.g. preposition deletion in English & Swedish constructions; Ralph, 1975; Boas, 2008). Construction grammar offers a way of bridging this gap, in terms of linguistic levels. There is also, however, a gap in perspective between (theoretical) grammar and (applied) lexicography. Accommodating the two traditions is a key feature of constructicon development.

1.3   Frame-based computational lexicography

One form of lexicography that seems particularly relevant to consider in this context is FrameNet, for two reasons: First, because of the close historical, practical, and to some extent theoretical connections between constructicons and framenets (see Lee-Goldman & Petruck, this volume; Lyngfelt, this volume; Lyngfelt, Bäckström et al., this volume; Ohara, this volume; Torrent et al., this volume). Second, due to the fact that FrameNet methodology has been extensively employed for bi- and multilingual lexicography – either using the English FrameNet infrastructure as a starting point to develop framenets for other languages (Boas, 2002, 2005; Subirats & Petruck, 2003) or creating frame-based multilingual resources (Sato, 2008; Schmidt, 2009; Boas & Dux, 2013; Torrent, Salomão et al., 2014; Peron-Corrêa et al., 2016).

In FrameNet, lexical meaning is characterized in relation to semantic frames, which are schematic scenarios including not only the words evoking the frame but also the participants involved in the situation, so-called Frame Elements (see e.g. Fillmore & Baker, 2010; Lee-Goldman & Petruck, this volume). Lexical units with the same background meaning, in the sense of evoking the same frame, may and do differ in other respects, regarding both semantics and morpho-syntax. Accordingly, multilingual framenet application is usually based on the assumption that at least some frames apply to different languages and the cross-language differences may be

attributed either to the lexical units instantiating the frames, or accounted for by ed-
iting parts of the frame structure. Thus, frames, their internal structure (the Frame
elements, FEs) and the relations among them are taken from Berkeley FrameNet
and applied – with the needed changes and adaptations – to the target language (a
so-called expand approach, Vossen, 1998, p. 83f.; cf. Section 4.1 below).[5]

This approach has been used in the development of interlingual lexicographic
analyses. As an example, while presenting the German FrameNet (GFN), Boas
(2002) advocates in favor of defining German lexical units (LUs) based on the set
of English LUs. With the aid of bilingual and monolingual dictionaries, and also
taking into consideration the valence descriptions provided by Berkeley FrameNet,
the GFN lexicographer would have the task of finding the best-fit equivalent to the
English LU. After, the lexicographer would survey the German LU in corpora and
check the analysis against language use evidence.

On a different, but related series of efforts, researchers in the field of multi-
lingual lexicography have used FrameNet as a foundation for the development of
Multilingual Lexical Databases (MLDs), some of which take advantage of already
expanded framenets.

Sato (2008), for example, developed multilingual features for the FrameSQL
tool. In such an application, databases from Berkeley FrameNet (Fillmore, Johnson,
& Petruck, 2003; Fillmore et al., 2003), Spanish FrameNet (Subirats & Petruck,
2003), Japanese FrameNet (Ohara et al., 2004) and the German SALSA project
(Burchardt et al., 2006) are aligned and fully searchable through multiple query
types. Users can list LUs evoking a given frame in all the languages covered by
the tool, as well as search for specific FEs and see how they are instantiated across
languages. For defining lexical equivalences, FrameSQL (1) searches an electronic
bilingual dictionary for words that are listed as equivalents to the source word, (2)
searches the Berkeley FrameNet database for the LUs evoking the frame evoked
by the target word, and (3) creates a set of the words that are listed as outputs of
both (1) and (2).

Schmidt (2009) developed the Kicktionary, a trilingual lexicon of the language
of football. In this resource, LUs are grouped in frames, which, in turn, are grouped
in scenes. According to Schmidt, this approach is useful for linking multilingual
lexica because scenes and frames, at least those modeling football, are language
independent. He points out, however, issues such as differences in lexicalization
patterns, problems with the creation of frames for entity nouns, lack of clear-cut
boundaries between scenes and frames, and difficulties in defining which frames

---

5.   Such adaptations include, for example, new sets of syntactic and morphological categories
for the analyses (see Torrent & Ellsworth, 2013, for a detailed explanation of such a process).

would be included in a scene. Moreover, he claims that Kicktionary's scenes-and-frames approach does not include typical lexical relations such as equivalence. To address this issue, WordNet synsets (Fellbaum, 1998) were used to model cross-linguistic lexical relations.

Working on the domains of football – also – and tourism, Torrent, Salomão et al. (2014) developed the FrameNet Brasil World Cup Dictionary, a trilingual – English, Spanish and Brazilian Portuguese – electronic dictionary focused on non-specialist users. Unlike the Kicktionary, the World Cup Dictionary relies on framenet structure to automatically suggest equivalences between LUs in the three languages. In this software, the valence patterns derived from the annotation of sentences containing verbs and eventive nouns are compared between languages as a means of providing best-fit translations for the LU being searched by the user. Peron-Corrêa et al. (2016) describe the computational process involved and discuss its limitations.

Adopting a perspective centered on language pedagogy, Boas & Dux (2013) developed G-FOL – the German Frame-Semantic Online Lexicon – a tool for helping foreign language learners in vocabulary acquisition. In a G-FOL pilot study, the `Personal_relationship` frame was used to evaluate if students exposed to an adapted version of FrameNet would perform better in vocabulary acquisition. The authors show that the group who used G-FOL performed better than the control group in all vocabulary-related tests conducted, which focused on the semantic description of the vocabulary item, its situational uses, and syntactic properties.

The works in frame-based interlingual lexicography surveyed here, despite their very different goals, all approach some key issues in the field. First, they all claim that frames, the background cognitive systems relative to which lexical meaning is built, are a useful tool for comparing lexica across languages. Second, they all recognize that, although some frames may be very similar across languages, that doesn't hold for every frame in every language. Third, they all, at some point, recognize that even crosslingual frames may show differences in their LUs in regard to morphosyntactic properties, situational implications, distribution and frequency.

Hence, research on frame-based interlingual lexicography is usually carried out amidst the tension between the recognition of frames as crosslinguistically valid analytical tools and close consideration of the differences in lexicalization patterns. As noted by Boas (2005, p. 464), "although bilingual lexicon fragments might match in terms of their semantic and syntactic valences, they might differ in terms of domain, frequency, connotation, and collocation in the two languages."

Frame-based interlingual lexicography relates to interlingual constructicography in at least two ways: First, grammatical constructions may evoke frames just like words (lexical cxns) presumably do. Therefore, it is possible that cross-linguistic relations between cxns can be established via frames in the same basic manner as

in frame-based lexicography. Such an approach is discussed by Bäckström, Lyngfelt & Sköldberg (2014) and Laviola (2015). However, it would hardly be equally applicable to all cxns; it seems that some cxns evoke frames and some do not (cf. Ohara, this volume; Lyngfelt, Bäckström et al., this volume).

Second, framenets have adopted a constructionally inspired annotation process (Torrent, Lage et al., 2014), in which (1) multilayer analyses map semantic information (the Frame Elements – FEs) to the linguistic realization of the Lexical Unit being analyzed (which may include Grammatical Functions and Phrase Types associated to the FEs), and (2) the meaning-form correspondence patterns derived from the analyses represent the minimal valence of the lexical construction being analyzed (see Fillmore, 2013, for a discussion of minimal valences as properties of lexical constructions).

Thus, framenet analyses include a lot of constructional information that may be useful for constructicon development in several ways, within and possibly across languages. How well a frame-based approach to interlingual constructicography would actually work remains to be tested. Its main usefulness should concern the semantics of the cxns (frames being foremost semantic units), whereas their morpho-syntactic structure involves features beyond what lexicography – frame-based or not – is usually concerned with.[6]

2. Comparing constructions across languages

As a step towards connecting constructicons for different languages, we have conducted a three-part comparison between English, Swedish, and Brazilian Portuguese. The study is based on previous bilingual comparisons between English and Swedish (Bäckström, Lyngfelt, & Sköldberg, 2014), and between English and Brazilian Portuguese (Laviola, 2015). As in both these studies, our point of departure is the English FrameNet Constructicon (cf. Fillmore, Lee-Goldman, & Rhomieux, 2012), exploring to what extent there are Swedish and Brazilian Portuguese counterparts to the English construction entries. The comparison is thus unidirectional in the sense that English is always the source language; consequently, Swedish and Portuguese are only compared indirectly.

---

6. While FrameNet analyses yield valence patterns that may be regarded as lexical valence constructions, their formal realization would vary considerably across languages, making interlingual frame-based constructicography considerably more complex than ditto lexicography, even in the case of such lexical constructions. One must also take into consideration the tendency of framenets – and of lexicography in general – to adopt shortcuts so as to fit constructional phenomena into lexicographic analyses (cf. Ruppenhofer et al., 2016, p. 27).

The English FrameNet Constructicon (henceforth EngCcn) consists of 73 construction entries. Of these, 50 entries are fully developed and 23 are in a more preliminary stage, but all but seven were found explicit enough for the purposes of this study. The excluded entries were either too incomplete or too abstract to serve as a base for interlingual comparison. By "too abstract" we mean constructions that are defined irrespective of language-particular properties, for example `head-complements`, with the definition "A head selects for a set of complements". Such an entry concerns the model of grammar assumed rather than properties of the language described, and it would be rather pointless to ask what the counterparts of this "English" source cxn would be in the target languages. The other cxn entries excluded from the comparison are `bare_arg_ellipsis`, `bare_noun_phrase.role`, `modifier-head`, `subject-predicate`, `the_ubiquitous_noun`, and `valence_sharing.raising`. Consequently, the comparison concerns 66 English construction entries. In the following these will be referred to as the *source constructions* (or *source cxns* for short).

2.1 A four step comparison

Determining equivalence between words, let alone constructions, is no trivial task. From the very basic assumption that equivalence is considered a relation, with a certain value, between (at least) two entities, Adamska-Sałaciak (2010, p. 387) derives the following seven questions for a bilingual lexicographer to consider:

1. where (at what level of organisation) should we look for the entities between which the relationship obtains?
2. what exactly are those entities?
3. what is the nature of the relationship between them (e.g. identity, interchangeability, similarity, correspondence)?
4. what is the feature according to which the relationship is established or measured (e.g. meaning, reference, message, effect)?
5. is equivalence a unitary concept or should different types thereof be recognised?
6. is equivalence 'discovered' (does it exist prior to being established by the lexicographer) or is it 'created' by the lexicographer's act?
7. are our answers to 1–6 in agreement with the findings of linguists and translation theorists?

Given that the entities on the source side of the relationship were established beforehand, in our case the construction entries in EngCcn, our objective was to first look for more or less equivalent entities in the target languages and then, by comparative analysis, discern the relationships between the source and target entities in terms of meaning/function, form, and organization. As an operationalization of

this task, each English source construction was compared to Swedish and Brazilian Portuguese with respect to four questions:

1. Is there a corresponding construction, or set of constructions, in the target language?
2. Is there one construction in the target language, that covers the full functional range of the source construction and is not a superordinate construction?
3. Are the source construction and the closest target construction formally similar, except for lexical differences?
4. Do all formal differences follow from other constructions not of the same type as the source-target constructional pair?

