10 Construction Grammar

Like LFG and HPSG, *Construction Grammar* (CxG) forms part of West Coast linguistics. It has been influenced considerably by Charles Fillmore, Paul Kay and George Lakoff (all three at Berkeley) and Adele Goldberg (who completed her PhD in Berkeley and is now in Princeton) (Fillmore 1988; Fillmore, Kay & O'Connor 1988; Kay & Fillmore 1999; Kay 2002, 2005; Goldberg 1995, 2006).

Fillmore, Kay, Jackendoff and others have pointed out the fact that, to a large extent, languages consist of complex units that cannot straightforwardly be described with the tools that we have seen thus far. In frameworks such as GB, an explicit distinction is made between core grammar and the periphery (Chomsky 1981a: 8), whereby the periphery is mostly disregarded as uninteresting when formulating a theory of Universal Grammar. The criticism leveled at such practices by CxG is justified since what counts as the 'periphery' sometimes seems completely arbitrary (Müller 2014d) and no progress is made by excluding large parts of the language from the theory just because they are irregular to a certain extent.

In Construction Grammar, idiomatic expressions are often discussed with regard to their interaction with regular areas of grammar. Kay & Fillmore (1999) studied the *What's X doing Y?*-construction in their classic essay. (1) contains some examples of this construction:

(1) a. What is this scratch doing on the table?
    b. What do you think your name is doing in my book?

The examples show that we are clearly not dealing with the normal meaning of the verb *do*. As well as the semantic bleaching here, there are particular morphosyntactic properties that have to be satisfied in this construction. The verb *do* must always be present and also in the form of the present participle. Kay and Fillmore develop an analysis explaining this construction and also capturing some of the similarities between the WXDY-construction and the rest of the grammar.

There are a number of variants of Construction Grammar:

- Berkeley Construction Grammar (Fillmore 1988; Kay & Fillmore 1999; Fried 2015)
- Goldbergian/Lakovian Construction Grammar (Lakoff 1987; Goldberg 1995, 2006)
- Cognitive Grammar (Langacker 1987, 2000, 2008; Dąbrowska 2004)
- Radical Construction Grammar (Croft 2001)
- Embodied Construction Grammar (Bergen & Chang 2005)

- Fluid Construction Grammar (Steels & Beule 2006; Steels 2011)
- Sign-Based Construction Grammar (Sag 2010, 2012)

The aim of Construction Grammar is to both describe and theoretically explore language in its entirety. In practice, however, irregularities in language are often given far more importance than the phenomena described as 'core grammar' in GB. Construction Grammar analyses usually analyze phenomena as phrasal patterns. These phrasal patterns are represented in inheritance hierarchies (e. g. Croft 2001; Goldberg 2003b). An example for the assumption of a phrasal construction is Goldberg's analysis of resultative constructions. Goldberg (1995) and Goldberg & Jackendoff (2004) argue for the construction status of resultatives. In their view, there is no head in (2) that determines the number of arguments.

(2) Willy watered the plants flat.

The number of arguments is determined by the construction instead, that is, by a rule or schema saying that the subject, verb, object and a predicative element must occur together and that the entire complex has a particular meaning. This view is fundamentally different from analyses in GB, Categorial Grammar, LFG[1] and HPSG. In the aforementioned theories, it is commonly assumed that arguments are always selected by lexical heads and not independently licensed by phrasal rules. See Simpson (1983), Neeleman (1994), Wunderlich (1997), Wechsler (1997), and Müller (2002a) for corresponding work in LFG, GB, Wunderlich's Lexical Decomposition Grammar and HPSG.

Like the theories discussed in Chapters 5–9, CxG is also a non-transformational theory. Furthermore, no empty elements are assumed in most variants of the theory and the assumption of lexical integrity is maintained as in LFG and HPSG. It can be shown that these assumptions are incompatible with phrasal analyses of resultative constructions (see Section 21.2.2 and Müller 2006, 2007c). This point will not be explained further here. Instead, I will discuss the work of Fillmore and Kay to prepare the reader to be able to read the original articles and subsequent publications. Although the literature on Construction Grammar is now relatively vast, there is very little work on the basic formal assumptions or analyses that have been formalized precisely. Examples of more formal works are Kay & Fillmore (1999); Kay (2002), Michaelis & Ruppenhofer (2001), and Goldberg (2003b). The book by Jean-Pierre Koenig (1999) (formerly Berkeley) has been heavily influenced by CxG. Fillmore and Kay's revisions of their earlier work took place in close collaboration with Ivan Sag. The result was a variant of HPSG known as Sign-Based Construction Grammar (SBCG) (Sag 2010, 2012). See Section 10.6.2 for further discussion.

John Bryant, Nancy Chang, Eva Mok have developed a system for the implementation of Embodied Construction Grammar[2]. Luc Steels is working on the simulation of language evolution and language acquisition (Steels 2003). In personal communication

[1] See Alsina (1996) and Asudeh, Dalrymple & Toivonen (2008, 2013), however. For more discussion of this point, see Sections 21.1.3 and 21.2.2.

[2] See http://www.icsi.berkeley.edu/~jbryant/old-analyzer.html and Bryant (2003).

(p. M. 2007) he stated that is is a long way to go until robots finally will be able to learn to speak but the current state of the art is already impressive. Steels can use robots that have a visual system (camera and image processing) and use visual information paired with audio information in simulations of language acquisition. The implementation of Fluid Construction Grammar is documented in Steels (2011) and Steels (2012). The second book contains parts about German, in which the implementation of German declarative clauses and *w* interrogative clauses is explained with respect to topological fields (Micelli 2012). The FCG system, various publications and example analyses are available at: http://www.fcg-net.org/. Jurafsky (1996) developed a Construction Grammar for English that was paired with a probabilistic component. He showed that many performance phenomena discussed in the literature (see Chapter 15 on the Competence/Performance Distinction) can be explained with recourse to probabilities of phrasal constructions and valence properties of words. Bannard, Lieven & Tomasello (2009) use a probabilistic context-free grammar to model grammatical knowledge of two and three year old children.

10.1 General remarks on the representational format

In this section, I will discuss the mechanisms of Berkeley Construction Grammar (BCG). As I pointed out in Müller (2006), there are fundamental problems with the formalization of BCG. The details will be given in Section 10.6.1. While the framework was developed further into Sign-Based Construction Grammar (see Section 10.6.2) by its creators Kay and Fillmore, there are still authors working in the original framework (for instance Fried 2013). I will therefore present the basic mechanisms here to make it possible to understand the original ideas and put them into a broader context.

As we saw in Section 9.1.2, dominance relations in HPSG are modeled like other properties of linguistic objects using feature-value pairs. In general, CxG uses feature-value pairs to describe linguistic objects, but dominance relations are represented by boxes (Kay & Fillmore 1999; Goldberg 2003b):

(3) $\boxed{\begin{array}{l} \text{phon} \langle \textit{the man} \rangle \\[1ex] \boxed{\text{phon} \langle \textit{the} \rangle} \quad \boxed{\text{phon} \langle \textit{man} \rangle} \end{array}}$

The structure can be written using feature-value pairs as follows:

(4) $\begin{bmatrix} \textsc{phon} & \langle \textit{the man} \rangle \\ \textsc{dtrs} & \langle\, [\, \textsc{phon} \, \langle \textit{the} \rangle \,], [\, \textsc{phon} \, \langle \textit{man} \rangle \,] \,\rangle \end{bmatrix}$

10.1.1 The head-complement construction

Kay & Fillmore (1999) assume the following construction for the combination of heads with their complements:

(5) Head plus Complements Construction (HC)

role head
lex +

role filler
loc +

\+

A head is combined with at least one complement (the '+' following the box stands for at least one sign that fits the description in that box). LOC+ means that this element must be realized locally. The value of ROLE tells us something about the role that a particular element plays in a construction. Unfortunately, here the term *filler* is used somewhat differently than in GPSG and HPSG. Fillers are not necessarily elements that stand in a long-distance dependency to a gap. Instead, a *filler* is a term for a constituent that fills the argument slot of a head.

The verb phrase construction is a sub-construction of the head-complement construction:

(6) Verb phrase Construction:

cat v

role head
lex +

role filler
loc +
gf ¬subj

\+

The syntactic category of the entire construction is V. Its complements cannot have the grammatical function subject.

The VP construction is a particular type of head-complement construction. The fact that it has much in common with the more general head-complement construction is represented as follows:

(7) Verb phrase Construction with inheritance statement:

INHERIT HC
cat v

gf ¬subj

\+

This representation differs from the one in HPSG, aside from the box notation, only in the fact that feature descriptions are not typed and as such it must be explicitly stated in the representation from which superordinate construction inheritance takes place. HPSG – in addition to the schemata – has separate type hierarchies specifying the inheritance relation between types.

10.1.2 Representation of valence information

In Kay and Fillmore, valence information is represented in a set (VAL). The Valence Principle states that local filler-daughters have to be identified with an element in the valence set of the mother.[3] The Subset Principle states that the set values of the head-daughter are subsets of the corresponding sets of the mother. This is the exact opposite approach to the one taken in Categorial Grammar and HPSG. In HPSG grammars, valence lists at the mother nodes are shorter, whereas in Berkeley CxG at least as many elements are present on the mother node as on the head-daughter.

10.1.3 Semantics

Semantics in CxG is handled exactly the same way as in HPSG: semantic information is contained in the same feature structure as syntactic information. The relation between syntax and semantics is captured by using the same variable in the syntactic and semantic description. (8) contains a feature description for the verb *arrive*:

(8) Lexical entry for *arrive* following Kay & Fillmore (1999: 11):

$$\boxed{\begin{array}{ll} \text{cat} & \text{v} \\ \text{sem} & \left\{ \left[^{\text{I}} \begin{array}{ll} \text{\textsc{frame}} & \text{ARRIVE} \\ \text{\textsc{args}} & \{ \text{A} \} \end{array} \right] \right\} \\ \text{val} & \{ [\text{\textsc{sem}} \{ \text{A} \}] \} \end{array}}$$

Kay & Fillmore (1999: 9) refer to their semantic representations as a notational variant of the Minimal Recursion Semantics of Copestake, Flickinger, Pollard & Sag (2005). In later works, Kay (2005) explicitly uses MRS. As the fundamentals of MRS have already been discussed in Section 9.1.6, I will not repeat them here. For more on MRS, see Section 19.3.

10.1.4 Adjuncts

For the combination of heads and modifiers, Kay and Fillmore assume further phrasal constructions that are similar to the verb phrase constructions discussed above and create a relation between a head and a modifier. Kay and Fillmore assume that adjuncts also contribute something to the VAL value of the mother node. In principle, VAL is nothing more than the set of all non-head daughters in a tree.

10.2 Passive

The passive has been described in CxG by means of so-called linking constructions, which are combined with lexical entries in inheritance hierarchies. In the base lexicon, it is only listed which semantic roles a verb fulfils and the way in which these are realized

[3] Sets in BCG work differently from those used in HPSG. A discussion of this is deferred to Section 10.6.1.

is determined by the respective linking constructions with which the basic lexical entry is combined. Figure 10.1 gives an example of a relevant inheritance hierarchy. There is a

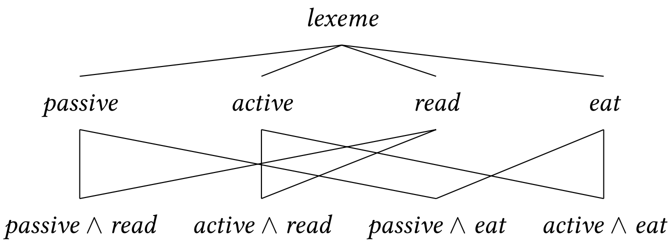

Figure 10.1: Passive and linking constructions

linking construction for both active and passive as well as lexical entries for *read* and *eat*. There is then a cross-classification resulting in an active and a passive variant of each verb.