These questions follow an ordered sequence where the first is a blocking question – i.e. in the case of a negative answer, no further questions were asked – and the fourth question was only asked in case of a negative answer to the third (see Figure 1). All four of them are polarity questions, to enable a numerical score for each pairing. Positive answers give a score of 1 and negative answers a score of 0, except for the fourth question where a positive answer renders 0.5 (see below). Thus, the maximum score for each pairing, in the case of both formal and functional equivalence, is 3.

1. Correspondence	2. Functional equivalence	3. Formal similarity	4. All differences external
No: 0	–	–	–
Yes: 1	Yes: 1 / No: 0	Yes: 1	–
		No: 0	Yes: 0.5 / No: 0

Figure 1. Comparison flowchart

Question 1, whether there is a corresponding cxn (or set of cxns), concerns *constructional equivalence* in the same sense as lexical equivalence in a bidirectional dictionary. In a very practical sense, it means whether there are any constructions to present as target equivalents in an interlingual constructicon. This does not presuppose (full) formal or functional equivalence, but requires cxns similar enough to be considered linkable. A negative answer to this question renders the following questions irrelevant and results in a total score of 0. Examples of such non-pairings are `subject_auxiliary_inversion`, which is not naturally employed in Portuguese, and `be_present-participle`, which is lacking in Swedish (cf. Section 3 below).

    Question 2 regards *functional equivalence*, but only in a unidirectional sense. It addresses whether there is a target cxn covering the full functional range of the source cxn, but not whether the same holds in the other direction. The latter would

require investigations beyond the scope of this study (see Section 2.2). At the same time, superordinate cxns are ruled out, in order not to collapse the distinction between functional equivalence and inclusion (which would follow from an unrestricted application of a unidirectional approach).[7]

Furthermore, we are well aware that complete functional equivalence is a rare thing, if it exists at all, especially if distribution is taken into account. This is not, however, what question 2 is meant to capture, even had we had the time to conduct the distributional investigations this would require. Rather, the purpose is to distinguish cases where a target cxn only covers part of the function of the source cxn or where the coverage is split between several target cxns. Hence, functional equivalence in this context is to be understood as absence of clear functional differences, not as full identity.

Question 3 concerns *formal similarity*, by which we mean similarity in morphosyntactic structure. Purely lexical differences between corresponding construction elements of the same part-of-speech are disregarded. For example, the Swedish counterpart to `comparison_inequality` (as in *harder than Kryptonite*) is considered formally similar, since the only difference involved is that between *than* (En.) and *än* (Sw.). The Brazilian Portuguese counterpart, on the other hand, is considered formally different, since it differs not only by its use of *que* but also with regard to the comparative adjective phrase. Whereas English and Swedish employs both morphological and periphrastic comparative (e.g. *sturdier* vs. *more sturdy*) productively, the morphological pattern is used with only four adjectives in Brazilian Portuguese, and even those adjectives also occur in the periphrastic variant.[8]

Question 4 concerns whether the formal differences are construction-specific or *follow from other constructions*. For example, there are general word order differences between especially Portuguese and English/Swedish; both Portuguese and Swedish generally display gender agreement, whereas English basically lacks grammatical gender, etc. Such wide-ranging differences affect a large number of

---

7. The exclusion of superordinate constructions is operationalized in the following way: if a target cxn corresponds to a cxn in the source language that is superordinate to the source cxn, the target cxn is considered superordinate. Note that a superordinate target cxn may still qualify as a constructional equivalent (according to question 1), but it would not be considered functionally equivalent. Such a case is the `what-with_absolute` cxn, which is a special variant of `with_absolute`, where the same target cxn is employed for both cxns in both Swedish and Brazilian Portuguese (see Section 3). The alternative to this approach would be to conclude that `what-with_absolute` is lacking in these languages and consequently offer no constructional equivalent in an interlingual constructicon.

8. While there are clearly distributional differences between English and Swedish in the use of morphological vs. periphrastic comparative, that level of detail is not taken into account here. For present purposes, the relevant fact is that both variants are commonly used in both languages.

cxns without necessarily being distinguishing properties of each of them. Arguably, such features are associated with very general cxns and then hold for more specific ones by inheritance. Such differences are still noted with respect to question 3, but, in cases where all formal differences are external in this sense, the comparison is adjusted for by a positive 0.5 score for question 4.

For a formal difference to count as external, it has to be inherited by a construction not too closely related to the source and target cxns. Consider for instance the family of `adjective_as nominal` cxns, which all share the property of lacking a nominal head. While the specific variants `Adjective_as nominal.abstract` (*the inevitable*), `Adjective_as nominal.anaphoric` (context-dependent), and `adjective_as nominal.people` (*the poor*) may be said to inherit this property from their mother cxn, it is still a salient feature of all of them. To the extent they display number-, gender- and definiteness marking, however, such agreement features are rather associated with general noun phrase cxns. Hence, external differences are defined as those that are not salient characteristics of the particular cxns at hand.

2.2   Methodological remarks

Before turning to the results of the trilingual comparison, there are a few methodological issues to consider. In particular, we will address (a) the numerical notation, (b) the deep-seated bias ensuing from our unidirectional approach, (c) why functional and formal differences are recorded, and (d) the role of theoretical and editorial considerations.

To begin with (a), the numerical scores are not intended as a measure of relative similarity. The individual features are essentially binary, with no attempt at grading the differences recognized but giving equal weight to minor and major ones. Instead, the scores serve as a tool for sorting different kinds of correspondences and indicating what kinds of differences would have to be taken into account for multilingual constructicon applications. In addition, despite several rounds of fine-tuning the criteria and harmonizing the approaches of different analysts, some subjective component to the scoring clearly remains. Consequently, any attempt to interpret the deceptively distinct numbers as measures of similarity between cxns must be undertaken with caution. Interpreting them as a measure of relative similarity between the languages involved should not be done at all.

Secondly (b), the comparison is unidirectional, which means that we have only studied to what extent entries in the English constructicon (EngCcn) are matched by Swedish and Brazilian Portuguese cxns, not the degree of equivalence in the opposite direction. Thus, the comparison is based on constructional distinctions in EngCcn and biased accordingly. For example, the closest Swedish equivalent to the

English `let_alone` cxn is `för_att_inte_tala_om` (lit. 'for to not speak about'). This does not, however, imply that the converse relation holds; an alternative close at hand would be the cognate cxn `not_to_mention`. How well these two (and potentially other) English expressions correspond to Swedish `för_att_inte_tala_om` will not be addressed here.

Furthermore, the comparison is not only influenced by English conditions but also specifically by EngCcn, i.e. based on English constructions *as they are presented in EngCcn*. This means that editorial decisions in EngCcn are in principle treated as general facts about English, on the one hand, and that nuances and variants that for some reason are not presented in EngCcn are not taken in consideration, on the other. We will return to this issue under (d) below.

Adopting a unidirectional approach follows standard practice in interlingual lexicography and was the only feasible way to perform a comparison over the whole set of cxn entries in EngCcn. Still, any methodological choice has consequences, and the bias in this case seems to be towards similarity. While the similarities between the source and the target cxn are often straight-forwardly noticeable, identifying the differences require further analysis. The closer the analysis, the more differences were found.[9]

Thirdly (c), it may not be obvious why formal and functional differences are recorded. Since the comparison is to pave the way for eventually connecting the constructicons, why not merely determine which cxns to link to? However, this is an investigation, not just a matching procedure, and correspondence is not a binary property. Even from a linking perspective, it is highly relevant to establish not only which cxns match but also how well they match and in what respects they differ. This is all the more the case since the links must connect cxn descriptions, not just their names (which are usually not fully transparent). Therefore information about salient differences matter for how the cxns are to be represented in a cross-linguistically adequate format.

Furthermore, both formal and functional differences concern the network relations within the resources. On the one hand, not all matches are one-to-one relations. On the other hand, some similarities and differences follow from inheritance, and it is not *a priori* obvious if, where and how such information should be presented. In addition, the investigation is not only an internal affair with a narrow focus on application development. The contrastive results should also be of interest to the linguistic community.

---

9.   After comparing EngCcn to Swedish (Bäckström, Lyngfelt, & Sköldberg, 2014), several Swedish construction entries were developed from the comparison to their English counterparts. The additional investigation involved in that process usually revealed more differences than the initial comparison. The same holds for Brazilian Portuguese (Laviola 2015).

Finally (d), it should be stressed that the comparison depends not only on cross-linguistic contrasts but also on a number of theoretical and editorial decisions – and not only those concerning the source resource EngCcn. An illustrative example is the `be_recip` cxn, which comes in two variants, one symmetrical (*Watson and Sherlock are close friends*) and one asymmetrical (*Watson is close friends with Sherlock*), both with the relational noun in the plural. In both Brazilian Portuguese and Swedish, the symmetrical case would be expressed similarly, whereas a singular noun would be used in the asymmetrical structure. One way to picture the contrast is that all three languages have a plural reciprocal cxn, but the ones in the target languages are more restricted (a functional difference). An alternative account is that they all have a general reciprocal cxn but differ in how it is expressed (a formal difference). A third possibility would have been to handle the symmetrical and asymmetrical patterns separately, but that option is precluded by the existing, unified treatment of `be_recip` in EngCcn. Thus, the choice is partly linguistically grounded – a unified account makes more sense for English than for the other languages – partly *ad hoc*.

Thus, even this minor a difference is enough to raise questions about how to delimit the cxns. In this case, the practical effects hardly matter. The fact that symmetrical and asymmetrical reciprocals correspond differently will have to be specified in an interlingual constructicon, however the cxn entries are delimited.[10] Nonetheless, the dependence on partially *ad hoc* choices remains a factor to be wary of and to take into account.

A particular issue with potentially far-reaching consequences is the relation to FrameNet. EngCcn is aligned with the English FrameNet and, where applicable, constructional distinctions are therefore aligned with (English) lexical frames. This means that some decisions in EngCcn may ultimately be grounded in lexical rather than constructional properties;[11] the FrameNet analysis of the English lexicon has thus influenced the analysis of English (grammatical) cxns, which in turn restricts the interlingual cxn comparison. This favors compatibility between framenets and constructicons but is also a potential source for mismatches. One case at hand concerns `rate` cxns in English and Swedish; see Lyngfelt, Bäckström et al. (this volume; cf. also Bäckström, Lyngfelt, & Sköldberg, 2014).

To conclude, the unidirectional approach makes the comparison somewhat biased, the more so given the dependence on not entirely objective analytical

---

10. The solution chosen in this case was to assume a functional difference, partly because the asymmetric variant in Portuguese and Swedish is not strictly reciprocal but tending towards the English pattern *a friend **of** Sherlock*.

11. Even presuming that lexical items are (lexical) constructions, they do not equal the source constructions but are one step further removed from the target cxns.

decisions. This source of error should be borne in mind, both to minimize the bias and to avoid drawing too strong conclusions from the results. At the same time, however, the effects of the bias should not be exaggerated. On the whole, the comparison presented in the following section should give an accurate account of the state of affairs.

3.   Comparison of English, Swedish and Brazilian Portuguese constructions

In this section, we present the results of the contrastive analyses. After an account of the overall results, we will in turn discuss relations of high (Section 3.1), low (Section 3.2) and medium (Section 3.3) equivalence. A numeric summary of all the analyses is presented in the appendix.