The idea behind this analysis goes back to work by Fillmore and Kay between 1995 and 1997[4], but variants of this analysis were first published in Koenig (1999: Chapter 3) and Michaelis & Ruppenhofer (2001: Chapter 4). Parallel proposals have been made in TAG (Candito 1996; Clément & Kinyon 2003: 188; Kallmeyer & Osswald 2012: 171–172) and HPSG (Koenig 1999; Davis & Koenig 2000; Kordoni 2001).

Michaelis & Ruppenhofer (2001: 55–57) provide the following linking constructions:[5]

(9) a. *Transitive Construction*:

$$\begin{bmatrix} \textsc{syn} & \begin{bmatrix} \textsc{cat} & v \\ \textsc{voice} & \textit{active} \end{bmatrix} \\ \textsc{val} & \left\{ \begin{bmatrix} \textsc{role} & \begin{bmatrix} \textsc{gf} & \textit{obj} \\ \textsc{da} & - \end{bmatrix} \end{bmatrix} \right\} \end{bmatrix}$$

b. the *Subject Construction*:

$$\begin{bmatrix} \textsc{syn} & \begin{bmatrix} \textsc{cat} & v \end{bmatrix} \\ \textsc{val} & \left\{ \begin{bmatrix} \textsc{role} & \begin{bmatrix} \textsc{gf} & \textit{subj} \end{bmatrix} \end{bmatrix} \right\} \end{bmatrix}$$

[4] http://www.icsi.berkeley.edu/~kay/bcg/ConGram.html. 03.05.2010.

[5] In the original version of the transitive construction in (9a), there is a feature θ that has the value DA−, however, DA is a feature itself and − is the value. I have corrected this in (9a) accordingly.

In the following structures, GF stands for *grammatical function* and DA for *distinguished argument*. The distinguished argument usually corresponds to the subject in an active clause.

c. the *Passive Construction*:

$$\left[\begin{array}{ll} \textsc{syn} & \left[\begin{array}{ll} \textsc{cat} & v \\ \textsc{form} & \textit{PastPart} \end{array}\right] \\ \textsc{val} & \left\{ \left[\begin{array}{ll} \textsc{role} & \left[\begin{array}{ll} \textsc{gf} & \textit{obl} \\ \textsc{da} & + \end{array}\right] \\ \textsc{syn} & \text{P[von]}/\textit{zero} \end{array}\right] \right\} \end{array}\right]$$

The structure in (9a) says that the valence set of a linguistic object that is described by the transitive construction has to contain an element that has the grammatical function *object* and whose DA value is '−'. The DA value of the argument that would be the subject in an active clause is '+' and '−' for all other arguments. The subject construction states that an element of the valence set must have the grammatical function *subject.* In the passive construction, there has to be an element with the grammatical function *oblique* that also has the DA value '+'. In the passive construction the element with the DA value '+' is realized either as a *by*-PP or not at all (*zero*).

The interaction of the constructions in (9) will be explained on the basis of the verb *schlagen* 'to beat':

(10) Lexical entry for *schlag-* 'beat':

$$\left[\begin{array}{ll} \textsc{syn} & \left[\begin{array}{ll} \textsc{cat} & v \end{array}\right] \\ \textsc{val} & \left\{ \left[\begin{array}{ll} \textsc{role} & \left[\begin{array}{ll} \theta & \textit{agent} \\ \textsc{da} & + \end{array}\right] \end{array}\right], \left[\begin{array}{ll} \textsc{role} & \left[\begin{array}{ll} \theta & \textit{patient} \end{array}\right] \end{array}\right] \right\} \end{array}\right]$$

If we combine this lexical entry with the transitive and subject constructions, we arrive at (11a) following Fillmore, Kay, Michaelis, and Ruppenhofer, whereas combining it with the subject and passive construction yields (11b):[6]

(11) a. *schlag-* + Subject and Transitive Construction:

$$\left[\begin{array}{ll} \textsc{syn} & \left[\begin{array}{ll} \textsc{cat} & v \\ \textsc{voice} & \textit{active} \end{array}\right] \\ \textsc{val} & \left\{ \left[\begin{array}{ll} \textsc{role} & \left[\begin{array}{ll} \theta & \textit{agent} \\ \textsc{gf} & \textit{subj} \\ \textsc{da} & + \end{array}\right] \end{array}\right], \left[\begin{array}{ll} \textsc{role} & \left[\begin{array}{ll} \theta & \textit{patient} \\ \textsc{gf} & \textit{obj} \\ \textsc{da} & - \end{array}\right] \end{array}\right] \right\} \end{array}\right]$$

[6] This assumes a particular understanding of set unification. For criticism of this, see Section 10.6.1.

b. *schlag-* + Subject and Passive Construction:

$$\begin{bmatrix} \textsc{syn} & \begin{bmatrix}\textsc{cat} & v\\ \textsc{form} & \textit{PastPart}\end{bmatrix}\\ \textsc{val} & \left\{ \begin{bmatrix} \textsc{role} & \begin{bmatrix}\theta & \textit{agent}\\ \textsc{gf} & \textit{obl}\\ \textsc{da} & +\end{bmatrix}\\ \textsc{syn} & \text{P[von]}/\textit{zero}\end{bmatrix}, \begin{bmatrix}\textsc{role} & \begin{bmatrix}\theta & \textit{patient}\\ \textsc{gf} & \textit{subj}\end{bmatrix}\end{bmatrix}\right\}\end{bmatrix}$$

Using the entries in (11), it is possible to analyze the sentences in (12):

(12) a. Er schlägt den Weltmeister.
he beat the world.champion
'He is beating the world champion.'

b. Der Weltmeister wird (von ihm) geschlagen.
the world.champion is by him beaten
'The world champion is being beaten (by him).'

This analysis is formally inconsistent as set unification cannot be formalized in such a way that the aforementioned constructions can be unified (Müller 2006; Müller 2007b: Section 7.5.2, see also Section 10.6.1 below). It is, however, possible to fix this analysis by using the HPSG formalization of sets (Pollard & Sag 1987; Pollard & Moshier 1990). The Subject, Transitive and Passive Constructions must then be modified such that they can say something about what an element in VAL looks like, rather than specifying the VAL value of a singleton set.

(13) The *Subject Construction* with Pollard and Moschier's definition of sets:

$$\begin{bmatrix}\textsc{syn}|\textsc{cat}\ \text{v}\\ \textsc{val}\ \boxed{1}\end{bmatrix} \wedge \left\{ \begin{bmatrix}\textsc{role} & \begin{bmatrix}\textsc{gf}\ \textit{subj}\end{bmatrix}\end{bmatrix} \right\} \subset \boxed{1}$$

The restriction in (13) states that the valence set of a head has to contain an element that has the grammatical function *subj*. By these means, it is possible to suppress arguments (by specifying SYN as *zero*), but it is not possible to add any additional arguments to the fixed set of arguments of *schlagen* 'to beat'.[7] For the analysis of Middle Constructions such as (14), inheritance-based approaches do not work as there is no satisfactory way to add the reflexive pronoun to the valence set:[8]

[7] Rather than requiring that *schlagen* 'to beat' has exactly two arguments as in HPSG, one could also assume that the constraint on the main lexical item would be of the kind in (11a). One would then require that *schlagen* has at least the two members in its valence set. This would complicate everything considerably and furthermore it would not be clear that the subject referred to in (13) would be one of the arguments that are referred to in the description of the lexical item for *schlagen* in (11a).

[8] One technically possible solution would be the following: one could assume that verbs that occur in middle constructions always have a description of a reflexive pronoun in their valence set. The Transitive Construction would then have to specify the SYN value of the reflexive pronoun as *zero* so that the additional reflexive pronoun is not realized in the Transitive Construction. The middle construction would suppress

(14) Das Buch liest sich gut.
the book reads REFL good
'The book reads well / is easy to read.'

If we want to introduce additional arguments, we require auxiliary features. An analysis using auxiliary features has been suggested by Koenig (1999). Since there are many argument structure changing processes that interact in various ways and are linked to particular semantic side-effects, it is inevitable that one ends up assuming a large number of syntactic and semantic auxiliary features. The interaction between the various linking constructions becomes so complex that this analysis also becomes cognitively implausible and has to be viewed as technically unusable. For a more detailed discussion of this point, see Müller (2007b: Section 7.5.2).

The following empirical problem is much more serious: some processes like passivization, impersonalization and causativization can be applied in combination or even allow for multiple application, but if the grammatical function of a particular argument is determined once and for all by unification, additional unifications cannot change the initial assignment. We will first look at languages which allow for a combination of passivization and impersonalization, such as Lithuanian (Timberlake 1982: Section 5), Irish (Noonan 1994), and Turkish (Özkaragöz 1986; Knecht 1985: Section 2.3.3). I will use Özkaragöz's Turkish examples in (15) for illustration (1986: 77):

(15) a. Bu şato-da boğ-ul-un-ur.
this château-LOC strangle-PASS-PASS-AOR
'One is strangled (by one) in this château.'

b. Bu oda-da döv-ül-ün-ür.
this room-LOC hit-PASS-PASS-AOR
'One is beaten (by one) in this room.'

c. Harp-te vur-ul-un-ur.
war-LOC shoot-PASS-PASS-AOR
'One is shot (by one) in war.'

-In, *-n*, and *-Il* are allomorphs of the passive/impersonal morpheme.[9]

Approaches that assume that the personal passive is the unification of some general structure with some passive-specific structure will not be able to capture double passivization or passivization + impersonalization since they have committed themselves to a certain structure too early. The problem for nontransformational approaches that state syntactic structure for the passive is that such a structure, once stated, cannot be

the subject, but realizes the object and the reflexive.

This solution cannot be applied to the recursive processes we will encounter in a moment such as causativization in Turkish, unless one wishes to assume infinite valence sets.

[9] According to Özkaragöz, the data is best captured by an analysis that assumes that the passive applies to a passivized transitive verb and hence results in an impersonal passive. The cited authors discussed their data as instances of double passivization, but it was argued by Blevins (2003) that these and similar examples from other languages are impersonal constructions that can be combined with personal passives.

modified. That is, we said that the underlying object is the subject in the passive sentence. But in order to get the double passivization/passivization + impersonalization, we have to suppress this argument as well. What is needed is some sort of process (or description) that takes a representation and relates it to another representation with a suppressed subject. This representation is related to a third representation which again suppresses the subject resulting in an impersonal sentence. In order to do this one needs different strata as in Relational Grammar (Timberlake 1982; Özkaragöz 1986), metarules (Gazdar, Klein, Pullum & Sag 1985), lexical rules (Dowty, 1978: 412; 2003: Section 3.4; Bresnan 1982c; Pollard & Sag 1987; Blevins 2003; Müller 2003b), transformations (Chomsky 1957), or just a morpheme-based morphological analysis that results in items with different valence properties when the passivization morpheme is combined with a head (Chomsky 1981a).

The second set of problematic data that will be discussed comes from causativization in Turkish (Lewis 1967):

(16) Öl-dür-t-tür-t-
'to cause someone to cause someone to kill someone'
(kill = cause someone to die)

The causative morpheme *-t* is combined three times with the verb (*tür* is an allomorph of the causative morpheme). This argument structure-changing process cannot be modeled in an inheritance hierarchy since if we were to say that a word can inherit from the causative construction three times, we would still not have anything different to what we would have if the inheritance via the causative construction had applied only once. For this kind of phenomenon, we would require rules that relate a linguistic object to another, more complex object, that is, lexical rules (unary branching rules which change the phonology of a linguistic sign) or binary rules that combine a particular sign with a derivational morpheme. These rules can semantically embed the original sign (that is, add *cause* to *kill*).

The problem of repeated combination with causativization affixes is an instance of a more general problem: derivational morphology cannot be handled by inheritance as was already pointed out by Krieger & Nerbonne (1993) with respect to cases like *preprepreversion*.

If we assume that changes in valence such as passive, causativization and middles should be described with the same means across languages, then evidence from Lithuanian and Turkish form an argument against inheritance-based analyses of the passive (Müller 2006, 2007c; Müller & Wechsler 2014a). See also Section 21.2.2 for the discussion of an inheritance-based approach to passive in LFG and Section 21.4.2 for the discussion of an inheritance-based approach in Simpler Syntax.

10.3 Verb position

At present, I only know of one article in the framework of CxG that has dealt with the sentence structure in German. This is the article by Vanessa Micelli (2012), where she describes a computer implementation of a German grammar in Fluid Construction Grammar. This fragment is restricted to declarative V2-clauses and *wh*-questions. In her analysis, the middle field forms a constituent comprising exactly two constituents (the direct and indirect object).[10] The right sentence bracket and the postfield are empty. Long-distance dependencies are not discussed. It is only possible for arguments of the verb in the left sentence bracket to occur in the prefield. Micelli's work is an interesting starting point but one has to wait and see how the analysis will be modified when the grammar fragment is expanded.

In the following, I will not discuss Micelli's analysis further, but instead explore some of the possibilities for analyzing German sentence structure that are at least possible in principle in a CxG framework. Since there are neither empty elements nor transformations, the GB and HPSG analyses as well as their variants in Categorial Grammar are ruled out. The following options remain:

- an analysis similar to LFG with an optional verb
- an entirely flat analysis as proposed in GPSG
- an analysis with binary branching but variable verb position like that of Steedman (2000: 159)

Different variants of CxG make different assumptions about how abstract constructions can be. In Categorial Grammar, we have very general combinatorial rules which combine possibly complex signs without adding any meaning of their own (see rule (2) on page 244 for example). (17) shows an example in which the abstract rule of forward application was used:

(17) [[[[Gibt] der Mann] der Frau] das Buch]
give the man the woman the book
'Does the man give the woman the book?'

If we do not want these kinds of abstract combinatorial rules, then this analysis must be excluded.

The LFG analysis in Section 7.3 is probably also unacceptable on a CxG view as it is assumed in this analysis that *der Mann der Frau das Buch* forms a VP although only three NPs have been combined. CxG has nothing like the theory of extended head domains that was presented in Section 7.3.

Thus, both variants with binary-branching structures are ruled out and only the analysis with flat branching structures remains. Sign-based CxG, which is a variant of HPSG (Sag 2010: 486), as well as Embodied Construction Grammar (Bergen & Chang 2005: 156) allow for a separation of immediate dominance and linear order so that it would be possible to formulate a construction which would correspond to the dominance rule in (18) for transitive verbs:[11]

[10] Note that none of the constituent tests that were discussed in Section 1.3 justifies such an analysis and that no other theory in this book assumes the *Mittelfeld* to be a constituent.

[11] In principle, this is also Micelli's analysis, but she assumed that the middle field forms a separate constituent.

(18) S → V, NP, NP

Here, we have the problem that adjuncts in German can occur between any of the arguments. In GPSG, adjuncts are introduced by metarules. In formal variants of CxG, lexical rules, but not metarules, are used.[12] If one does not wish to expand the formalism to include metarules, then there are three options remaining:

- Adjuncts are introduced in the lexicon (van Noord & Bouma 1994; Bouma, Malouf & Sag 2001a) and treated as arguments in the syntax,
- Constructions always have slots available for an arbitrary number of adjuncts, or
- Constructions can be discontinuous

Kasper (1994) has proposed an analysis of the first type in HPSG: adjuncts and arguments are combined with the head in a flat structure. This corresponds to the dominance rule in (19), where the position of adjuncts is not stated by the dominance rule.

(19) S → V, NP, NP, Adj*

If we want to say something about the meaning of the entire construction, then one has to combine the original construction (transitive, in the above example) with the semantics contributed by each of the adjuncts. These computations are not trivial and require relational constraints (small computer programs), which should be avoided if there are conceptually simpler solutions for describing a particular phenomenon.

The alternative would be to use discontinuous constructions. Analyses with discontinuous constituents have been proposed in both HPSG (Reape 1994) and Embodied Construction Grammar (Bergen & Chang 2005). If we apply Bergen and Chang's analysis to German, the italicized words in (20) would be part of a ditransitive construction.

(20) *Gibt* *der Mann* morgen *der Frau* unter der Brücke *das Geld*?
gives the man tomorrow the woman under the bridge the money
'Is the man going to give the woman the money under the bridge tomorrow?'

The construction has been realized discontinuously and the adjuncts are inserted into the gaps. In this kind of approach, one still has to explain how the scope of quantifiers and adjuncts is determined. While this may be possible, the solution is not obvious and has not been worked out in any of the CxG approaches to date. For further discussions of approaches that allow for discontinuous constituents see Section 11.7.2.2.

[12] Goldberg (2014: 116) mentions metarule-like devices and refers to Cappelle (2006). The difference between metarules and their CxG variant as envisioned by Cappelle and Goldberg is that in CxG two constructions are related without one construction being basic and the other one derived. Rather there exists a mutual relation between two constructions.

10.4 Local reordering

If we assume flat branching structures, then it is possible to use the GPSG analysis for the order of arguments. However, Kay (2002) assumes a phrasal construction for so-called Heavy-NP-Shift in English, which means that there is a new rule for the reordering of heavy NPs in English rather than one rule and two different ways to linearize the daughters.

In CxG, it is often argued that the usage contexts of certain orders differ and we therefore must be dealing with a different construction. Accordingly, one would have to assume six constructions for sentences with ditransitive verbs in final position (also see page 185). An alternative would be to assume that the ordering variants all have a similar structure and that the information-structural properties are dependent on the position of constituents in the respective structure (see De Kuthy (2000) for German and Bildhauer (2008) for Spanish).

10.5 Long-distance dependencies

Kay & Fillmore (1999: Section 3.10) discuss long-distance dependencies in their article. Since the number of arguments is not specified in the verb phrase construction, it is possible that an argument of the verb is not locally present. Like the LFG and GPSG analyses in previous chapters, there are no empty elements assumed for the analysis of long-distance dependencies. In the *Left Isolation Construction* that licenses the entire sentence, there is a left daughter and a right daughter. The left daughter corresponds to whatever was extracted from the right daughter. The connection between the fronted element and the position where it is missing is achieved by the operator VAL. VAL provides all elements of the valence set of a linguistic object as well as all elements in the valence set of these elements and so on. It is thereby possible to have unrestricted access to an argument or adjunct daughter of any depth of embedding, and then identify the fronted constituent with an open valence slot.[13] This approach corresponds to the LFG analysis of Kaplan & Zaenen (1989) based on functional uncertainty.

10.6 New developments and theoretical variants

Berkeley Construction Grammar was already discussed in the main part of this chapter. The discussion of the formal underpinnings was deferred until the theoretical variants section, since it is more advanced. I made some comments on set unification in Müller (2006: 858), but the longer discussion is only available in Müller (2007b: Section 7.5.2), which is in German. Therefore, I include Section 10.6.1 here, which discusses the formal

[13] Note again, that there are problems with the formalization of this proposal in Kay & Fillmore's paper. The formalization of VAL, which was provided by Andreas Kathol, seems to presuppose a formalization of sets as the one that is used in HPSG, but the rest of Fillmore & Kay's paper assumes a different formalization, which is inconsistent. See Section 10.6.1.

underpinnings of Berkeley Construction Grammar in more detail and shows that they are not suited for what they were intended to do.

Section 10.6.2 discusses Sign-Based Construction Grammar, which was developed in joint work by Charles Fillmore, Paul Kay and Ivan Sag. It embodies ideas from BCG without having its formal flaws. Section 10.6.3 deals with Embodied Construction Grammar, which is based on work by Charles Fillmore, Paul Kay and George Lakoff. Section 10.6.4 deals with Fluid Construction Grammar.

10.6.1 Berkeley Construction Grammar

Section 10.2 discussed the valence representation in BCG and linking constructions for active and passive. Kay & Fillmore (1999) represent valence information in sets and I deferred the discussion of the formal properties of sets in BCG to this section. Fillmore and Kay's assumptions regarding set unification differ fundamentally from those that are made in HPSG. Kay and Fillmore assume that the unification of the set { a } with the set { b }, where a and b do not unify, results in the union of the two sets, that is { a, b }. Due to this special understanding of sets it is possible to increase the number of elements in a set by means of unification. The unification of two sets that contain compatible elements is the disjunction of sets that contain the respective unifications of the compatible elements. This sounds complicated, but we are only interested in a specific case: the unification of an arbitrary set with a singleton set:

(21) { NP[*nom*], NP[*acc*] } ∧ { NP[*nom*] } = { NP[*nom*], NP[*acc*] }

According to Fillmore & Kay the unification of a set with another set that contains a compatible element does not result in an increase of the number of list elements.

(22) illustrates another possible case:

(22) { NP, NP[*acc*] } ∧ { NP[*nom*] } = { NP[*nom*], NP[*acc*] }

The first NP in (22) is underspecified with respect to its case. The case of the NP in the second set is specified as nominative. NP[*nom*] does not unify with NP[*acc*] but with NP.

This particular conception of unification has consequences. Unification is usually defined as follows:

(23) The unification of two structures FS_1 and FS_2 is the structure FS_3, that is subsumed by both FS_1 and FS_2 where there is no other structure that subsumes FS_1 and FS_2 and is subsumed by FS_3.

A structure FS_1 is said to subsume FS_3, iff FS_3 contains all feature value pairs and structure sharings from FS_1. FS_3 may contain additional feature value pairs or structure sharings. The consequence is that the subsumption relations in (24b,c) have to hold, if unification of valence sets works as in (24a):

(24) Properties of the set unification according to Kay & Fillmore (1999):
a. { NP[*nom*] } ∧ { NP[*acc*] } = { NP[*nom*], NP[*acc*] }
b. { NP[*nom*] } ⪰ { NP[*nom*], NP[*acc*] }
c. { NP[*acc*] } ⪰ { NP[*nom*], NP[*acc*] }

(24b) means that a feature structure with a valence set that contains just one NP[*nom*] is more general than a feature structure that contains both an NP[*nom*] and an NP[*acc*]. Therefore the set of transitive verbs is a subset of the intransitive verbs. This is rather unintuitive, but compatible with Fillmore & Kay's system for the licensing of arguments. However, there are problems with the interaction of valence specifications and linking constructions, which we turn to now.

We have seen the result of combining lexical items with linking constructions in (11a) and (11b), but the question of how these results are derived has not been addressed so far. Kay (2002) suggests an automatic computation of all compatible combinations of maximally specific constructions. Such a procedure could be used to compute the lexical representations we saw in Section 10.2 and these could be then used to analyze the well-formed sentences in (12).

However, problems would result for ungrammatical sentences like (25b). *grauen* 'to dread' is a subjectless verb. If one would simply combine all compatible linking constructions with *grauen*, the Kay & Fillmoreian conception of set unification would cause the introduction of a subject into the valence set of *grauen*. (25b) would be licensed by the grammar:

(25) a. Dem Student graut vor der Prüfung.
the.DAT student dreads before the exam
'The student dreads the exam.'

b. * Ich graue dem Student vor der Prüfung.
I dread the.DAT student before the exam

One could solve this problem by specifying an element with the grammatical function *subject* in the lexical entry of *grauen* 'to dread'. In addition, it would have to be stipulated that this subject can only be realized as an overt or covert expletive (The covert expletive would be SYN *zero*). For the covert expletive, this means it has neither a form nor a meaning. Such expletive pronouns without phonological realization are usually frowned upon in Construction Grammar and analyses that can do without such abstract entities are to be preferred.

Kay & Fillmore (1999) represent the semantic contribution of signs as sets as well. This excludes the possibility of preventing the unwanted unification of linking constructions by referring to semantic constraints since we have the same effect as we have with valence sets: if the semantic descriptions are incompatible, the set is extended. This means that in an automatic unification computation all verbs are compatible with the Transitive Construction in (9a) and this would license analyses for (26) in addition to those of (25b).

(26) a. * Der Mann schläft das Buch.
the man sleeps the book

b. * Der Mann denkt an die Frau das Buch.
the man thinks at the woman the book

An intransitive verb was unified with the Transitive Construction in the analysis of (26a) and in (26b) a verb that takes a prepositional object was combined with the Transitive Construction. This means that representations like (11) cannot be computed automatically as was intended by Kay (2002). Therefore one would have to specify subconstructions for all argument structure possibilities for every verb (active, passive, middle, …). This does not capture the fact that speakers can form passives after acquiring new verbs without having to learn about the fact that the newly learned verb forms one.