Out of 66 English construction entries, linkable construction equivalents were found for all but five cxns in Brazilian Portuguese and four in Swedish, as shown in Table 1. Furthermore, the vast majority of the constructional pairings (56 for Portuguese and 54 for Swedish) qualify as functionally equivalent, which means the relation can be handled as a one-to-one correspondence.

Table 1. Correspondence and functional equivalence

	Brazilian Portuguese	Swedish
1. Correspondence	61	62
2. Functional equivalence	56	54

The data also allow for an indirect comparison between Brazilian Portuguese and Swedish: Corresponding cxns were found in both languages in 58 cases.[12] Out of those, 52 matches can be indirectly deduced as potentially functionally equivalent.[13] Note, however, that these numbers only indicate potential matches, since they are derived from an indirect comparison via English. The indirect score for

---

12. One cxn (`subject-auxiliary-inversion.exclamation`) is missing in both target languages, four additional cxns are lacking in Brazilian Portuguese and three in Swedish, for a total of eight.

13. In 50 cases, both languages scored 1 for functional equivalence to the English source cxn. There are also two cases where both languages scored 0 for the same reason: neither language has a specific counterpart of the particular what-with_absolute cxn variant (see footnote 6 in relation to the functional equivalence discussion in Section 2.1), and both languages have a more restricted counterpart of the be_recip cxn (see Section 2.2). In one case, inversion_with_preposed_element, both languages scored 0 but for different reasons. The particular scorings are shown in the appendix.

correspondence is probably still fairly accurate, but the indications of functional equivalence can be no more than a promising starting point for future investigation (all the more so since the initial comparisons with English are unidirectional; see Section 2.2). Nevertheless, the relatively high number of matches is promising for future work towards linking the resources.

Turning to the formal side of the cxns, we find more differences, as shown in Table 2. In both languages, more cxns display formal differences than not. It is striking – and somewhat surprising – that Brazilian Portuguese displays far fewer cases of construction-specific formal differences vs. English. While we could expect more external differences in Portuguese, these do not *per se* preclude there also being cxn-specific differences. In general, slightly more differences have been noted for Swedish than for Brazilian Portuguese, which should of course not be taken to indicate that Swedish is less similar to English than Portuguese is. Rather, it seems to be due to a combination of random effects of the sample, the human factor involved in the analyses, and perhaps different editorial policies of the Swedish and Brazilian constructicon projects.

Table 2. Formal similarity

	Brazilian Portuguese	Swedish
Formally similar (1)	18	18
All differences external (0,5)	20	13
Formally different (0)	23	31
Total	61	62

The significance of these formal differences depends on what applications are considered. For identification tasks, formal differences may or may not be important. From a production perspective, they certainly are. Whether only construction specific differences matter, or also those inherited from more general cxns, depends not only on the intended application but also on the network structure of the resource. We will return to such issues in Section 4, after addressing the results in terms of degree of equivalence.

In the following sections, results are presented and discussed according to the level of equivalence found between constructions, even though we recognize that equivalence is a tricky concept for lexical let alone grammatical constructions as repeatedly pointed out throughout this chapter.

As is the case for lexical items, full equivalence would only hold if the items being compared had not only similar form and meaning, but also allowed for the same pragmatic inferences and presented similar distribution and contextual restrictions (Boas, 2005; Farø, 2004; cf. Atkins & Rundell, 2008; Svensén, 2009). As

already pointed out in Section 1.1, equivalence is directly grounded in the four criteria adopted in the comparison, and since none of them address issues such as distribution and contextual restrictions, we are actually dealing with some kind of idealized, or potential, equivalence between constructions.

Three levels of equivalence are proposed: high, medium and low. Each level translates into a score range, respectively, 3–2.5, 2–1.5, and 1–0. As mentioned above (Section 2.2), it should be borne in mind that the differences behind the numbers are not graded. For instance, the Brazilian Portuguese counterpart of the `degree_so` cxn, differs from the English source cxn in several respects, but since the differences are all inherited from more general cxns the pairing still qualifies for a high level of equivalence with a score of 2.5 (see Section 3.1) – whereas the Swedish counterpart of `measurement_plus_prepositional_phrase`, although similar to English in most respects, displays one small functional difference and one small formal difference, hence earning a score of 1 and a status of low equivalence (see Section 3.2). In other words, the numbers only indicate *whether* there are functional or formal differences, not how many or how big they are.

We start by presenting and discussing in Sections 3.1 and 3.2 the high and low points in the continuum, and then move, in Section 3.3, to constructions with medium equivalence.

3.1   Constructions with high equivalence

Constructions presenting the final score of 3 and 2.5 are taken as having high equivalence to the English source constructions. These are pairings judged to be both functionally equivalent and formally similar or where the only formal differences are those that follow from more general cxns.[14] Table 3 shows the number of constructions presenting final scores of 3 and 2.5 for each language.

Table 3. Final scores per language – scores 3 and 2.5

Score	Language	
	Brazilian Portuguese	**Swedish**
3	17	17
2.5	18	13

---

14. As noted in Section 2.1, purely lexical differences are disregarded, such as choice of preposition or the obvious but in this context trivial difference between the conjunctions *and* (English), *e* (Brazilian Portuguese), and *och* (Swedish).

Pairings converging on both form and function (score 3) typically concern fairly basic functions that are not dependent on particular morphosyntactic marking. A typical example is the `coordination` cxn, which is defined in EngCcn as follows: "Coordinates – units of identical or like types – are coordinated by a Conjunction. If more than two Coordinates are present, then all but the last must be followed by a Conjunction, or alternatively, only the penultimate Coordinate must be followed by a Conjunction." Such a definition holds for both Swedish and Brazilian Portuguese, and, therefore, the `Coordination` construction in English is taken as highly equivalent to those in Swedish and Brazilian Portuguese, both in terms of function and form. Other typical examples of 3 score pairings in both Brazilian Portuguese and Swedish are `gapping`, `integrated_appositive` and `tautology`.[15]

Matches that receive a score of 3 in only one of the languages typically display some formal difference in the deviating language. For instance, `noun-noun_compound` in Swedish stands out by the use of a linking morpheme, and `rather-than_coordination` in Portuguese differs by the use of a different grammatical structure linking the conjuncts being coordinated.

A score of 2.5 means that some formal difference can be found between the source and the target constructions, but only differences following from more general constructions. Such differences are typically basic morphosyntactic properties such as general agreement patterns. The `degree_so` construction (*so long/terrible* (*that*) S) exemplifies this situation for both Swedish and Brazilian Portuguese. Examples (1)–(2) present constructs licensed by `degree_so` and its closest counterparts in the target languages, (1) comparing English and Swedish and (2) comparing English and Portuguese.

(1) a. These horns can be so long and incurved that there is a danger of damage to the animal's cheeks (if they are not carefully trimmed at the tips)

   b. Hornen   kan   bli   så långa   och inåtsvängda att
       horn.PL.DEF can.PRS become.INF so long.PL and incurved.PL that
       det   är   risk för skada   på djurens   kinder
       there be.PRS risk for damage on animal.PL.DEF.GEN cheek.PL

The relevant difference in Swedish concerns the adjectival head, which agrees in number and gender with its antecedent noun. A case of number agreement is illustrated in (1b), where the *-a* suffix marks the adjectives for plural. This property is not particular to `degree_so`, but is a general feature of adjectival cxns in Swedish (in adnominal position, adjectives also agree with respect to definiteness).

---

15. For those English construction entries that are only referred to here, full accounts are available online at <http://www1.icsi.berkeley.edu/~hsato/cxn00/21colorTag/index.html>. Several of them are also described in Fillmore, Lee-Goldman, & Rhomieux (2012).

Agreement on the adjective also applies to Portuguese. In addition, the complementizer is mandatory, not optional as in English (and Swedish). This is illustrated in (2):

(2)   a.   The smell is so terrible you want to throw up
b.   O   cheiro   é   tão ruim   que   você vai   querer
the smell.SG be.PRS so bad.SG that you go.PRS want.INF
vomitar.
throw_up.INF

Neither of these differences are related to this construction exclusively, but to adjectives and complementizers in general. Also note that the existence of more differences in Portuguese than in Swedish does not affect the numerical score, since they are all external.

3.2   Constructions with low equivalence

Pairings of low equivalence are on the one hand cases where a corresponding target cxn is missing (score 0), and on the other hand cxns that differ in both form and function (score 1). As mentioned above, there are five non-pairings for Brazilian Portuguese and four for Swedish (where one English source cxn is missing in both languages). Pairings that differ in both form and function are six in Swedish and only two in Brazilian Portuguese.

Table 4. Final scores per language – scores 0 and 1

Score	Language	
	Brazilian Portuguese	**Swedish**
0	5	4
1	2	6

The cxns missing in Brazilian Portuguese are two `subject_auxiliary_inversion` cxns (out of seven) and the three *way* cxns. `Inversion` is actually lacking in general, but for the other five inversion cxns in EngCcn, Portuguese performs the same functions with different means. As for the *way* cxns, they all employ the structure [NP V one's way PP] to express self motion,[16] where *one's* is a possessive pronoun coindexed with the NP agent and the PP is the goal or direction, as illustrated in (3). There are three subtypes of this cxn, based on the meaning of the verb: means (a), manner (b) or neutral (c).

---

16. In FrameNet terms, they evoke the `Motion` frame.

(3) a. He squeezed his way down the stairs.
   b. She danced her way into Europe.
   c. My problem was to make my way from the entrance to a vacant table.

In Portuguese there is no construction with the same functional characteristics. To translate the sentences in (3), one would have to use two or more argument structure constructions. In Swedish, on the other hand, there are reflexive cxns filling roughly the same functions as the English *way* cxns.[17]

The four cxns missing in Swedish are `be_present-participle`, `tagged_sentence_subjectless,` `rate.milage` and `subject_auxiliary_inversion.exclamation` (ex. *Don't I know it!*). The participial cxn (*be* V-*ing*) is missing because Swedish does not mark aspect systematically, at least not in a similar way, whereas `tagged_sentence_subjectless` (ex. *Giving you trouble, was he?*) is a kind of particular cxn simply not expected to have counterparts in every language. `rate.milage` is one of four rate cxns in EngCcn. Swedish has counterparts of the other three, but this particular variant is missing.[18]

The `inversion` cxn is a somewhat more complex case. Since Swedish is a V2 language (verb second), VS word order, or "inversion", is the standard pattern whenever the subject is not clause initial. In English, on the other hand, there are a number of specific inversion patterns (seven inversion cxns in EngCcn), which are scattered remnants from an older V2 system (cf., e.g., Fischer et al. 2000, Chapter 4). Most of them have Swedish counterparts, but this is a cxn where the striking non-SVO order has come to be associated with certain pragmatic effects that seem to depend on that word order being non-ordinary; it is also particularly based on *do*-support, which is not employed in Swedish.[19]

Turning to cxns that do have correspondences in the target languages but the target cxns differ in both form and function, there is one English cxn where this is the case for both Brazilian Portuguese and Swedish, namely `what-with_absolute` (ex. *What with health budgets being pruned and cut back I'm asking the health board if staff shortages perhaps were a contributory factor here*). This is an informal variant of the `with_absolute` cxn, where both target languages have a counterpart of the standard cxn but lack the informal variant. Hence, the corresponding general cxns are the closest counterparts of the special case as well.

---

17. The Swedish reflexive counterpart of `way_manner` is more restricted, but there is a particle cxn covering the remaining cases.

18. Note that the Swedish counterparts to the other rate cxns do not cover `rate.milage`; hence the 0 score (unlike the `what-with_absolute` cxn; see below).

19. Other English cxns with *do*-support do (sic!) have Swedish counterparts, although without *do*.

The data contains one more case where a Brazilian Portuguese target cxn differs in both form and function from its English counterpart (`inversion_with_preposed_element`), whereas another five were found in Swedish. There is nothing special about these cxns – they are simply a diverse set of cases where different formal and functional restrictions have been conventionalized in the respective languages. An illustrative Swedish example is the `measurement_plus_prepositional_phrase` cxn, as *in seven feet in width* and *twelve years of age*. The corresponding Swedish cxn is similar in most respects, but differs in form by the complement of the preposition usually being definite (lit. 'on the width') and is functionally restricted in not occurring with age expressions. These are instead expressed with an adjective, which is an option in English as well (cf. the `measurement_plus_adjective` cxn; ex. *twelve years old*). Thus, even minor differences may result in a low score.

What is note-worthy is not the existence of such cases but their relative rarity in the material, especially in Brazilian Portuguese. The main reason is probably the way that the comparison was carried out; had the investigation of functional equivalence involved more detailed distributional analyses, more functional differences would clearly have been discerned (see Section 2.1).

3.3   Constructions with medium equivalence

The middle group mainly consists of cxns that differ in either function or form. More specifically, it includes cases that are formally similar but functionally different (score 2), cases that are functionally equivalent but have cxn-particular formal differences (score 2), and cases with functional differences and no cxn-specific formal differences but displaying formal differences that follow from more general cxns (score 1.5).

As shown in Table 5, this middle group contains strikingly few cases of functional difference, only two for each language. These include the counterparts of the aforementioned `be_recip` cxn in both languages (see Section 2.2 above), with a score of 2 for Swedish and 1.5 for Portuguese. The other Swedish cxn in this group is `inversion_with_preposed_element` (score 2), and the other Portuguese case is `postpositive_adjective` (score 2). The latter is defined as follows in EngCcn:[20]

> A Noun is modified by an Adjective_phrase that appears entirely following the Noun. This construction is required for some adjectives (e.g., *galore*), and for most adjectives with complements (*people late to the party*). Adjectives with obligatory complements, such as *bent* (on) also must combine with nouns via this construction.

Table 5. Final scores per language – scores 2 and 1.5

Score	Languages	
	Brazilian Portuguese	**Swedish**
2 (functional difference)	1	2
2 (formal difference)	22	24
1.5	1	0

As can be seen from the definition, this construction is required by a specific class of adjectives in English, which present a syntactic behavior that deviates from that of the majority of English adjectives. In Brazilian Portuguese, however, the norm is precisely to put adjectives after the nouns they modify. Hence, although there is a construction for postpositive adjectives in Brazilian Portuguese, it does not have the same function as the one in English.[21]

While the low numbers for functional differences are striking, both in this section and the previous one, it is perhaps not so surprising given the present method of comparison. A more detailed investigation of distributional properties would clearly have revealed more functional differences. Such an investigation, however, would go beyond the present purpose of exploring the basic preconditions for aligning constructicons.

Turning to formal differences, these are saliently associated with the cxns in question, as opposed to the inherited differences mainly treated in Section 3.1 above). In some cases this is because these are the cxns from which the differences are inherited, such as `determined_noun_phrase`, where agreement patterns are specified that affect most other cxns involving noun phrases.

The majority, however, concern more particular properties. For example, the Swedish counterparts of the two `location_in_calendar` cxns in EngCcn behave differently with regard to definiteness. Past time locations are typically definite

---

21. One might also view this pairing as a case where the target cxn covers the full range of the source cxn, which would then count as functional equivalence according to the unidirectional approach employed in this study. It is also somewhat hazardous to assume a functional difference when the function of the source cxn is not clearly defined. Still, there is arguably a functional difference following from the fact that this word order pattern is marked in English but unmarked in Portuguese, lending it more of a rhetoric potential in the former language, hence the score of 2. Also, it correlates directly with the valence of the adjectival head, which is not a factor influencing the adjective phrase construction in Portuguese.

A similar reasoning applies to subject-auxiliary_inversion in Swedish: on the one hand, due to being a more general pattern, it has less of a rhetorical effect than in English; on the other hand, there are also more clearly defined differences in that certain rhetoric patterns have been conventionalized (see Section 3.4).

(*förra veckan* 'last week.DEF') whereas future ones are usually indefinite (*nästa vecka* 'next week.INDEF'). Although there are general differences between English and Swedish regarding definiteness marking, this particular distinction is specific to time expressions. Another Swedish example is tag questions, which are not of the form [be + Negation + Pronoun] as in English (*isn't it*, *aren't you*, etc.) but instead have the fixed form *eller hur* (lit. 'or how').

A Brazilian Portuguese example is the there.presentational cxn. While, in English, this construction features the word *there* as the subject of the verb *to be* after which an entity is presented, in Portuguese, the verbs *haver* 'exist' or *ter* 'have' are used with a null subject. Having null subjects is not a general property of these verbs – especially for *ter* – but a specific configuration they take in this construction.

While these formal differences are often as particular as the cxns they apply to, there are also cases that derive from more general differences but are still saliently associated with the cxns in question. Such a case is the Brazilian counterpart of subject_auxiliary_inversion.conditional, which differs from the English source cxn in not involving inversion. This is clearly a general property of Brazilian Portuguese, but it is also a salient difference regarding this particular pairing. Therefore it counts as a construction-specific difference.

4.   Prospects for multilingual constructicography

We will now turn to the prospects for future alignment of the three constructicons – and for eventually involving corresponding resources for other languages as well. In general, the issue can be broken down into a matter of linking, on the one hand, and questions of representation, on the other. Both aspects, especially representation, play out somewhat differently depending on the intended uses and users. The major distinction in this regard concerns the difference between language technology applications and human users. In the case of human users there is also the added dimension of presentation: metalanguage, visual appearance, instructional text, etc.

Regarding the linking issues, the results presented in the previous section are mostly encouraging for future attempts to align the three resources, while also indicating several non-trivial problems to handle. For all but a few of the English construction entries, linkable construction equivalents were found in both Brazilian Portuguese and Swedish. Furthermore, the vast majority of the constructional pairings qualify as functionally equivalent, in the sense that the relation can be treated as a one-to-one correspondence. Hence, it seems that the mapping should be a relatively simple matter in most of the cases, at least as long as the source language is English.

It should be borne in mind, however, that the established correspondences are unidirectional. Just because a link holds from English to Brazilian Portuguese or from English to Swedish it does not necessarily follow that the relation is equally straight-forward in the opposite direction, let alone between Brazilian Portuguese and Swedish. This will have to be tested. Still, the high degree of correspondence found suggests that such analogous explorations should be fruitful. At the same time, however, the remaining – blessedly few – non-matches, and the non one-to-one mappings, are challenges that will need to be handled.

Turning to representation, the main issue is how to represent the structure of the constructions in a cross-linguistically applicable way. To what extent, and in what way, structural differences should be explicitly indicated depends on the intended application. Whereas functional differences (non straight-forward linking relations) should be highly relevant for most purposes, formal differences, to the extent they are relevant, may be derivable from the descriptions as such – depending on the description format. For some purposes, however, it is probably useful to point out salient formal differences directly, especially regarding any forms of language pedagogy.

In the following, we will treat computational alignment and adaptation for human users separately. The linking issues will mainly be treated in the language technology Section (4.1), since the resources are digital and the linking will thus be handled computationally even for human-oriented uses. Matters of representation will be addressed from both perspectives, whereas questions regarding presentation are particular to application for humans (Section 4.2).

4.1   Computational alignment

The comparison presented in Section 3 is concerned with linguistic units and linguistic structure, not the database structure of the three resources. Computational alignment of constructicons, on the other hand, would deal with relations between database entries. Thus, the core issue would be how the linguistic relations distinguished in the analyses are to be modeled. In terms of linking, the mapping relations to implement (or at least consider) are of three kinds:

1. one-to-one (functionally equivalent pairings)
2. non one-to-one (correspondence but not functional equivalence)
3. non correspondence (possibly a non-mapping, but still something that needs to be handled).

The other side of the alignment problem is how to represent the entities (the cxn entries) in a useful way. What properties of the cxns need to be represented and how are these to be formalized?

To make an informed decision on which aspects of constructions should be modeled and which kinds of links should be proposed, one would need first to determine the purpose of the alignment effort. For example, if the alignment is meant for language understanding and not language production, then partially correspondent constructions may be linked in a less strict way, allowing the system to generate semantic interpretations of the input. On the other hand, if the task also involves language generation, as is the case for machine translation systems, then, partial correspondences should either involve language internal rules or rely on a statistical model of the target language to reorder text sequences.

In computational terms, a constructicon may be implemented in a relational database where constructions and their constituents are represented as entries in tables. Relations between those can be either directly written in the database, or modeled using intermediate tables. In the first case, the existence of some connection between the two entries being related is stated, but no information about the relation itself is provided. In the second case, it is possible to model information about the relation itself, such as the type of the relation or the constraints applying to it. Relations of this type can model both constructional constituency, that is, represent how a construction is composed in terms of its elements, and links between constructions.

In this scenario, the alignment of constructional databases can be implemented by proposing new relation types connecting them. Following Vossen (1998), Lönneker-Rodman (2007) presents two methodologies for aligning lexical resources: the expand approach and the merge approach. In the first, already mentioned in 1.2, the structural backbone of the source resource is transferred to the target language and populated with language-specific data. Adaptations of the original structure may be needed, based on the analysis of the language material in the target resource. In the second, two already existing resources are mapped to each other based on correspondences in their nodes. Alignment may be partial, since the resources were built independently.

The comparison between the entries in the EngCcn and their corresponding pairs in the Brazilian Portuguese and Swedish pursued in the previous session indicates that no trivial expand approach from the EngCcn is possible if one wants to both link the similar aspects of constructions and stress the relevant differences between them. Therefore, the merge approach seems to be the most suitable for aligning constructional resources.

The easiest cases in this scenario are those in which a given construction in the source language finds a correspondent structure in the target language, and such

correspondence holds for both the functional scope and the formal aspects of the construction. To computationally represent these one-to-one form-function correspondences, a relation between the two constructions could be modeled. If the databases being connected are relational, it means that the construction entries in them are entities, and relations between those entities would be modeled through tables building correspondences between the entities' internal structures.

Moreover, one would also need to decide whether the type of correspondence relation just sketched is uni- or bidirectional.

Additional complications arise, when the kind of link to be modeled holds between constructions with partial correspondence, that is, constructions that overlap to some extent in their functional scope but also play somewhat different roles in their respective languages. Such cases usually lead to one-to-many links in the database, since the source construction may end up being partially correspondent to two or more structures in the target language.

Finally, cases of non-correspondence, such as those involving the *way* cxns for Brazilian Portuguese and the `be_present-participle` cxn for Swedish, for example, call for yet another decision: that of whether cases of non-correspondence will be addressed at all. Taking the *way* cxn as an example, Brazilian Portuguese would use two constructions to properly express the general meaning of this construction. If one decides to computationally represent that sort of relation, besides modeling the links between the source construction and the target constructions, one would also need to model a language internal rule connecting the two or more target constructions.