Michaelis & Ruppenhofer (2001) do not use sets for the representation of semantic information. Therefore they could use constraints regarding the meaning of verbs in the Transitive Construction. To this end, one needs to represent semantic relations with feature descriptions as it was done in Section 9.1.6. Adopting such a representation, it is possible to talk about two-place relations in an abstract way. See for instance the discussion of (26) on page 277. However, the unification with the Subject Construction cannot be blocked with reference to semantics since there exist so-called raising verbs that take a subject without assigning a semantic role to it. As is evidenced by subject verb agreement, *du* 'you' is the subject in (27), but the subject does not get a semantic role. The referent of *du* is not the one who *seems*.

(27) Du scheinst gleich einzuschlafen.
you seem.2SG soon in.to.sleep
'You seem like you will fall asleep soon.'

This means that one is forced to either assume an empty expletive subject for verbs like *grauen* or to specify explicitly which verbs may inherit from the subject construction and which may not.

In addition to (27), there exist object raising constructions with accusative objects that can be promoted to subject in passives. The subject in the passive construction does not get a semantic role from the finite verb:

(28) a. Richard lacht ihn an.
Richard laughs him towards
'Richard smiles at him.'

b. Richard fischt den Teich leer.
Richard fishes the pond empty

The objects in (28) are semantic arguments of *an* 'towards' and *leer* 'empty', respectively, but not semantic arguments of the verbs *lacht* 'laughs' and *fischt* 'fishes', respectively. If one wants to explain these active forms and the corresponding passive forms via the linking constructions in (9), one cannot refer to semantic properties of the verb. Therefore, one is forced to postulate specific lexical entries for all possible verb forms in active and passive sentences.

10.6.2 Sign-Based Construction Grammar

In more recent work by Fillmore, Kay, Michaelis and Sag, the Kay & Fillmore formalization of the description of valence using the Kay & Fillmore version of sets was abandoned in favor of the HPSG formalization (Kay 2005; Michaelis 2006; Sag 2012; Sag, Boas & Kay 2012: 10–11). Sign-Based Construction Grammar was developed from the Berkeley variant of CxG. Sign-Based Construction Grammar is a variant of HPSG (Sag 2010: 486) and as such uses the formal apparatus of HPSG (typed feature structures). Valence and saturation are treated in exactly the same way as in standard HPSG. Changes in valence are also analyzed as in HPSG using lexical rules (Sag, Boas & Kay 2012: Section 2.3). The analysis of long-distance dependencies was adopted from HPSG (or rather GPSG). Minimal Recursion Semantics (MRS; Copestake, Flickinger, Pollard & Sag 2005) is used for the description of semantic content. The only difference to works in standard HPSG is the organization of the features in feature structures. A new feature geometry was introduced to rule out constructions that describe daughters of daughters and therefore have a much larger locality domain in contrast to rules in phrase structure grammars, LFG, and GPSG. I do not view this new feature geometry as particularly sensible as it can be easily circumvented and serves to complicate the theory. This will be discussed in Section 10.6.2.1. Other changes regard the omission of the LOCAL feature and the omission of valence features. These changes are discussed in Section 10.6.2.2 and 10.6.2.4, respectively.

10.6.2.1 Locality and MOTHER

Sag, Wasow & Bender (2003: 475–489) and Sag (2007, 2012) suggest using a MOTHER feature in addition to daughter features. The Head-Complement Construction would then have the form in (29):

(29) Head-Complement Construction following Sag, Wasow & Bender (2003: 481):

$$\textit{head-comp-cx} \rightarrow \begin{bmatrix} \text{MOTHER}|\text{SYN}|\text{VAL}|\text{COMPS}\ \langle\rangle \\ \text{HEAD-DTR}\ \boxed{0} \begin{bmatrix} \text{SYN}|\text{VAL}|\text{COMPS}\ \boxed{A} \\ \textit{word} \end{bmatrix} \\ \text{DTRS}\ \langle\ \boxed{0}\ \rangle \oplus \boxed{A}\ \textit{nelist} \end{bmatrix}$$

The value of COMPS is then a list of the complements of a head (see Section 9.6.1). Unlike in standard HPSG, it is not *synsem* objects that are selected with valence lists, but rather signs. The analysis of the phrase *ate a pizza* takes the form in (30).[14]

[14] SBCG uses a FORM feature in addition to the PHON feature, which is used for phonological information as in earlier versions of HPSG (Sag 2012: Section 3.1, Section 3.6). The FORM feature is usually provided in example analyses.

(30)
$$
\begin{bmatrix}
\text{MOTHER} & \begin{bmatrix} \text{FORM} & \langle \textit{ate, a, pizza} \rangle \\ \text{SYN} & \begin{bmatrix} \text{HEAD} & \textit{verb} \\ \text{SPR} & \langle \text{NP}[\textit{nom}] \rangle \\ \text{COMPS} & \langle\rangle \end{bmatrix} \\ \text{SEM} & \ldots \\ \textit{phrase} & \end{bmatrix} \\
\text{HEAD-DTR} \; \boxed{1} & \begin{bmatrix} \text{FORM} & \langle \textit{ate} \rangle \\ \text{SYN} & \begin{bmatrix} \text{HEAD} & \textit{verb} \\ \text{SPR} & \langle \text{NP}[\textit{nom}] \rangle \\ \text{COMPS} & \langle \boxed{2}\, \text{NP}[\textit{acc}] \rangle \end{bmatrix} \\ \text{SEM} & \ldots \\ \textit{word} & \end{bmatrix} \\
\text{DTRS} & \langle \boxed{1}, \boxed{2} \rangle \\
\textit{head-comp-cx} &
\end{bmatrix}
$$

The difference between HPSG in the version of Pollard & Sag (1994) is that for Sag, Wasow & Bender, signs do not have daughters and this makes the selection of daughters impossible. As a result, the SYNSEM feature becomes superfluous (selection of the PHON value and of the value of the newly introduced FORM feature is allowed in Sag, Wasow & Bender (2003) and Sag (2012)). The information about the linguistic objects that contribute to a complex sign is only represented in the very outside of the structure. The sign represented under MOTHER is of the type *phrase* but does not contain any information about the daughters. The object described in (30) is of course also of another type than the phrasal or lexical signs that can occur as its daughters. We therefore need the following extension so that the grammar will work (Sag, Wasow & Bender 2003: 478):[15]

(31) Φ is a Well-Formed Structure according to a grammar G if and only if:
  1. there is a construction C in G, and
  2. there is a feature structure I that is an instantiation of C, such that Φ is the value of the MOTHER feature of I.

For comparison, a description is given here with the feature geometry that was assumed in Section 9.6.1:

[15] A less formal version of this constraint is given as the Sign Principle by Sag (2012: 105): "Every sign must be listemically or constructionally licensed, where: a sign is listemically licensed only if it satisfies some listeme, and a sign is constructionally licensed if it is the mother of some well-formed construct."

(32)
$$\begin{bmatrix} \mathrm{PHON}\ \langle \textit{ate, a, pizza} \rangle \\ \mathrm{SYNSEM|LOC} \begin{bmatrix} \mathrm{CAT} & \begin{bmatrix} \mathrm{HEAD} & \textit{verb} \\ \mathrm{SPR} & \langle \mathrm{NP}[\textit{nom}] \rangle \\ \mathrm{COMPS} & \langle \rangle \end{bmatrix} \\ \mathrm{CONT} & \ldots \end{bmatrix} \\ \mathrm{HEAD\text{-}DTR} \begin{bmatrix} \mathrm{PHON}\ \langle \textit{ate} \rangle \\ \mathrm{SYNSEM|LOC} \begin{bmatrix} \mathrm{CAT} & \begin{bmatrix} \mathrm{HEAD} & \textit{verb} \\ \mathrm{SPR} & \langle \mathrm{NP}[\textit{nom}] \rangle \\ \mathrm{COMPS} & \langle \boxed{2}\ \mathrm{NP}[\textit{acc}] \rangle \end{bmatrix} \\ \mathrm{CONT} & \ldots \end{bmatrix} \\ \textit{word} \end{bmatrix} \\ \mathrm{NON\text{-}HEAD\text{-}DTRS}\ \langle [\ \mathrm{SYNSEM}\ \boxed{2}\ ] \rangle \\ \textit{head-complement-phrase} \end{bmatrix}$$

In (32), the features HEAD-DTR and NON-HEAD-DTRS belong to those features that phrases of type *head-complement-phrase* have. In (30), however, the phrase corresponds only to the value of the MOTHER feature and therefore has no daughters represented in the sign itself. Using the feature geometry in (32), it is in principle possible to formulate restrictions on the daughters of the object in the NON-HEAD-DTRS list, which would be completely ruled out under the assumption of the feature geometry in (30) and the restriction in (31).

There are several arguments against this feature geometry, which will be discussed in the following subsections. The first one is an empirical one: there may be idioms that span clauses. The second argument concerns the status of the meta statement in (31) and the third one computational complexity.

10.6.2.1.1 Idioms that cross constituent boundaries

In Müller (2007b: Chapter 12) I conjectured that the locality restrictions may be too strong since there may be idioms that require one to make reference to daughters of daughters for their description. Richter & Sailer (2009) discuss the following idioms as examples:

(33) a. nicht wissen, wo X_Dat der Kopf steht
not know where X the head stands
'to not know where x's head is at'

b. glauben, X_Acc tritt ein Pferd
believe X kicks a horse
'be utterly surprised'

c. aussehen, als hätten X_Dat die Hühner das Brot weggefressen
look as.if had X the chicken the bread away.eaten
'to look confused/puzzled'

d. look as if butter wouldn't melt [in X's mouth]
'to look absolutely innocent'

In sentences containing the idioms in (33a–c), the X-constituent has to be a pronoun that refers to the subject of the matrix clause. If this is not the case, the sentences become ungrammatical or lose their idiomatic meaning.

(34) a. Ich glaube, mich / # dich tritt ein Pferd.
I believe me.ACC you.ACC kicks a horse

b. Jonas glaubt, ihn tritt ein Pferd.[16]
Jonas believes him kicks a horse
'Jonas is utterly surprised.'

c. # Jonas glaubt, dich tritt ein Pferd.
Jonas believes you kicks a horse
'Jonas believes that a horse kicks you.'

In order to enforce this co-reference, a restriction has to be able to refer to both the subject of *glauben* 'believe' and the object of *treten* 'kick' at the same time. In SBCG, there is the possibility of referring to the subject since the relevant information is also available on maximal projections (the value of a special feature (XARG) is identical to the subject of a head). In (33a–c), we are dealing with accusative and dative objects. Instead of only making information about one argument accessible, one could represent the complete argument structure on the maximal projection (as is done in some versions of HPSG, see page 304 and pages 553–554). This would remove locality of selection, however, since if all heads project their argument structure, then it is possible to determine the properties of arguments of arguments by looking at the elements present in the argument structure. Thus, the argument structure of *wissen* 'to know' in (35) would contain the description of a *dass* clause.

(35) Peter weiß, dass Klaus kommt.
Peter knows that Klaus comes
'Peter knows that Klaus is coming.'

Since the description of the *dass* clause contains the argument structure of *dass*, it is possible to access the argument of *dass*. *wissen* 'to know' can therefore access *Klaus kommt*. As such, *wissen* also has access to the argument structure of *kommt* 'to come', which is why *Klaus* is also accessible to *wissen*. However, the purpose of the new, more restrictive feature geometry was to rule out such nonlocal access to arguments.

An alternative to projecting the complete argument structure was suggested by Kay et al. (2015: Section 6): instead of assuming that the subject is the XARG in idiomatic constructions like those in (33), they assume that the accusative or dative argument is the XARG. This is an interesting proposal that could be used to fix the cases under discussion,

[16] http://www.machandel-verlag.de/der-katzenschatz.html, 06.07.2015.

but the question is whether it scales up if interaction with other phenomena are considered. For instance, Bender & Flickinger (1999) use XARG in their account of question tags in English. So, if English idioms can be found that require a non-subject XARG in embedded sentences while also admitting the idiom parts in the embedded sentence to occur as full clause with question tag, we would have conflicting demands and would have to assume different XARGs for root and embedded clauses, which would make this version of the lexical theory rather unattractive, since we would need two lexical items for the respective verb.