Lönneker-Rodman (2007) also points out that there are two scales against which the effort of aligning resources should be evaluated: organizational similarity and interrelatedness. The first refers to the underlying principles that guided the development of the resource, while the second concerns the possibility of connecting not only the resources and entities in them as wholes, but also their constituent parts. Hence, all the choices regarding how to link constructions across language models are dependent on how the resources represent constructions computationally, both in terms of their constituency, and in terms of granularity, that is, in terms of which aspects of the form and the meaning of the constructions will be expressed in the representation.

So far, no attempt has been made of computationally aligning constructional resources. However, following the path designed for aligning lexical resources, FrameNet Brasil included a set of relations and constraints in the Brazilian Portuguese Constructicon (BPCcn) that may be useful for pursuing the computational alignment of constructicons. In the following, we will outline the way such relations and constraints are modeled in FrameNet Brasil.

4.1.1 Relations and constraints in FrameNet Brasil

FrameNet Brasil has been developing computational tools to model relations and constraints applying to constructions. Two relations and four constraint types have already been modeled:

1. the Constructional Inheritance relation
2. the Construction to Frame relation
3. the Construction Element to Construction constraint
4. the Construction Element to Frame constraint
5. the Construction Element to Frame Family constraint
6. the Construction Element ordering constraint.

Constructional Inheritance models the fact that a given construction in the database has all the properties of its parent construction plus some other specific properties. This relation reduces the modeling effort, to the extent that general properties of a given construction have to be modeled only once. Through the Construction to Frame relation, on the other hand, constructions can be linked to the frame(s) they evoke – if they evoke a frame at all – and, if this is the case, a CE to FE mapping can be proposed. For example, consider the Brazilian Portuguese `non-agentive_intransitive` cxn in (4).

(4) [O vaso$_{Subj}$] [quebrou$_{Pred}$]
    The vase break.PST
    The vase broke

This construction features two constituents: the SUBJECT and the PREDICATE. Any regular constructionist approach would claim that the SUBJECT of the `non-agentive_intransitive` cxn must have the property of being a Patient affected by the verb in the PREDICATE.

Because the BPCcn is directly connected to the frames database of FrameNet Brasil, it is possible to associate this construction to the `Undergoing` frame,[22] which is defined as follows: "An ENTITY is affected by an EVENT". The Evoking relation would then connect the CEs in the construction to their corresponding meaning, represented as the FEs in the frame: the SUBJECT CE is mapped to the ENTITY FE, while the PREDICATE CE is mapped to the EVENT FE. Such a mapping models the external semantic properties of the construction, while the constraints model internal aspects of both its form and meaning.

---

22. Frame names are written in a sans serif font (`Consolas`) with an initial capital letter, similar to cxns in the first regard but distinctive from them in the second. Frame elements (FEs) are distinguished from construction elements (CEs) in the same manner, both being written in SMALL CAPS but only FEs with an initial capital letter.

The CE to Construction constraint stores the information that a given CE is licensed by another construction. For the non-agentive_intransitive cxn, this constraint models the fact that the SUBJECT CE is a determined_NP, while the PREDICATE CE is an intransitive_VP. Figure 2 shows a graphic representation of the relation and the constraint discussed so far.
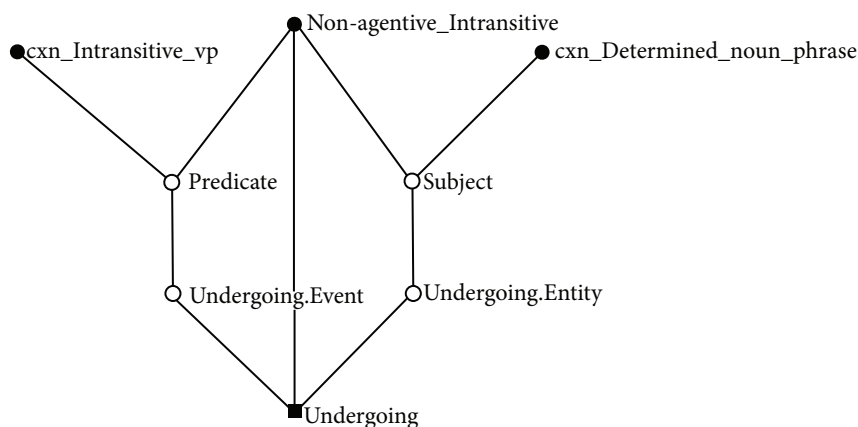
Figure 2. Graphic representation of the non-agentive_intransitive cxn in BPCcn

The following two constraints model slot filling restrictions in constructions. In general, they delimit the set of lexical items that can fill the head of a given CE to, in the case of (4), the LUs evoking a given frame, and, in the case of (5), the LUs evoking a given frame plus all the LUs evoking frames that inherit from that. Those constraints were conceived as a means to model semantic restrictions on the constituents of a construction. For example, when a given construction requires its subject to be human, or its verbal head to be of a certain semantic type.

The last constraint models constituent order and is applied to constructions in which word order is fixed.

Although these relations and constraints were initially conceived for better representing the properties of constructions, they can play a role in alignment across languages. The assumption that two constructions in two different languages evoke the same frame but have formal differences could be modeled, for example, as both of them having an evoke relation with the same frame, but, on the other hand, have different constraint configurations. For non frame-evoking constructions (see Fillmore, 1999; Fillmore, Lee-Goldman, & Rhomieux, 2012; Lyngfelt, Bäckström et al., this volume; Ohara, this volume), other relations and constraints would be required. Similarly, the kinds of external differences analyzed in Section 3.1 and 3.3 – those leading to the 2.5 and 1.5 scores – could to some extent be modeled via

the Inheritance relations connecting constructions to their parents. Finally, new cross-linguistically oriented constraints may be added to the database, as a means of highlighting salient differences or similarities between constructions. How far the current set holds, and what additions will be needed, is to be determined through actual implementation.

4.2   Resources for human users

From the view-point of a human user, the representation of constructions and their properties is secondary to the presentation of the resource in general. The first crucial factor is the basic meta-language. All the existing constructicons (ccns) are monolingual resources, mainly presented in the same language as they are describing.[23] This means that a simple link from, say, a Swedish cxn entry to a corresponding Brazilian Portuguese entry would take the user from a resource presented in Swedish to one in Portuguese. Hence, any multilingual application useful for humans would require either translations between the languages involved or a meta-language common to all the connected resources – in which case English would be the only realistically plausible *lingua franca*.

An almost equally important factor is the design of the user interface. Even if all the ccns concerned were translated into English, the fact that they are structured and presented differently would remain a daunting threshold. A user being directed from one ccn to the other would be cast into a new kind of environment, the interpretation of which would require a great deal of adaptation, even if the same language is employed in both ccns.

Consequently, any multilingual application intended for human users would require the development of a multilingual infrastructure with some kind of common interface. Merely connecting the existing ccns could never be sufficient, regardless of how it is done.

Presuming these obstacles are overcome, the development of a usable, let alone user-friendly, multilingual resource would also face the same challenges as any comparable monolingual resource: making it answer to the needs and convenience of the intended users. Thus, the design features required would depend on what the resource is to be used for, on the one hand; and on what previous knowledge can be expected from the user, on the other. In this regard, constructicography is largely analogous to lexicography (cf. Atkins & Rundell, 2008; Svensén, 2009), especially e-lexicography (L'Homme, 2014), except that constructicography also concerns grammatical structure.

---

23. Both the Swedish and the Russian constructicon also provide some metatext in English.

This is where the representation format comes in. The conventional representations of syntactic structure employed in valence dictionaries (e.g. Herbst et al., 2004), while attractively simple, would need to be enriched to be able to account for more complex cxns, and they are clearly insufficient to account for structural differences between languages. Linguistic standards such as the Leipzig glossing rules (Lehmann, 1982), on the other hand, are well adapted for the latter purpose but require a familiarity with linguistic meta-language rare outside academia. Hence, the general knowledge of linguistic terminology and description formats is a strongly limiting factor, and the usability of some functions may in practice be restricted to language professionals or even linguists. That is, unless some mode of representation less dependent on technical terminology is developed.

5.   Concluding remarks

In this chapter we have approached the prospects for multilingual constructicon application, chiefly by means of a comparison between English, Brazilian Portuguese and Swedish. Starting out from the existing cxn entries in the English ccn, we have explored to what extent corresponding target cxns may be distinguished in Brazilian Portuguese and Swedish. After establishing (or failing to establish) approximate correspondences, we have also for each pairing recorded functional and formal differences between the source and the target cxn.

This highly explorative study was meant to give an indication of whether alignment of ccns across languages is a fruitful path to pursue, and shed some light on what kinds of possibilities and challenges it involves. Naturally, a desired outcome was also that the dataset obtained could serve as startup material for such a development, and the study was designed accordingly.

A secondary purpose was to develop a format for comparison that is useful for future work in this direction. This means on the one hand that the methodology is applicable to comparison with other languages as well as other cxns, and on the other hand that the information recorded is both relevant and sufficient for the actual alignment. We are, however, well aware that the last point is highly dependent on the purpose of the alignment, but hope that the present format may at least serve as a point of departure for future adjustments.

Two important limitations should be borne in mind here. First, correspondence in this context is a matter of linkability between resources, in the same basic sense as in bilingual lexicography. It does not imply any presumptions of full equivalence. Thus, the study is not fully comparable with typical contrastive CxG accounts, due to its focus on approximate correspondence and its fairly rough mode of comparison. Neither is it fully comparable with bilingual lexicography, which is mainly

concerned with semantic (and, to a lesser extent, pragmatic) correspondence, whereas the current work involves grammatical structure as well and thus also a formal comparison. Rather, the approach is a blend between the two, which is why we have adopted the label (interlingual) *constructicography*.

Second, the comparison is unidirectional, not only being conducted on the terms of English but also taking the analyses in the Berkeley English ccn for granted. Such asymmetry between the source language and the target languages follows traditional practice in bilingual lexicography, and it was justified in the days of paper lexicography, when directionality was a precondition for a (paper) dictionary. Electronic resources are not subject to the same limitations, and the user can go back and forth, effectively making the role distinction between source and target language apply to individual operations rather than to the resources in general. However, the methods in lexicography have not yet caught up with the technical developments, and in constructicography even less so. It is obvious that many desired uses for multilingual ccn application would require the present work to be complemented by additional investigations taking the opposite perspective or, better still, taking both directions into account.

Bearing this in mind, the results from the comparison are mostly encouraging, at least as regards linkability and LT application. With only a scant few exceptions, corresponding target cxns could be established in both Brazilian Portuguese and Swedish, and the matchings obtained should serve as a good test set for developing a connecting infrastructure. Such testing will include how to handle formal and functional discrepancies, as well as the development of a cross-linguistically applicable description format. An important aspect to consider in this regard is suitability for and adaptation to different kinds of application.

For applications pertaining to human users, on the other hand, the basic problem to solve has less to do with linking and more with presentation. Clearly, the user of one ccn resource, in a certain language and with a certain user's interface, would in most cases need better help than a mere link to another ccn resource in a different language and with a different interface. Therefore, multilingual ccn applications for humans would require a common platform, preferably with the same organization, description format, and interface for all the languages involved, and either using the same meta-language or having translations of all the information presented. Hence, the main task is not about connecting existing resources but rather developing a new one. While a desirable goal for future development, it is probably not the immediate next step.

Acknowledgments

References

Adamska-Sałaciak, A. (2010). Examining equivalence. *International Journal of Lexicography*, 23(4), 387–409. doi:10.1093/ijl/ecq024

Almeida, V. G. (2016). *Identificação Automática de Construções de Estrutura Argumental: um experimento a partir da modelagem linguístico-computacional das construções Transitiva Direta Ativa, Ergativa e de Argumento Cindido* ['Automatic Identification of Argument Structure Constructions: an experiment using the Direct Active Transitive Cxn, the Ergative Cxn and the Split Argument Construction']. M.A. Thesis. Federal University of Juiz de Fora.

Atkins, B. T. S., & Rundell, M. (2008). *The Oxford Guide to Practical Lexicography*. Oxford/New York: Oxford University Press.

Bäckström, L., Lyngfelt, B., & Sköldberg, E. (2014). Towards Interlingual Constructicography. On correspondence between constructicon resources for English and Swedish. *Constructions and Frames*, 6(1), 9–32. doi:10.1075/cf.6.1.02bak

Barðdal, J. (2004). The semantics of the impersonal construction in Icelandic, German and Faroese: Beyond thematic roles. In W. Abraham (Ed.), *Focus on Germanic Typology* (pp. 101–130). Berlin: Akademie Verlag.

Boas, F. (1911). Introduction. *Handbook of American Indian languages*, Vol. 1 (pp. 5–83). Washington: Government Print Office.

Boas, H. C. (2002). Bilingual FrameNet Dictionaries for Machine Translation. In *Proceedings of the Third International Conference on Language Resources and Evaluation* (pp. 1364–1371). Las Palmas, Spain: European Language Resources Association.

Boas, H. C. (2005). Semantic Frames as Interlingual Representations for Multilingual Lexical Databases. *International Journal of Lexicography*, 18(4), 445–478. doi:10.1093/ijl/eci043

Boas, H. C. (2008). Determining the structure of lexical entries and grammatical constructions in Construction Grammar. *Annual Review of Cognitive Linguistics*, 6, 113–144. doi:10.1075/arcl.6.06boa

Boas, H. C. (Ed.) (2009). *Multilingual FrameNets in Computational Lexicography: Methods and Applications*. Berlin/New York: Mouton de Gruyter. doi:10.1515/9783110212976

Boas, H. C. (2010a). Comparing constructions across languages. In H. C. Boas (Ed.), *Contrastive Studies in Construction Grammar* (pp. 1–20). Amsterdam/Philadelphia: John Benjamins. doi:10.1075/cal.10.02boa

Boas, H. C. (Ed.) (2010b). *Contrastive Studies in Construction Grammar*. Amsterdam/Philadelphia: John Benjamins. doi:10.1075/cal.10

Boas, H. C., & Dux, R. (2013). Semantic frames for foreign-language education: Towards a German frame-based dictionary. *Veredas*, 17(1), 81–100.

Burchardt, A., Erk, K., Frank, A., Kowalski, A. Padó, S., & M. Pinkal, M. (2006). The SALSA Corpus: a German corpus resource for lexical semantics. *Proceedings of the LREC 2006*. Genoa, Italy: European Language Resources Association.

Croft, W. (2001). *Radical Construction Grammar. Syntactic theory in typological perspective*. Oxford/New York: Oxford University Press. doi:10.1093/acprof:oso/9780198299554.001.0001

Dooley, S. (2014). The Swedish Comparative Correlative Construction: *Ju … Desto …* and Variations. In *Constructions* (special issue on Swedish constructions, Ed. by B. Lyngfelt & C. Wide). Available online at <http://constructions-journal.com/>

Farø, K (2004). Hvornår går man over åen efter vand? Idiomatiske ækvivalensproblemer i leksikografi og leksikologi. Med dansk og tysk som exempel ['When do you cross the river for water? Issues of idiomaticity and equivalence in lexicography and lexicology. The example of Danish and German']. *LexicoNordica*, 11, 85–108.

Fellbaum, C. (1998). *WordNet: an electronic lexical database*. Cambridge, Mass.: MIT Press.

Fillmore, C. J. (1999). Inversion and Constructional Inheritance. In G. Webelhuth, J. P. Koenig & A. Kathol. *Lexical and Constructional Aspects of Linguistic Explanation* (pp. 113–128). Stanford: CSLI.

Fillmore, C. J. (2008). Border Conflicts: FrameNet Meets Construction Grammar. In E. Bernal & J. DeCesaris (Eds.), *Proceedings of the XIII EURALEX International Congress* (pp. 49–68). Barcelona: Universitat Pompeu Fabra.

Fillmore, C. J. (2013). Berkeley Construction Grammar. In Th. Hoffmann & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 111–132). Oxford/New York: Oxford University Press.

Fillmore, C. J., & Baker, C. (2010). A Frames Approach to Semantic Analysis. In B. Heine & H. Narrog (Eds.), *Oxford Handbook of Linguistic Analysis*. Oxford: Oxford University Press.

Fillmore, C. J., Johnson, C. R., & Petruck, M. R. L. (2003). Background to FrameNet. *International Journal of Lexicography*, 16(3), 235–250. doi:10.1093/ijl/16.3.235

Fillmore, C. J., Petruck, M. R. L., Ruppenhofer, J., & Wright, A. (2003). FrameNet in action: the case of attaching. *International Journal of Lexicography*, 16(3), 297–332. doi:10.1093/ijl/16.3.297

Fillmore, C. J., Lee-Goldman, R., & Rhomieux, R (2012). The FrameNet Constructicon. In H. C. Boas & I. A. Sag (Eds.), *Sign-Based Construction Grammar* (pp. 309–372). Stanford: CSLI Publications.

Fischer, O., van Kemenade, A., Koopman, W., & van der Wurff, W. (2000). *The Syntax of Early English*. Cambridge/New York: Cambridge University Press.

FrameNet. <https://framenet.icsi.berkeley.edu/fndrupal/>

Fried, M. (2006). Agent back-grounding as a functional domain: reflexivization and passivization in Czech and Russian. In B. Lyngfelt & T. Solstad (Eds.), *Demoting the Agent: passive, middle, and other voice phenomena* (pp. 83–109). Amsterdam/Philadelphia: John Benjamins. doi:10.1075/la.96.06fri

Fried, M., & Östman, J.-O. (Eds.) (2004). *Construction Grammar in a cross-language perspective*. Amsterdam/Philadelphia: John Benjamins. doi:10.1075/cal.2

Goldberg, A. E. (2013). Constructionist approaches. In Th. Hoffmann & G. Trousdale (Eds.), *The Oxford Handbook of Construction Grammar* (pp. 15–31). Oxford/New York: Oxford University Press.

Hanks, P. (2000). Do word meanings exist? *Computers and the Humanities*, 34, 205–215. doi:10.1023/A:1002471322828

Hannesdóttir, A. H. (2015). What is a target language in an electronic dictionary? In I. Kosem, M. Jakubiček, J. Kallas & S. Krek (Eds.), *Electronic lexicography in the 21st century: linking lexical data in the digital age*. Proceedings of the eLex 2015 conference, 11–13 August 2015,

United Kingdom. Ljubljana/Brighton: Trojina, Institute for Applied Slovene Studies/Lexical Computing Ltd. <https://elex.link/elex2015/proceedings/eLex_2015_15_Hannesdottir.pdf>

Hasegawa, Y., Lee-Goldman, R., & Fillmore, C. J. (2014). On the universality of frames: Evidence from English-to-Japanese translations. *Constructions and Frames*, 6(2), 170–201. doi:10.1075/cf.6.2.03has

Haspelmath, M. (2007). Pre-established categories don't exist: Consequences for language description and typology. *Linguistic Typology*, 11(1), 119–132. doi:10.1515/LINGTY.2007.011

Herbst, Th., Heath, D., Roe, I. F., & Götz, D. (2004). *A Valence Dictionary of English*. Berlin: de Gruyter Mouton. doi:10.1515/9783110892581

Hilpert, M. (2008). *Germanic future constructions*. Amsterdam/Philadelphia: John Benjamins. doi:10.1075/cal.7

Hilpert, M., & Östman, J. -O. (Eds.) (2014). Reflections on Constructions across Grammars. Special issue of *Constructions and Frames*, 6(2). doi:10.1075/cf.6.2.01int

Höder, S. (2012). Multilingual constructions: A diasystematic approach to common structures. In K. Braunmüller & Ch. Gabriel (Eds.), *Multilingual individuals and multilingual societies* (pp. 241–257). Amsterdam/Philadelphia: John Benjamins. doi:10.1075/hsm.13.17hod

Höder, S. (2014). Phonological elements and Diasystematic Construction Grammar. *Constructions and Frames*, 6(2), 202–231. doi:10.1075/cf.6.2.04hod

Kilgariff, A. (1997). I don't believe in word senses. *Computers and the Humanities*, 31(2), 91–113. doi:10.1023/A:1000583911091

Koller, W. (2011). *Einführung in die Übersetzungswissenschaft*. 8., neuarbeitete Auflage. Unter Mitarbeit von Kjetil Berg Henjum. Tübingen & Basel: A Francke Verlag.

Kromann, H. -P., Riiber, Th., & Rosbach, P. (1991). Principles of Bilingual Lexicography. In H. Steger & H. E. Wiegand (Eds.), *Wörterbücher/Dictionaries/Dictionnaires. Ein internationals Handbuch zur Lexikographie*, 1989–91 (Handbücher zur Sprach- und Kommunikationswissenschaft), Band 5.1–3 (pp. 2711–2728). Berlin & New York: Walter de Gruyter.

Kuzar, R. (2012). *Sentence Patterns in English and Hebrew*. Amsterdam/Philadelphia: John Benjamins. doi:10.1075/cal.12

Lakoff, G. (1987). *Women, Fire, and Dangerous Things: What categories reveal about the mind*. Chicago: University of Chicago Press. doi:10.7208/chicago/9780226471013.001.0001

Laviola, A. B. (2015). *Frames e Construções em Contraste: uma análise comparativa português-inglês no tangente à implementação de constructicons* ['Frames and Constructions in Contrast: a Portuguese-English comparative analysis in regards to the implementation of constructicons']. MA Thesis, Federal University of Juiz de Fora.

Lehmann, Ch. (1982). Directions for interlinear morphemic translations. *Folia Linguistica*, 16, 199–224. doi:10.1515/flin.1982.16.1-4.199

Lew, R. (2013). Identifying, Ordering and Defining Senses. *The Bloomsbury Companion to Lexicography* (pp. 284–302). London, New Delhi, New York, Sydney: Bloomsbury.

Lönneker-Rodman, B. (2007). *Multilinguality and FrameNet*. Technical Report. Berkeley: International Computer Science Institute.