(33d) is especially interesting, since here the X that refers to material outside the idiom is in an adjunct. If such cases existed, the XARG mechanism would be clearly insufficient since XARG is not projected from adjuncts. However, as Kay et al. (2015) point out the X does not necessarily have to be a pronoun that is coreferent with an element in a matrix clause. They provide the following example:

(36) Justin Bieber—Once upon a time ∅ butter wouldn't melt in little Justin's mouth. Now internationally famous for being a weapons-grade petulant brat …

So, whether examples of the respective kind can be found is an open question.

Returning to our *horse* examples, Richter & Sailer (2009: 313) argue that the idiomatic reading is only available if the accusative pronouns is fronted and the embedded clause is V2. The examples in (37) do not have the idiomatic reading:

(37) a. Ich glaube, dass mich ein Pferd tritt.
I believe that me a horse kicks
'I believe that a horse kicks me.'

b. Ich glaube, ein Pferd tritt mich.
I believe a horse kicks me
'I believe that a horse kicks me.'

Richter & Sailer assume a structure for *X_Acc tritt ein Pferd* in (33b) that contains, among others, the constraints in (38).

(38)
$$
\begin{bmatrix}
\text{SYNSEM|LOC} & \begin{bmatrix} \text{CAT|LISTEME} & \textit{very-surprised} \\ \text{CONT|MAIN} & \textit{surprised}'(x_{\boxed{2}}) \end{bmatrix} \\
\text{DTRS} & \begin{bmatrix} \text{FILLER-DTR} & \begin{bmatrix}\text{SYNSEM|LOC} & \boxed{1} \\ \textit{word} & \end{bmatrix} \\
\text{H-DTR} & \begin{bmatrix} \text{LF|EXC} & \text{‘a horse kicks } x_{\boxed{2}}\text{’} \\ \text{(DTRS|H-DTR)}^{+} & \begin{bmatrix} \text{SYNSEM|LOC|CAT} & \begin{bmatrix}\text{HEAD} & \begin{bmatrix}\text{TENSE} & \textit{pres}\end{bmatrix} \\ \text{LISTEME} & \textit{treten}\end{bmatrix} \\ \text{ARG-ST} & \left\langle \text{NP[LISTEME } \textit{pferd}\text{, DEF } -, \textit{sg}], \begin{bmatrix}\text{LOC } \boxed{1} & \begin{bmatrix}\text{CAT|HEAD|CASE } \textit{acc} \\ \text{CONT} \begin{bmatrix}\text{INDEX } \boxed{2} \\ \textit{ppro}\end{bmatrix}\end{bmatrix}\end{bmatrix} \right\rangle \\ \textit{word} & \end{bmatrix}\end{bmatrix}\end{bmatrix} \\
\textit{phrase} &
\end{bmatrix}
$$

The feature geometry in (38) differs somewhat from what was presented in Chapter 9 but that is not of interest here. It is only of importance that the semantic contribution of the entire phrase is *surprised′*($x_{\boxed{2}}$). The following is said about the internal structure of the phrase: it consists of a filler-daughter (an extracted element) and also of a head daughter corresponding to a sentence from which something has been extracted. The head daughter means ‘a horse kicks $x_{\boxed{2}}$’ and has an internal head somewhere whose argument structure list contains an indefinite NP with the word *Pferd* ‘horse’ as its head. The second element in the argument structure is a pronominal NP in the accusative whose LOCAL value is identical to that of the filler ($\boxed{1}$). The entire meaning of this part of the sentence is *surprised′*($x_{\boxed{2}}$), whereby $\boxed{2}$ is identical to the referential index of the pronoun. In addition to the constraints in (38), there are additional ones that ensure that the partial clause appears with the relevant form of *glauben* ‘to believe’ or *denken* ‘to think’. The exact details are not that important here. What is important is that one can specify constraints on complex syntactic elements, that is, it must be possible to refer to daughters of daughters. This is possible with the classical HPSG feature geometry, but not with the feature geometry of SBCG. For a more general discussion of locality, see Section 18.2.

The restrictions on *Pferd* clauses in (38) are too strict, however, since there are variants of the idiom that do not have the accusative pronoun in the *Vorfeld*:

(39) a. ich glaub es tritt mich ein Pferd wenn ich einen derartigen Unsinn lese.[17]
I believe EXPL kicks me a horse when I a such nonsense read
‘I am utterly surprised when I read such nonsense.’

[17] http://www.welt.de/wirtschaft/article116297208/Die-verlogene-Kritik-an-den-Steuerparadiesen.html, 10.12.2015.

b. omg dieser xBluuR der nn ist wieder da ey nein ich glaub es tritt
   omg this XBluuR he is again there no I believe EXPL kicks
   mich ein Pferd!![18]
   me a horse

c. ich glaub jetzt tritt mich ein pferd[19]
   I believe now kicks me a horse

   'I am utterly surprised now.'

In (40a–b) the *Vorfeld* is filled by an expletive and in (40c) an adverb fills the *Vorfeld* position. While these forms of the idiom are really rare, they do exist and should be allowed for by the description of the idiom. So, one would have to make sure that *ein Pferd* 'a horse' is not fronted, but this can be done in the lexical item of *tritt* 'kicks'. This shows that the cases at hand cannot be used to argue for models that allow for the representation of (underspecified) trees of depth greater one, but I still believe that such idioms can be found. Of course this is an open empirical question.

What is not an open empirical question though is whether humans store chunks with complex internal structure or not. It is clear that we do and much Construction Grammar literature emphasizes this. Constructional HPSG can represent such chunks, but SBCG cannot since linguistic signs do not have daughters. So here Constructional HPSG and TAG are the theories that can represent complex chunks of linguistic material with its internal structure, while other theories like GB, Minimalism, CG, LFG and DG can not.

10.6.2.1.2 Complicated licensing of constructions

In addition to these empirical problems, there is a conceptual problem with (31): (31) is not part of the formalism of typed feature structures but rather a meta-statement. Therefore, grammars which use (31) cannot be described with the normal formalism. The formalization given in Richter (2004) cannot be directly applied to SBCG, which means that the formal foundations of SBCG still have to be worked out.[20] Furthermore, the original problem that (31) was designed to solve is not solved by introducing the new feature geometry and the meta statement. Instead, the problem is just moved to another level since we now need a theory about what is a permissible meta-statement and what is not. As such, a grammarian could add a further clause to the meta statement stating that Φ is only a well-formed structure if it is true of the daughters of a relevant construction C that they are the MOTHER value of a construction C′. It would be possible to formulate constraints in the meta-statement about the construction C′ or individual values inside the corresponding feature structures. In this way, locality would have been abandoned

[18] http://forum.gta-life.de/index.php?user/3501-malcolm/, 10.12.2015.

[19] http://www.castingshow-news.de/menowin-frhlich-soll-er-zum-islam-konvertieren-7228/, 10.12.2015.

[20] A note of caution is necessary since there were misunderstandings in the past regarding the degree of formalizations of SBCG: in comparison to most other theories discussed in this book, SBCG is well-formalized. For instance it is easy to come up with a computer implementation of SBCG fragments. I implemented one in the TRALE system myself. The reader is referred to Richter (2004) to get an idea what kind of deeper formalization is talked about here.

since it is possible to refer to daughters of daughters. By assuming (31), the theoretical inventory has been increased without any explanatory gain.

10.6.2.1.3 Computational complexity

One motivation behind restrictions on locality was to reduce the computational complexity of the formalism (Ivan Sag, p. c. 2011, See Chapter 17 on computational complexity and generative power). However, the locality restrictions of SBCG can be circumvented easily by structure sharing (Müller 2013b: Section 9.6.1). To see this consider a construction with the following form:

(40)
$$\left[\begin{array}{ll} \text{MOTHER} & \left[\begin{array}{ll} \text{PHON} & \textit{phonological-object} \\ \text{FORM} & \textit{morphological-object} \\ \text{SYN} & \textit{syntactic information} \\ \text{SEM} & \textit{semantic information} \\ \text{NASTY} & \boxed{1} \\ \multicolumn{2}{l}{\textit{sign}} \end{array}\right] \\ \text{DTRS} & \boxed{1}\ \textit{list of signs} \end{array}\right]$$

The feature NASTY in the MOTHER sign refers to the value of DTRS and hence all the internal structure of the sign that is licensed by the constructional schema in (40) is available. Of course one could rule out such things by stipulation – if one considered it to be empirically adequate, but then one could as well continue to use the feature geometry of Constructional HPSG (Sag 1997) and stipulate constraints like "Do not look into the daughters." An example of such a constraint given in prose is the Locality Principle of Pollard & Sag (1987: 143–144).

10.6.2.2 Lexical extraction and the LOCAL feature

A main theme in Ivan Sag's work was to eliminate empty elements from grammar. He co-developed a lexical analysis of extraction that did not use empty elements (van Noord & Bouma 1994; Bouma, Malouf & Sag 2001a). Rather than assuming a trace as in earlier versions of HPSG (see Section 9.5) a lexical rule (lexical construction) or a special mapping between list-valued features (for instance ARG-ST and VALENCE) is assumed so that a lexical item is licensed that has an element in the GAP list (SLASH in earlier versions of HPSG). The LOCAL feature in Standard HPSG is used to bundle all those information that is shared between a filler and the gap. The lexical entry for the trace was given as (53) on page 297 and is repeated here for convenience:

(41) Extraction trace:

$$\left[\begin{array}{ll}\textsc{phon} & \langle\rangle \\ \textsc{synsem} & \left[\begin{array}{ll}\textsc{loc} & \boxed{1} \\ \textsc{nonloc} & \left[\begin{array}{ll}\textsc{inher}|\textsc{slash} & \langle\, \boxed{1}\, \rangle \\ \textsc{to-bind}|\textsc{slash} & \langle\rangle\end{array}\right]\end{array}\right] \\ \textit{word} & \end{array}\right]$$

For the trace-based analysis of extraction it is crucial that only information under CATEGORY and CONT is shared between filler and gap. This information is bundled under LOCAL. Information about daughters and phonology of the filler is not shared. Since traces are not pronounced, their PHONOLOGY value would be incompatible with any PHONOLOGY value of a filler.

In a lexical approach, one would assume an additional lexical item with an element in SLASH or GAP.

(42) Lexical introduction of nonlocal dependencies in SBCG according to Sag (2012: 163):

$$\left[\begin{array}{ll}\textsc{form} & \langle\, \textit{like}\, \rangle \\ \textsc{arg-st} & \left\langle \boxed{1} \left[\begin{array}{ll}\text{NP} & \\ \textsc{gap} & \langle\rangle\end{array}\right], \boxed{2} \left[\begin{array}{ll}\text{NP} & \\ \textsc{gap} & \langle\rangle\end{array}\right] \right\rangle \\ \textsc{syn} & \left[\begin{array}{ll}\textsc{val} & \langle\, \boxed{1}\, \rangle \\ \textsc{gap} & \langle\, \boxed{2}\, \rangle\end{array}\right]\end{array}\right]$$

The analysis of (43) then results in a linguistic object in which the second element of the ARG-ST list of *like* has the FORM value ⟨ *bagels* ⟩.

(43) Bagels, I like.

In a trace-based account, the FORM value of the extracted element would be the empty list.

Now, the problem is that not everybody agrees with the traceless analysis of unbounded dependencies. For instance, Levine & Hukari (2006) wrote a monograph discussing various versions of traceless accounts of extraction and argue against them. Chaves (2009) suggested solutions to some of the puzzles, but does not solve them entirely. While feature geometries that include a LOCAL feature allow researchers to assume trace-based analyses, the SBCG geometry makes this impossible. So those who use traces in their theories will never adopt the SBCG geometry. See Section 19 for more on empty elements.

A further advantage of having a package of information that is shared in nonlocal dependencies is that some information can be excluded from sharing by specifying it outside of such a package. This was used by Höhle (1994), Müller (1996d, 2002a), and Meurers (1999a) to account for partial verb phrase fronting in German. The generalization about partial verb phrase fronting in German is that verbs can be fronted together with some or all of their objects even though the verb does not form a VP in other contexts. For instance, *erzählen* 'tell' and *wird* 'will' in (44a) usually form a complex that may not be separated by scrambling a projection of *erzählen* to the left.

(44) a. dass er seiner Tochter ein Märchen erzählen wird
that he his daughter a fairy.tale tell will
'that he will tell his daughter a fairy tale'

b. * dass er seiner Tochter ein Märchen erzählen nie wird
that he his daughter a fairy.tale tell never will

Hinrichs & Nakazawa (1994) account for this by assuming a special schema for verbal complexes and by assuming that such verbal complexes are marked in a certain way (later this was encoded with the LEX feature with the value '+') and that heads that form a verbal complex select for appropriately marked elements. So in the analysis of (44a) *wird* 'will' selects a LEX + element and the word *erzählen* 'tell' satisfies this requirement.

The problem for many accounts of nonlocal dependencies is that the fronted material is not necessarily a single verb:

(45) a. Erzählen wird er seiner Tochter ein Märchen.
tell will he.NOM his.DAT daughter a.ACC fairytale
'He will tell his daughter a fairytale.'

b. Ein Märchen erzählen wird er seiner Tochter.
a fairytale.ACC tell will he.NOM his.DAT daughter

c. Seiner Tochter erzählen wird er das Märchen.
his.DAT daughter tell will he.NOM the.ACC fairytale

d. Seiner Tochter ein Märchen erzählen wird er.
his.DAT daughter a.ACC fairytale tell will he.NOM

If the local requirement of *wird* would be shared between filler and gap, the sentences in (45b)–(45d) would be ruled out, since the projections in the *Vorfeld* are not lexical elements but complex phrasal projections that are LEX −. Now, as pointed out by the authors mentioned above, this turns into a non-issue if not all information is shared between filler and gap. If LEX is outside LOCAL, the filler maybe LEX −, while the local constraints on the extracted element require the LEX value +. If all information is shared as in SBCG, no such solution is possible. Of course one could stick to the SBCG feature geometry and say that not the entire ARG-ST element is shared with the GAP element, but then one would end up with single structure sharings of all the features at the outermost level of a sign, that is, PHON, FORM, SYN, SEM, and CNTXT (see Sag 2012: 98 for these features)[21] and the whole point of grouping features is to avoid such multiple individual structure sharings in situations in which there is a systematic relation between feature values.

[21] Note that these structure sharings are necessary. It is not an option to leave the values of these features unspecified in the argument structure list. Since we have a model-theoretic view on things, an unspecified value would make the respective structures infinitely ambiguous, since infinitely many instantiation of these features are possible.

10.6.2.3 Selection of PHON and FORM values

The feature geometry of constructional HPSG has the PHON value outside of SYNSEM. Therefore verbs can select for syntactic and semantic properties of their arguments but not for their phonology. For example, they can require that an object has accusative case but not that it starts with a vowel. SBCG allows for the selection of phonological information (the feature is called FORM here) and one could try to justify this by pointing out that there are cases of selection in which the form of the selected argument depends on the form of the head. For example, the indefinite article in English has to be either *a* or *an* depending on whether the noun it is combined with starts with a vowel or not:

(46) a. an institute
     b. a house

So *institute* could select for *an*, while *house* selects for *a*. However, the situation is more complicate since prenominal adjectives influence the choice of determiner:

(47) a. a famous institute
     b. an interesting house

In (47) a phrasal nominal constituent is combined with the determiner and this shows that the phonology of the determiner has to be determined on the phrasal level and cannot be lexically encoded (unless one is willing to represent information about preceding adjuncts in the head noun somehow). The general question is whether there are phenomena that require a selection of FORM values by heads or whether all phonological phenomena are best handled by constraints on phonology on the phrasal level.

Note also that the treatment of raising and nonlocal dependencies in SBCG admits nonlocal selection of phonology values, since the FORM value of the filler is present at the ARG-ST list of the head from which the argument is extracted. In earlier HPSG versions only LOCAL information is shared and elements in valence lists do not have a PHON feature. In principle, SBCG could be used to model languages in which the phonology of a filler is relevant for a head from which it is extracted. So for instance *likes* can see the phonology of *bagels* in (48):

(48) Bagels, I think that Peter likes.

It would be possible to state constraints saying that the filler has to contain a vowel or two vowels or that it ends with a consonant. In addition all elements on the extraction path (*that* and *think*) can see the phonology of the filler as well. While there are languages that mark the extraction path, I doubt that there are languages that have phonological effects across long distances.

Similarly, the analysis of raising in SBCG assumes that the element on the valence list of the embedded verb is identical to an element in the ARG-ST list of the matrix verb (Sag 2012: 159). Hence, both verbs in (49) can see the phonology of the subject:

(49) Kim can eat apples.

In principle there could be languages in which the form of the downstairs verb depends on the presence of an initial consonant in phonology of the subject. English allows for long chains of raising verbs and one could imagine languages in which all the verbs on the way are sensitive to the phonology of the subject. Such languages probably do not exist.

Now, is this a problem? Not for me, but if one develops a general setup in a way to exclude everything that is not attested in the languages of the world (as for instance the selection of arguments of arguments of arguments), then it is a problem that heads can see the phonology of elements that are far away.

There are two possible conclusions for practitioners of SBCG: either the MOTHER feature could be given up since one agrees that theories that do not make wrong predictions are sufficiently constraint and one does not have to explicitly state what cannot occur in languages or one would have to react to the problem with nonlocally selected phonology values and therefore assume a SYNSEM or LOCAL feature that bundles information that is relevant in raising and nonlocal dependencies and does not include the phonology.[22] This supports the arguments I made on MOTHER and LOCAL in the previous subsections.

10.6.2.4 The VALENCE list

Another change from Constructional HPSG to SBCG involves the use of a single valence feature rather then three features SPR, SUBJ and COMPS that were suggested by Borsley (1987) to solve problems in earlier HPSG versions that used a single valence feature (SUBCAT). Borseley's suggestion was taken up by Pollard & Sag (1994: Chapter 9) and has been used in some version or other in other HPSG versions since then.

Sag (2012: 85) assumes that VPs are described by the following feature description:

(50) $\left[\text{SYN} \left[\text{VAL} \ \langle \text{NP} \rangle \right]\right]$

The problem with such an approach is that VPs differ from other phrasal projections in having an element on their VALENCE list. APs, NPs, and (some) PPs have an empty VALENCE list. In other versions of HPSG the complements are represented on the COMPS list and generalizations about phrases with fully saturated COMPS lists can be expressed directly. One such generalization is that projections with an empty COMPS list (NPs, PPs, VPs, adverbs, CPs) can be extraposed in German (Müller 1999a: Section 13.1.2).

10.6.2.5 Conclusion

Due to the conceptual problems with meta-statements and the relatively simple ways of getting around locality restrictions, the reorganization of features (MOTHER vs. SYN-

[22] If SYNSEM is reintroduced, the elements in the valence lists could be of type *synsem*. Information about phonology would not be part of the description of the selected elements. This would not solve the problem of partial verb phrase fronting though, since the LEX feature is selected for (hence part of the information under SYNSEM) but not shared with the filler. One would need a LOCAL feature in addition to SYNSEM. See Section 10.6.2.2.

SEM) does not bring with it any advantages. Since the grammar becomes more complex due to the meta-constraint, we should reject this change.[23] Other changes in the feature geometry (elimination of the LOCAL feature and use of a single valence feature) are problematic as well. However, if we do reject the revised feature geometry and revert to the feature geometry that was used before, then Sign-Based Construction Grammar and Constructional HPSG (Sag 1997) are (almost) indistinguishable.

10.6.3 Embodied Construction Grammar

Embodied Construction Grammar was developed by Bergen & Chang (2005) and there are some implementations of fragments of German that use this format (Porzel et al. 2006). In the following, I will briefly present the formalism using an example construction. (51) gives the DetNoun construction:[24]

(51)

Construction DetNoun		
subcase of RefExp		
constructional		
d:Determiner		
c:CommonNoun		
self.case	↔	d.case
self.number	↔	c.number
d.gender	↔	c.gender
d.case	↔	c.case
d.number	↔	c.number
form		
d.f **before** c.f		
meaning		
self.m	↔	c.m

This representational form is reminiscent of PATR-II grammars (Shieber, Uszkoreit, Pereira, Robinson & Tyson 1983): as in PATR-II, the daughters of a construction are given names. As such, (51) contains the daughters c and d. d is a determiner and c is a common noun. It is possible to refer to the construction itself via the object **self**. Constructions (and also their daughters) are feature-value descriptions. Structure sharing is repre-

[23] In Müller (2013b: 253) I claimed that SBCG needs an additional feature in comparison to other variants of HPSG. As Van Eynde (2015) points out this is not true for more recent variants of HPSG since they have the SYNSEM feature, which is now superfluous. (Van Eynde refers to the LOCAL feature, but the LOCAL feature was eliminated because it was considered superfluous because of the lexical analysis of extraction, see Section 10.6.2.2). If one simply omits the MOTHER feature from SBCG one is back to the 1987 version of HPSG (Pollard & Sag 1987), which also used a SYN and a SEM feature. What would be missing would be the locality of selection (Sag 2012: 149) that was enforced to some extent by the SYNSEM feature. Note that the locality of selection that is enforced by SYNSEM can be circumvented by the use of relational constraints as well (see Frank Richter and Manfred Sailer's work on collocations (Richter & Sailer 1999a; Soehn & Sailer 2008)). So in principle, we end up with style guides in this area of grammar as well.

[24] For a similar construction, see Bergen & Chang (2005: 162).

sented by path equations. For example, d.gender ↔ c.gender states that the value of the gender feature of the determiner is identical to the gender feature of the noun. As well as restrictions on features, there are restrictions on the form. d.f **before** c.f states that the form contribution of the determiner must occur before that of the noun. Bergen & Chang (2005) differentiate between immediate (**meets**) and non-immediate precedence (**before**). Part of the information represented under f is the orthographic form (f.orth). The inheritance relation is given explicitly in the construction as in Kay & Fillmore (1999).

The construction in (51) can be represented in a similar way to the format used in Chapter 6:

(52)
$$\begin{bmatrix} \text{F}|\text{ORTH} & \boxed{1} \oplus \boxed{2} \\ \text{CASE} & \boxed{3} \\ \text{NUMBER} & \boxed{4} \\ \text{M} & \boxed{5} \\ \text{DTRS} & \left\langle \begin{bmatrix} \text{F}|\text{ORTH} & \boxed{1} \\ \text{CASE} & \boxed{3} \\ \text{NUMBER} & \boxed{4} \\ \text{GENDER} & \boxed{6} \\ \multicolumn{2}{l}{\textit{Determiner}} \end{bmatrix}, \begin{bmatrix} \text{F}|\text{ORTH} & \boxed{2} \\ \text{CASE} & \boxed{3} \\ \text{NUMBER} & \boxed{4} \\ \text{GENDER} & \boxed{6} \\ \text{M} & \boxed{5} \\ \multicolumn{2}{l}{\textit{CommonNoun}} \end{bmatrix} \right\rangle \\ \multicolumn{2}{l}{\textit{DetNoun}} \end{bmatrix}$$

The structure in (52) corresponds to a construction where the determiner directly precedes the noun because the form contribution of the determiner has been combined with that of the noun. This strict adjacency constraint makes sense as the claim that the determiner must precede the noun is not restrictive enough since sequences such as (53b) would be allowed:

(53) a. [dass] die Frauen Türen öffnen
that the women doors open
'that the woman open doors'

b. * die Türen öffnen
Frauen

If discontinuous phrases are permitted, *die Türen* 'the doors' can be analyzed with the DetNoun Construction although another noun phrase intervenes between the determiner and the noun (Müller: 1999a: 424; 1999c). The order in (53b) can be ruled out by linearization constraints or constraints on the continuity of arguments. If we want the construction to require that the determiner and noun be adjacent, then we would simply use **meets** instead of **before** in the specification of the construction.