L'Homme, M. -C. (2014). Why Lexical Semantics Is Important for E-Lexicography and Why It Is Equally Important to Hide Its Formal Representations from Users of Dictionaries. *International Journal of Lexicography*, 27(4), 360–377. doi:10.1093/ijl/ecu019

Ohara, K. H., Fujii, S., Ohori, T., Suzuki, R., Saito, H., & Ishizaki, S. (2004). The Japanese FrameNet Project: An Introduction. Proceedings of the LREC 2004 Satellite Workshop: Building Lexical Resources from Semantically Annotated Corpora. Lisbon, Portugal.

Peron-Corrêa, S. R., Diniz da Costa, A., Lara, M. S., Matos, E. E. S., & Torrent, T. T. (2016). FrameNet-Based Automatic Suggestion of Translation Equivalents. In J. Silva, R. Ribeiro, P. Quaresma, A. Adami & A. Branco (Eds.), *Computational Processing of the Portuguese Language* (pp. 347–352). Berlin: Springer.

Ralph, B. (1975). On the Nature of Preposition Deletion in Swedish. In K. -H. Dahlstedt (Ed.), *The Nordic Languages and Modern Linguistics 2*. Proceedings of the Second International Conference of Nordic and General Linguistics. University of Umeå, June 14–19, 1973 (pp. 666–684). Stockholm.

Ruppenhofer, J., Ellsworth, M., Petruck, M. R. L., Johnson, C. R., & Scheffczyk, J. (2016). *FrameNet II: extended theory and practice*. Berkeley: ICSI. <https://framenet2.icsi.berkeley. edu/docs/r1.7/book.pdf>.

Sato, H. (2008). New Functions of FrameSQL for Multilingual FrameNets. In Proceedings of the Sixth International Language Resources and Evaluation Conference (pp. 758–762). Marrakech, Morocco: European Language Resources Association.

Schmidt, T. (2009). The Kicktionary – A Multilingual Lexical Resource of Football Language. In H. C. Boas (Ed.), *Multilingual FrameNets in Computational Lexicography* (pp. 101–134). Berlin: Mouton de Gruyter.

Snell-Hornby, M. (1986). The Bilingual Dictionary – Victim of its own tradition? In R. R. K. Hartmann (Ed.), *The history of lexicography. Papers from the dictionary research seminar at Exeter, March 1986* (pp. 207–218). Amsterdam/Philadelphia: John Benjamins.

Subirats, C., & Petruck, M. R. L. (2003). Surprise: Spanish FrameNet! In Proceedings of the 17th International Conference of Linguists. Prague, Czech Republic.

Svensén, B. (2009). *A Handbook of Lexicography. The Theory and Practice of Dictionary-Making*. Cambridge & New York: Cambridge University Press.

Torrent, T. T., & Ellsworth, M. (2013). Behind the labels: criteria for defining analytical categories in FrameNet Brasil. *Veredas*, 17(1), 44–65.

Torrent, T. T., Lage, L. M., Sampaio, T. F., Tavares, T. S., & Matos, E. E. S. (2014). Revisiting border conflicts between FrameNet and Construction Grammar: annotation policies for the Brazilian Portuguese Constructicon. *Constructions and Frames*, 6(1), 34–51. doi:10.1075/cf.6.1.03tor

Torrent, T. T., Salomão, M. M. M., Campos, F. C. A., Braga, R. M. M., Matos, E. E. S., Gamonal, M. A., Gonçalves, J., Souza, B. P., Gomes, D. S., & Peron-Corrêa, S. R. (2014). Copa 2014 FrameNet Brasil: a frame-based trilingual electronic dictionary for the Football World Cup. In: Proceedings of COLING 2014, The 25th International Conference on Computational Linguistics: System Demonstrations (pp. 10–14). Dublin, Ireland: ACL.

Vossen, P. (1998). Introduction to EuroWordNet. *Computers and the Humanities*, 32 (2–3), 73–89 doi:10.1023/A:1001175424222

Wassenscheidt, Ph. (2014). Constructions do not cross languages: On cross-linguistic generalizations of constructions. *Constructions and Frames*, 6(2), 305–337. doi:10.1075/cf.6.2.07was

Zgusta, L. (1971). *Manual of Lexicography*. Janua Linguarum. Series maior. 39. Prague: Academia; The Hague & Paris: Mouton. doi:10.1515/9783111349183

Appendix. Summary of the contrastive analyses[24]

Source lang. English	Target lang.	Criteria				
		Correspon- dence	Functional equivalence	Formal similarity	All differences external	Total score
absolute_clause	Pt-Br	1	1	1	–	3
	Swe	1	1	0	0	2
	Pt-Br: Similar Swe: Functional equivalence on the present level of abstraction, less so regarding subtypes. Formally, less hospitable to gerunds (among other differences)					
adjective_as_nominal. abstract	Pt-Br	1	1	0	0.5	2.5
	Swe	1	1	0	0.5	2.5
	Pt-Br: Agreement in gender and number Swe: Agreement in gender (neuter) and definiteness (definite)					
adjective_as_nominal. anaphoric	Pt-Br	1	1	0	0.5	2.5
	Swe	1	1	0	0.5	2.5
	Pt-Br: Agreement in gender and number Swe: Agreement in gender, number and definiteness (definite)					
adjective_as_nominal. people	Pt-Br	1	1	0	0.5	2.5
	Swe	1	1	0	0.5	2.5
	Pt-Br: Agreement in gender and number Swe: Agreement in number (plural) and definiteness (definite)					
as.role	Pt-Br	1	1	0	0	2
	Swe	1	1	0	0.5	2.5
	Pt-Br: A temporal conjunction can also be used instead of the correlative conjunction Swe: Typically no definite article, characteristic of predicative uses in general					
attributive_degree_ modification	Pt-Br	1	1	0	0	2
	Swe	1	1	0	0	2
	Pt-Br: Article may be definite or indefinite. Different CE order. Swe: No article (nor preposition) between the adjective phrase and the nominal head					
be_present-participle	Pt-Br	1	1	1	–	3
	Swe	0	–	–	–	0
	Pt-Br: Similar Swe: No general cxn marking progressive aspect					

---

24. For full accounts of the English construction entries, see <http://www1.icsi.berkeley.edu/~hsato/cxn00/21colorTag/index.html>.

Source lang. English	Target lang.	Criteria				
		Correspon-dence	Functional equivalence	Formal similarity	All differences external	Total score
be_recip	Pt-Br	1	0	0	0.5	1.5
	Swe	1	0	1	–	2
	Pt-Br: Only symmetrical valence is possible. Asymmetrical valence does not entail reciprocality. Swe: Target cxn covers only the symmetric variant					
comparison	Pt-Br	1	1	0	0	2
	Swe	1	1	0	0.5	2.5
	Pt-Br: Comparison operator generally involves an adverb before the first conjunct and a *wh*-word before the second Swe: Agreement in gender and number					
comparison_equality	Pt-Br	1	1	0	0	2
	Swe	1	1	0	0.5	2.5
	Pt-Br: Comparison operator involves the intensifier *tanto/tao* 'so' before the first conjunct and *quanto* 'how much' before the second Swe: Agreement in gender and number					
comparison_equality_ metalinguistic	Pt-Br	1	1	0	0	2
	Swe	1	1	0	0	2
	Pt-Br: Comparison operator involves the intensifier *tanto/tão* 'so' before the first conjunct and *quanto* 'how much' before the second Swe: Two word order variants with partially different lexical marking					
comparison_inequality	Pt-Br	1	1	0	0.5	2.5
	Swe	1	1	0	0.5	2.5
	Pt-Br: Comparison operator involves the intensifiers *mais* 'more' or *menos* 'less' before the first conjunct and *(do) que* '(of) that' before the second Swe: Gender and number agreement in the periphrastic variant					
coordination	Pt-Br	1	1	1	–	3
	Swe	1	1	1	–	3
	Pt-Br: Similar Swe: Basically similar (although tends to be more hospitable to ellipsis; cf. gapping)					
degree_modification	Pt-Br	1	1	0	0	2
	Swe	1	1	0	0.5	2.5
	Pt-Br: Different morphologic and syntactic strategies in all types of degree modification Swe: Number and gender agreement in most variants					

Source lang. English	Target lang.	Criteria				
		Correspon-dence	Functional equivalence	Formal similarity	All differences external	Total score
`degree_so`	Pt-Br:	1	1	0	0.5	2.5
	Swe:	1	1	0	0.5	2.5
	Pt-Br: Agreement in number and gender					
	Swe: Agreement in number and gender; complementizer obligatory					
`deictic_dayname_inverted`	Pt-Br	1	1	0	0.5	2.5
	Swe	1	0	0	0	1
	Pt-Br: Agreement in number and gender					
	Swe: Mostly lacking this pattern, instead relying on general use of the non-inverted counterpart; also, no preposition preceding the noun phrase					
`determined_noun_phrase`	Pt-Br	1	1	0	0	2
	Swe	1	1	0	0	2
	Pt-Br: Agreement in gender and number (salient property of this cxn; hence the score of 2, not 2.5 as in the cxns inheriting from this one)					
	Swe: Agreement in gender, number, and definiteness (salient property of this cxn; hence the score of 2, not 2.5 as in the cxns inheriting from this one)					
`determined_proper_name`	Pt-Br	1	1	0	0.5	2.5
	Swe	1	1	0	0.5	2.5
	Pt-Br: Agreement in number and gender					
	Swe: Agreement in number and definiteness					
`dimension_conjunction`	Pt-Br	1	1	0	0.5	2.5
	Swe	1	1	1	–	3
	Pt-Br: Dimension expressed by a PP, not an AP, which is a general property of this type of modifier, hence the 0.5					
	Swe: Only lexical difference, but different part of speech (*gånger* 'times' instead on Eng. *by*)					
`exocentric_adjectival_ compound`	Pt-Br	1	1	0	0	2
	Swe	1	1	0	0.5	2.5
	Pt-Br: Uses PP modification instead of a compound					
	Swe: Agreement in gender, number and definiteness					
`gapping`	Pt-Br	1	1	1	–	3
	Swe	1	1	1	–	3
	Pt-Br: Similar					
	Swe: Similar, except for distributional differences (less restricted in Swe)					

Source lang. English	Target lang.	Criteria				
		Correspon-dence	Functional equivalence	Formal similarity	All differences external	Total score
`have_with`	Pt-Br	1	1	1	–	3
	Swe	1	1	0	0	2
	Pt-Br: Essentially similar, except that the preposition *com* 'with' may be cliticized to the pronoun Swe: A reflexive in the complement of *med* 'with'					
`infinitival_relative_modal`	Pt-Br	1	1	1	–	3
	Swe	1	1	1	–	3
	Pt-Br: Basically similar, except for known distributional differences Swe: Basically similar, probably distributional differences					
`integrated_appositive`	Pt-Br	1	1	1	–	3
	Swe	1	1	1	–	3
	Pt-Br: Similar Swe: Similar					
`inversion_with_preposed_element`	Pt-Br	1	0	0	0	1
	Swe	1	0	1	–	2
	Pt-Br: Functional restriction limiting the adverb to 'so'. Formal difference because there is no auxiliary inversion Swe: Different functional restrictions; also less of a rhetorical effect due to this word order being less marked					
`let_alone`	Pt-Br	1	1	0	0.5	2.5
	Swe	1	1	1	–	3
	Pt-Br: Conjunction must include complementizer, as is generally the case for conjunctions in BP Swe: Similar, except for lexical and distributional differences					
`location_in_calendar_subunit`	Pt-Br	1	1	0	0.5	2.5
	Swe	1	1	0	0	2
	Pt-Br: Agreement in gender and number Swe: Normally definite expressions for past dates and indefinite for future ones (both this cxn and `location_in_calendar_unit` are covered by the same cxn in SweCcn)					
`location_in_calendar_unit`	Pt-Br	1	1	0	0.5	2.5
	Swe	1	1	0	0	2
	Pt-Br: Agreement in gender and number Swe: Normally definite expressions for past dates and indefinite for future ones (both this cxn and `location_in_calendar_subunit` are covered by the same cxn in SweCcn)					