This discussion has shown that (52) is more restrictive than (51). There are, however, contexts in which one could imagine using discontinuous constituents such as the deviant one in (53b). For example, discontinuous constituents have been proposed for verbal complexes, particle verbs and certain coordination data (Wells 1947). Examples for analyses with discontinuous constituents in the framework of HPSG are Reape (1994), Kathol (1995), Kathol (2000), Crysmann (2008), and Beavers & Sag (2004).[25] These analyses, which are discussed in Section 11.7.2.2 in more detail, differ from those previously presented in that they use a feature DOMAIN instead of or in addition to the daughters features. The value of the DOMAIN feature is a list containing the head and the elements dependent on it. The elements do not have to necessarily be adjacent in the utterance, that is, discontinuous constituents are permitted. Which elements are entered into this list in which way is governed by the constraints that are part of the linguistic theory. This differs from the simple **before** statement in ECG in that it is much more flexible and in that one can also restrict the area in which a given element can be ordered since elements can be freely ordered inside their domain only.

There is a further difference between the representation in (51) and the general HPSG schemata: in the ECG variant, linearization requirements are linked to constructions. In HPSG and GPSG, it is assumed that linearization rules hold generally, that is, if we were to assume the rules in (54), we would not have to state for each rule explicitly that shorter NPs tend to precede longer ones and that animate nouns tend to occur before inanimate ones.

(54) a. S → NP[nom], NP[acc], V
b. S → NP[nom], NP[dat], V
c. S → NP[nom], NP[dat], NP[acc], V
d. S → NP[nom], NP[acc], PP, V

It is possible to capture these generalizations in ECG if one specifies linearization constraints for more general constructions and more specific constructions inherit them from these. As an example, consider the Active-Ditransitive Construction discussed by

[25] Crysmann, Beaver and Sag deal with coordination phenomena. For an analysis of coordination in TAG that also makes use of discontinuous constituents, see Sarkar & Joshi (1996) and Section 21.6.2.

Bergen & Chang (2005: 170):

(55)

Construction Active-Ditransitive
subcase of Pred-Expr
constructional
agent:Ref-Expr
action:Verb
recipient:Ref-Expr
theme:Ref-Expr
…
form
agent.f **before** action.f
action.f **meets** recipient.f
recipient.f **meets** theme.f
meaning
…

These restrictions allow the sentences in (56a,b) and rule out those in (56c):

(56) a. Mary tossed me a drink.
b. Mary happily tossed me a drink.
c. * Mary tossed happily me a drink.

The restriction agent.f **before** action.f forces an order where the subject occurs before the verb but also allows for adverbs to occur between the subject and the verb. The other constraints on form determine the order of the verb and its object: the recipient must be adjacent to the verb and the theme must be adjacent to the recipient. The requirement that an agent in the active must occur before the verb is not specific to ditransitive constructions. This restriction could therefore be factored out as follows:

(57)

Construction Active-Agent-Verb
subcase of Pred-Expr
constructional
agent:Ref-Expr
action:Verb
form
agent.f **before** action.f

The Active-Ditransitive Construction in (55) would then inherit the relevant information from (57).

In addition to the descriptive means used in (51), there is the evokes operator (Bergen & Chang 2005: 151–152). An interesting example is the representation of the term hypotenuse: this concept can only be explained by making reference to a right-angled triangle (Langacker 1987: Chapter 5). Jerome Feldman (slides) gives the following formalization:

(58)

Schema hypotenuse
  subcase of line-segment
  evokes right-triangle **as** rt
constraints
  self ↔ rt.long-side

This states that a hypotenuse is a particular line segment, namely the longest side of a right-angled triangle. The concept of a right-angled triangle is activated by means of the evokes operator. Evokes creates an instance of an object of a certain type (in the example, *rt* of type *right-triangle*). It is then possible to refer to the properties of this object in a schema/construction.

The feature description in (59) is provided in the notation from Chapter 6. It is the equivalent to (58).

(59) $$\boxed{1}\begin{bmatrix}\textsc{evokes} & \left\langle \begin{bmatrix}\textsc{long-side} & \boxed{1}\\ \multicolumn{2}{l}{\textit{right-triangle}}\end{bmatrix} \right\rangle\\ \multicolumn{2}{l}{\textit{hypotenuse}}\end{bmatrix}$$

The type *hypotenuse* is a subtype of *line-segment*. The value of EVOKES is a list since a schema or construction can evoke more than one concept. The only element in this list in (59) is an object of type *right-triangle*. The value of the feature LONG-SIDE is identified with the entire structure. This essentially means the following: I, as a hypotenuse, am the long side of a right-angled triangle.

Before turning to FCG in the next subsection, we can conclude that ECG and HPSG are notational variants.

10.6.4 Fluid Construction Grammar

Van Trijp (2013, 2014) claims that SBCG and HPSG are fundamentally different from Fluid Construction Grammar (FCG). He claims that the former approaches are generative ones while the latter is a cognitive-functional one. I think that it is not legitimate to draw these distinctions on the basis of what is done in FCG.[26] I will comment on this at various places in this section. I first deal with the representations that are used in FCG, talk about argument structure constructions, the combination operations fusion and merging that are used in FCG and then provide a detailed comparison of FCG and SBCG/HPSG.

[26] Steels (2013: 153) emphasizes the point that FCG is a technical tool for implementing constructionist ideas rather than a theoretical framework of its own. However, authors working with the FCG system publish linguistic papers that share a certain formal background and certain linguistic assumptions. So this section addresses some of the key assumptions made and some of the mechanisms used.

10.6.4.1 General remarks on the representational format

Fluid Construction Grammar (FCG, Steels 2011) is similar to HPSG in that it uses attribute value matrices to represent linguistic objects. However, these AVMs are untyped as in LFG. Since there are no types, there are no inheritance hierarchies that can be used to capture generalizations, but one can use macros to reach similar effects. Constructions can refer to more general constructions (van Trijp 2013: 105). Every AVM comes with a name and can be depicted as follows:

(60)
$$\boxed{\begin{array}{l} \textit{unit-name} \\ \begin{bmatrix} \textsc{feature}_1 & \textit{value}_1 \\ \dots & \\ \textsc{feature}_n & \textit{value}_n \end{bmatrix} \end{array}}$$

Linguistic objects have a form and a meaning pole. The two poles could be organized into a single feature description by using a SYN and a SEM feature, but in FCG papers the two poles are presented separately and connected via a double arrow. (61) is an example:

(61) *Kim* according to van Trijp (2013: 99):

$$\boxed{\begin{array}{l} \textit{Kim-unit}\ \text{(semantic pole)} \\ \begin{bmatrix} \textsc{meaning} & \begin{bmatrix} \textsc{individual} & \textit{kim} \end{bmatrix} \\ \textsc{sem-cat} & \begin{bmatrix} \textsc{class} & \textit{person} \end{bmatrix} \end{bmatrix} \end{array}}$$

$$\updownarrow$$

$$\boxed{\begin{array}{l} \textit{Kim-unit}\ \text{(syntactic pole)} \\ \begin{bmatrix} \textsc{form} & \begin{bmatrix} \textsc{string} & \textit{kim} \end{bmatrix} \\ \textsc{syn-cat} & \begin{bmatrix} \textsc{lex-class} & \textit{proper-noun} \end{bmatrix} \end{bmatrix} \end{array}}$$

Depending on the mode in which the lexical items are used, the syntactic pole or the semantic pole is used first. The first processing step is a matching phase in which it is checked whether the semantic pole (for generation) or the syntactic pole (for parsing) matches the structure that was build so far. After this test for unification, the actual unification, which is called merging, is carried out. After this step, the respective other pole (syntax for generation and semantics for parsing) is merged. This is illustrated in Figure 10.2 on the facing page.

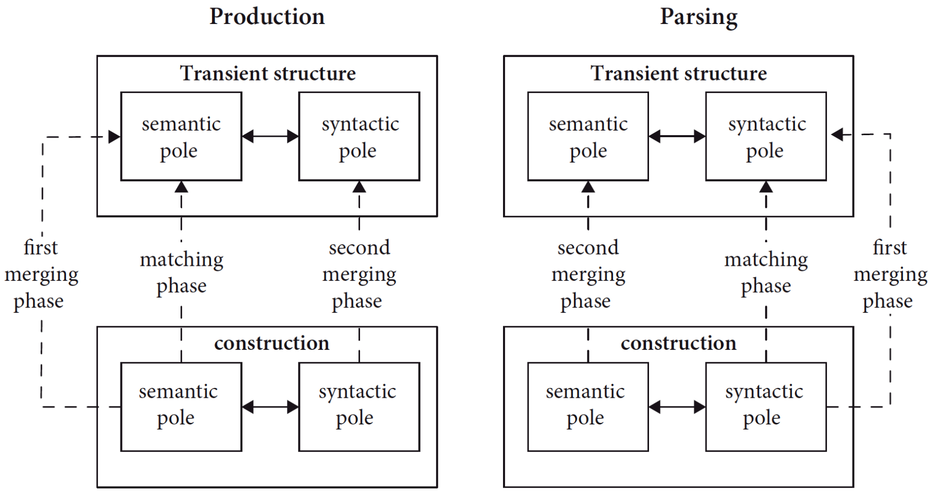

Figure 10.2: Generation and parsing in FCG (van Trijp 2011: 99)

10.6.4.2 Argument Structure Constructions

Fluid Construction Grammar assumes a phrasal approach to argument structure, that is, it is assumed that lexical items enter into phrasal configurations that contribute independent meaning (van Trijp 2011). The FCG approach is one version of implementing Goldberg's plugging approach to argument structure constructions (Goldberg 1995). Van Trijp suggests that every lexical item comes with a representation of potential argument roles like Agent, Patient, Recipient, and Goal. Phrasal argument structure constructions are combined with the respective lexical items and realize a subset of the argument roles, that is they assign them to grammatical functions. Figure 10.3 on the next page shows an example: the verb *sent* has the semantic roles Agent, Patient, Recipient, and Goal. Depending on the argument structure construction that is chosen, a subset of these roles is selected for realization.[27]

Note that under such an approach, it is necessary to have a passive variant of every active construction. For languages that allow for the combination of passive and impersonal constructions, one would be forced to assume a transitive-passive-impersonal construction. As was argued in Müller (2006: Section 2.6) free datives (commodi/incommodi)

[27] It is interesting to note here that van Trijp (2011: 141) actually suggests a lexical account since every lexical item is connected to various phrasal constructions via coapplication links. So every such pair of a lexical item and a phrasal construction corresponds to a lexical item in Lexicalized Tree Adjoining Grammar (LTAG). See also Müller & Wechsler (2014a: 25) on Goldberg's assumption that every lexical item is associated with phrasal constructions.

Note that such coapplication links are needed since without them the approach cannot account for cases in which two or more argument roles can only be realized together but not in isolation or in any other combination with other listed roles.