Source lang. English	Target lang.	Criteria				
		Correspondence	Functional equivalence	Formal similarity	All differences external	Total score
`measurement_plus_adjective`	Pt-Br	1	1	0	0	2
	Swe	1	1	0	0.5	2.5
	Pt-Br: BP uses only measurement_plus_PP Swe: Agreement in gender, number and definiteness					
`measurement_plus_prepositional_phrase`	Pt-Br	1	1	1	–	3
	Swe	1	0	0	0	1
	Pt-Br: Similar Swe: Not used with age expressions; complement of preposition is definite					
`noun-noun_compound`	Pt-Br	1	1	1	–	3
	Swe	1	1	0	0	2
	Pt-Br: Similar Swe: Commonly occurring with a linking morpheme					
`ones_very_eyes`	Pt-Br	1	1	0	0.5	2.5
	Swe	1	1	0	0	2
	Pt-Br: Agreement in gender and number Swe: Different idiom with a similar function					
`open_interrogative.non-subject`	Pt-Br	1	1	0	0	2
	Swe	1	1	0	0	2
	Pt-Br: No inversion Swe: Inversion not restricted to auxiliaries (hence, no *do*-support)					
`own_right`	Pt-Br	1	1	0	0	2
	Swe	1	0	0	0	1
	Pt-Br: Different idiom Swe: Related idioms with overlapping function but different restrictions					
`postpositive_adjective`	Pt-Br	1	0	1	–	2
	Swe	1	1	0	0	2
	Pt-Br: Postposed adjectives are the standard in BP, not a marked expression Swe: Different formal and distributional restrictions, but no obvious functional differences					
`proper_name_embellishments`	Pt-Br	1	1	0	0.5	2.5
	Swe	1	1	0	0	2
	Pt-Br: Agreement in gender and number Swe: Embellishments usually definite (except for certain titles)					

Source lang. English	Target lang.	Criteria				
		Correspon- dence	Functional equivalence	Formal similarity	All differences external	Total score
`rate.cost_time`	Pt-Br	1	1	1	–	3
	Swe	1	1	0	0	2
	Pt-Br: Similar Swe: Two different cxns, definite denominator in the most common version (same two cxns also corresponding to `rate.frequency` and `rate.speed`)					
`rate.frequency`	Pt-Br	1	1	0	0	2
	Swe	1	1	0	0	2
	Pt-Br: Only prepositional variant Swe: Two different cxns, definite denominator in the most common version (same two cxns also corresponding to `rate.cost_time` and `rate.speed`)					
`rate.mileage`	Pt-Br	1	1	0	0	2
	Swe	0	–	–	–	0
	Pt-Br: Only prepositional variant Swe: Structurally most similar correspondent takes opposite perspective; same perspective requires paraphrase					
`rate.speed`	Pt-Br	1	1	0	0	2
	Swe	1	1	0	0	2
	Pt-Br: Only prepositional variant Swe: Two different cxns, definite denominator in the most common version (same two cxns also corresponding to `rate.cost_time` and `rate.frequency`)					
`rather_than_coordination`	Pt-Br	1	1	0	0	2
	Swe	1	1	1	–	3
	Pt-Br: Different POS in the coordination structure Swe: Two lexical variants with slightly different meanings, both consistent with the general function of the English source cxn					
`shared_Completion`	Pt-Br	1	1	1	–	3
	Swe	1	1	1	–	3
	Pt-Br: Similar Swe: Similar					
`stripping`	Pt-Br	1	1	1	–	3
	Swe	1	1	1	–	3
	Pt-Br: Similar Swe: Similar					

Source lang. English	Target lang.	Criteria				
		Correspon-dence	Functional equivalence	Formal similarity	All differences external	Total score
subject_auxiliary_ inversion. closed_interrogative	Pt-Br	1	1	0	0	2
	Swe	1	1	0	0	2
	Pt-Br: No inversion Swe: Inversion not restricted to auxiliaries (hence, no *do*-support)					
subject_auxiliary_ inversion. conditional	Pt-Br	1	1	0	0	2
	Swe	1	1	1	–	3
	Pt-Br: No inversion Swe: Inversion not restricted to auxiliaries (hence, no *do*-support)					
subject_auxiliary_ inversion. emphatic_negative_ imperative	Pt-Br	1	1	0	0	2
	Swe	1	1	0	0	2
	Pt-Br: No auxiliary (yet, VS word order) Swe: No *do*-support (yet, VS word order)					
subject_auxiliary_ inversion. exclamation	Pt-Br	0	–	–	–	0
	Swe	0	–	–	–	0
	Pt-Br: No inversion Swe: Inversion not restricted to auxiliaries (hence, no *do*-support)					
subject_auxiliary_ inversion	Pt-Br	0	–	–	–	0
	Swe	1	0	0	0	1
	Pt-Br: No subject auxiliary inversion Swe: Functional differences due to general V2 (verb second); formally, not restricted to auxiliaries (hence, no *do*-support)					
subject_auxiliary_ inversion. optative	Pt-Br	1	1	0	0	2
	Swe	1	1	1	–	3
	Pt-Br: No inversion Swe: Despite general differences between inversion patterns in Eng and Swe, this particular cxn actually similar (except for distributional differences)					
superlative	Pt-Br	1	1	0	0.5	2.5
	Swe	1	1	0	0.5	2.5
	Pt-Br: Different N-Adj word order, derived from more general AdjP cxn Swe: Definite marking on adjective and noun, agreement in gender and number on the determiner					
supplement_ascriptional	Pt-Br	1	1	1	–	3
	Swe	1	1	1	–	3
	Pt-Br: Similar Swe: Similar					

Source lang. English	Target lang.	Criteria				
		Correspon-dence	Functional equivalence	Formal similarity	All differences external	Total score
supplement_ specificational	Pt-Br	1	1	1	–	3
	Swe	1	1	1	–	3
	Pt-Br: Similar Swe: Similar					
tagged_sentence. canonical	Pt-Br	1	1	0	0.5	2.5
	Swe	1	1	0	0	2
	Pt-Br: Usually does not express the subject, nor inverts its position with that of the auxiliary. Both distinctions, however, follow from more general cxns. Swe: Employs fixed phrase (*eller hur?*) instead of pronominal question					
tagged_sentence	Pt-Br	1	1	0	0.5	2.5
	Swe	1	1	0	0	2
	Pt-Br: Usually does not express the subject, nor inverts its position with that of the auxiliary. Both distinctions, however, follow from more general cxns. Swe: Employs fixed phrase (*eller hur?*) instead of pronominal question					
tagged_sentence_ subjectless	Pt-Br	1	1	0	0.5	2.5
	Swe	0	–	–	–	0
	Pt-Br: Usually does not express the subject, nor inverts its position with that of the auxiliary. Both distinctions, however, follow from more general cxns. Swe: Tagged sentences not used in this way; no directly corresponding cxn					
tautology.deep_tautology	Pt-Br	1	1	1	–	3
	Swe	1	1	1	–	3
	Pt-Br: Similar Swe: Similar					
tautology.nominal	Pt-Br	1	1	1	–	3
	Swe	1	1	1	–	3
	Pt-Br: Similar Swe: Similar					
tautology.nostalgia	Pt-Br	1	1	1	–	3
	Swe	1	1	1	–	3
	Pt-Br: Similar Swe: Similar					

Source lang. English	Target lang.	Criteria				
		Correspon-dence	Functional equivalence	Formal similarity	All differences external	Total score
`there.presentational`	Pt-Br	1	1	0	0	2
	Swe	1	1	0	0	2
	Pt-Br: Subjectless *haver* 'exist' or *ter* 'have' used instead of *there* 'to be' Swe: Less restricted than the English source cxn (does not count as a functional difference from a unidirectional Eng > Swe perspective); employs expletive *det* 'it' rather than *där* 'there'.					
`there_be_a_time_when`	Pt-Br	1	1	0	0	2
	Swe	1	1	0	0	2
	Pt-Br: Subjectless *haver 'exist'* or *ter 'have'* used instead of *there* 'to be' Swe: Different kind of expletive (*det* 'it'); different verbs (also different way of expressing future, which is, however, an external difference)					
`uniqueness`	Pt-Br	1	1	0	0.5	2.5
	Swe	1	1	0	0	2
	Pt-Br: Agreement in gender and number Swe: Finite relative clause instead of infinitival relative (also definiteness marking and agreement)					
`way_manner`	Pt-Br	0	–	–	–	0
	Swe	1	0	0	0	1
	Pt-Br: Two argument structure cxns needed to express this function Swe: Formally closest variant (reflexive without *way*) more restricted; there's also an alternative pattern with a particle cxn					
`way_means`	Pt-Br	0	–	–	–	0
	Swe	1	1	0	0	2
	Pt-Br: Two argument structure cxns needed to express this function Swe: Reflexive cxn without *way*					
`way_neutral`	Pt-Br	0	–	–	–	0
	Swe	1	1	0	0	2
	Pt-Br: Two argument structure cxns needed to express this function Swe: Reflexive cxn without *way*					
`what-with_absolute`	Pt-Br	1	0	0	0	1
	Swe	1	0	0	0	1
	Pt-Br: Cxn lacking, covered by general `with_absolute` cxn Swe: Cxn lacking, covered by general `with_absolute` cxn					

Source lang. English	Target lang.	Criteria				
		Correspondence	Functional equivalence	Formal similarity	All differences external	Total score
with_absolute	Pt-Br	1	1	1	–	3
	Swe	1	1	0	0	2

Pt-Br: Similar

Swe: When complement of *med* 'with' is associated with a contextually salient referent, which is commonly the case, it is marked by only a definite article, or sometimes a possessive reflexive, instead of an ordinary possessive pronoun (as in e.g. English)

General index

Index of constructions

In constructionist theory, a constructicon is an inventory of constructions making up the full set of linguistic units in a language. In applied practice, it is a set of construction descriptions – a "dictionary of constructions". The development of constructicons in the latter sense typically means combining principles of both construction grammar and lexicography, and is probably best characterized as a blend between the two traditions. We call this blend constructicography.

The present volume is a comprehensive introduction to the emerging field of constructicography. After a general introduction follow six chapters presenting constructicon projects for English, German, Japanese, Brazilian Portuguese, Russian, and Swedish, respectively, often in relation to a framenet of the language. In addition, there is a chapter addressing the interplay between linguistics and language technology in constructicon development, and a final chapter exploring the prospects for interlingual constructicography.

This is the first major publication devoted to constructicon development and it should be particularly relevant for those interested in construction grammar, frame semantics, lexicography, the relation between grammar and lexicon, or linguistically informed language technology.

John Benjamins Publishing Company