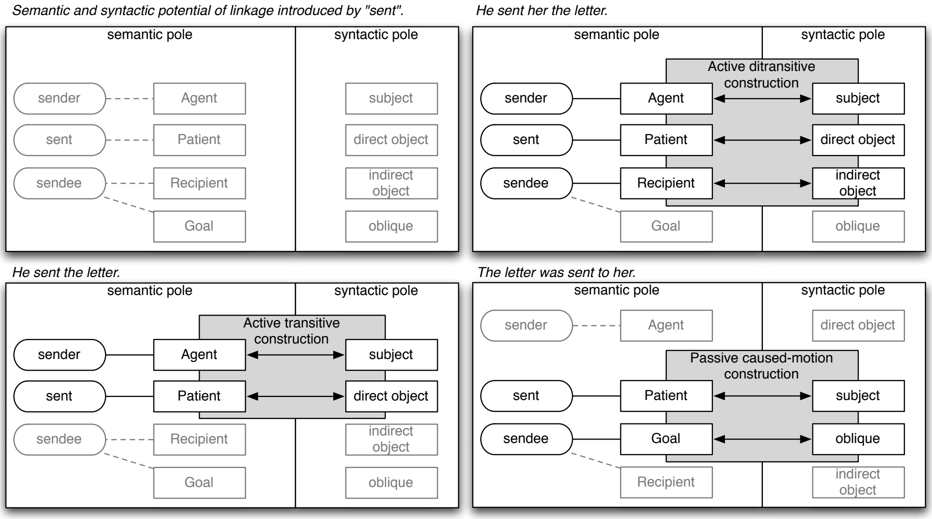

Figure 10.3: Lexical items and phrasal constructions. Figure taken from van Trijp (2011: 122)

in German can be added to almost any construction. They interact with the dative passive and hence should be treated as arguments. So, for the resultative construction one would need an active variant, a passive variant, a variant with dative argument, a variant with dative argument and dative passive, and a middle variant. While it is technically possible to list all these patterns and it is imaginable that we store all this information in our brains, the question is whether such listings really reflect our linguistic knowledge. If a new construction comes into existence, lets say an active sentence pattern with a nominative and two datives in German, wouldn't we expect that this pattern can be used in the passive? While proposals that establish relations between active and passive constructions would predict this, alternative proposals that just list the attested possibilities do not.

The issue of how such generalizations should be captured was discussed in connection with the organization of the lexicon in HPSG (Flickinger 1987; Meurers 2001). In the lexical world, one could simply categorize all verbs according to their valence and say that *loves* is a transitive verb and the passive variant *loved* is an intransitive verb. Similarly *gives* would be categorized as a ditransitive verb and *given* as a transitive one. Obviously this misses the point that *loved* and *given* share something: they both are related to their active form in a systematic way. This kind of generalization was called a horizontal generalization as compared to vertical generalizations, which describe classes in an inheritance hierarchy.

The issue is independent of the lexical organization of knowledge, it can be applied to phrasal representations as well. Phrasal constructions can be organized in hierarchies (vertical), but the relation between certain variants is not covered by this. The analog to the lexical rules in a lexical approach are GPSG-like metarules in a phrasal approach. So what seems to be missing in FCG is something that relates phrasal patterns, e. g. allo-constructions (Cappelle 2006; Goldberg 2014: 116, see also footnote 12).

10.6.4.3 Fusion, matching and merging

As was pointed out by Dowty (1989: 89–90), checking for semantic compatibility is not sufficient when deciding whether a verb may enter (or be fused with) a certain construction. The example is the contrast between *dine*, *eat*, and *devour*. While the thing that is eaten may not be realized with *dine*, its realization is optional with *eat* and obligatory with *devour*. So the lexical items have to come with some information about this.

Van Trijp (2011) and Steels & van Trijp (2011) make an interesting suggestion that could help here: every verb comes with a list of potential roles and argument structure constructions can pick subsets of these roles (see Figure 10.3). This is called *matching*: introducing new argument roles is not allowed. This would make it possible to account for *dine*: one could say that there is something that is eaten, but that no Theme role is made available for linking to grammatical functions. To account for the extension of argument roles as it is observed in the caused motion construction Steels & van Trijp (2011) suggest a process called *merging*. Merging is seen as a repair strategy: if an utterance involves an intransitive verb and some other material, the utterance cannot be processed with matching alone. For example, when processing (62), *he sneezed* could be parsed, but *the foam* and *off the cappuccino* would be unintegrated.

(62) He sneezed the foam off the cappuccino.

So, Steels & van Trijp (2011: 319–320) suggest that only if regular constructions cannot apply, merging is allowed. There are two problems with this: first, human language is highly ambiguous and in the case at hand this could result in situations in which there is a reading for an utterance, so that the repair strategy would never kick in. Consider (63):

(63) Er schlug den Mann tot.
he beat the man dead
'He beat the man to death.' or 'He beat the dead man.'

(63) has two readings: the resultative reading in which *tot* 'dead' expresses the result of the beating and another reading in which *tot* is a depictive predicate. So, the problem is that (63) has a reading and therefore repair is not triggered: *schlug* 'beat' is used with the transitive construction and *tot* is an adjunct (see Winkler 1997). However, the more likely analysis of (63) is the one with the resultative analysis, in which the valence frame is extended by an oblique element. So this means that one has to allow the application of merging independent of other analyses that might be possible. As Steels and van Trijp note (p. 320), if merging is allowed to apply freely, utterances like (64a) will be allowed and of course (64b) as well.

(64) a. * She sneezed her boyfriend.
b. * She dined a steak.

In (64) *sneeze* and *dined* are used in the transitive construction.

The way out of this dilemma is to establish information in lexical items that specifies in which syntactic environments a verb can be used. This information can be weighted and for instance the probability of *dine* to be used transitively would be extremely low. Steels and van Trijp would connect their lexical items to phrasal constructions via so-called coapplication links and the strength of this link would be very low for *dine* and the transitive construction and reasonably high for *sneeze* and the caused-motion construction. This would explain the phenomena (and in a usage-based way), but it would be a lexical approach, as it is common in CG, HPSG, SBCG, and DG.

10.6.4.4 Long-distance dependencies

Van Trijp (2014) compares the SLASH-based approaches that are used in GPSG, HPSG, and SBCG with the approach that he suggests within the framework of FCG. He claims that there are fundamental differences between SBCG and FCG and assigns SBCG to the class of generative grammars, while placing FCG in the class of cognitive-functional approaches. He claims that his cognitive-functional approach is superior in terms of completeness, explanatory adequacy, and theoretical parsimony (p. 2). What van Trijp (2014) suggests is basically an analysis that was suggested by Reape (2000) in unpublished work (see Reape (1994) for a published version of an linearization-based approach and Kathol (2000); Müller (1996c, 1999a, 2002a) for linearization-based approaches that despite of being linearization-based assume the SLASH approach for nonlocal dependencies). Van Trijp develops a model of grammar that allows for discontinuous constituents and just treats the serialization of the object in sentences like (65) as an alternative linearization option.

(65) a. This book, I read.
b. What did the boy hit?

Van Trijp's analysis involves several units that do not normally exist in phrase structure grammars, but can be modeled via adjacency constraints or represent relations between items which are part of lexical representations in HPSG/SBCG anyway. An example is the subject-verb anchor that connects the subject and the verb to represent the fact that these two items play an important functional role. Figure 10.4 on the facing page shows the analysis of (66).

(66) What did the boy hit?

As can be seen in the figure, van Trijp also refers to information structural terms like topic and focus. It should be noted here that the analysis of information structure has quite some history in the framework of HPSG (Engdahl & Vallduví 1996; Kuhn 1995, 1996; Günther et al. 1999; Wilcock 2001; De Kuthy 2002; Paggio 2005; Bildhauer 2008;

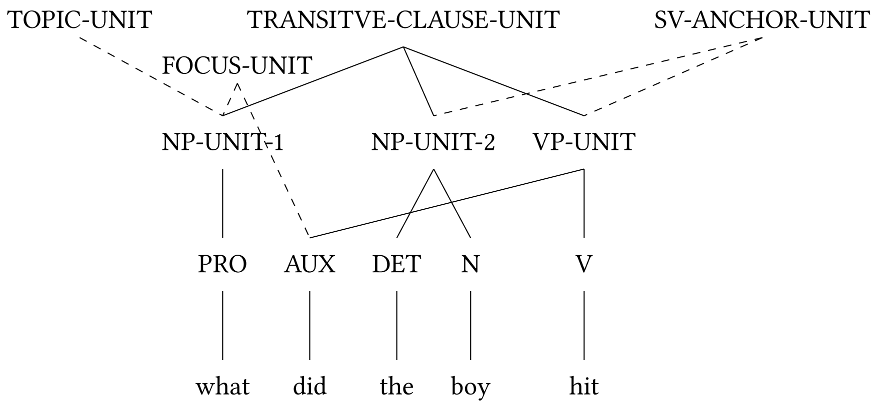

Figure 10.4: The analysis of *What did the boy hit?* according to van Trijp (2014)

Bildhauer & Cook 2010). The fact that information structure is not talked about in syntax papers like Sag (2012) does not entail that information structure is ignored or should be ignored in theories like HPSG and SBCG. So much for completeness. The same holds of course for explanational adequacy. This leaves us with theoretical parsimony, but before I comment on this, I want to discuss van Trijp's analysis in a little bit more detail in order to show that many of his claims are empirically problematic and that his theory therefore cannot be explanatory since empirical correctness is a precondition for explanatory adequacy.

Van Trijp claims that sentences with nonlocal dependency constructions in English start with a topic.[28] Bresnan's sentences in (2) and (3) were discussed on page 222 (Bresnan 2001: 97) and are repeated below for convenience:

(67) Q: What did you name your cat?
A: Rosie I named her. (*Rosie* = FOCUS)

(68) Q: What did you name your pets?
A: My dog, I named Harold. My cat, I named Rosie. (*my dog*, *my cat* = TOPIC)

These sentences show that the pre-subject position is not unambiguously a topic or a focus position. So, a statement saying that the fronted element is a topic is empirically not correct. If this position is to be associated with an information structural function, this association has to be a disjunction admitting both topics and focused constituents.

A further problematic aspect of van Trijp's analysis is that he assumes that the auxiliary *do* is an object marker (p. 10, 22) or a non-subject marker (p. 23). It is true that *do*

[28] He uses the following definitions for topic and focus: *Topicality is defined in terms of aboutness: the topic of an utterance is what the utterance is 'about'. Focality is defined in terms of salience: focus is used for highlighting the most important information given the current communicative setting.* (van Trijp 2014: 15)

support is not necessary in subject questions like (69a), but only in (69b), but this does not imply that all items that are followed by *do* are objects.

(69) a. Who saw the man?
b. Who did John see?

First, *do* can be used to emphasize the verb:

(70) Who *did* see the man?

Second all types of other grammatical functions can precede the verb:

(71) a. Where did you see the man? (adverbial)
b. How tall is the man? (predicative)
c. What did John consider Peter? (predicative)
d. What does this book cost? (adverbial)
e. About what did you talk? (prepositional object)

And finally, even a subject can appear in front of *do* if it is extracted from another clause:

(72) Who does he think saw this man? (subject)

There is a further empirical problem: approaches that assume that a filler is related to its origin can explain scope ambiguities that only arise when an element is extracted. Compare for instance the sentence in (73a) with the sentences in (73b, c): although the order of *oft* 'often' and *nicht* 'not' in (73a) and (73c) is the same, (73a) is ambiguous but (73c) is not.

(73) a. Oft liest er das Buch nicht.
often reads he the book not
'It is often that he does not read the book.' or 'It is not the case that he reads the book often.'
b. dass er das Buch nicht oft liest
that he the book not often reads
'that it is not the case that he reads the book often'
c. dass er das Buch oft nicht liest
that he the book often not reads
'that it is often that he does not read the book'

(73a) has the two readings that correspond to (73b) and (73c). A purely linearization-based approach probably has difficulties to explain this. A SLASH-based approach can assume that (73a) has a gap (or some similar means for the introduction of nonlocal dependencies) at the position of *oft* in (73b) or (73c). The gap information is taken into account in the semantic composition at the site of the gap. This automatically accounts for the observed readings.

Another empirical problem that has to be solved is the existence of extraction path marking languages. Bouma, Malouf & Sag (2001a) list a number of languages in which elements vary depending on the existence or absence of a gap in a constituent they attach to. For instance, Irish has complementizers that have one form if the clause they attach to has an element extracted and another form if it does not. SLASH-based proposals can account for this in a straight-forward way: the fact that a constituent is missing in a phrase is represented in the SLASH value of the trace and this information is percolated up the tree. So even complex structures contain the information that there is a constituent missing inside them. Complementizers that are combined with sentences therefore can select sentences with SLASH values that correspond to their inflection. Van Trijp's answer to this challenge is that all languages are different (van Trijp 2014: 22) and that the evidence from one language does not necessarily mean that the analysis for that language is also appropriate for another language. While I agree with this view in principle (see Section 13.1), I do think that extraction is a rather fundamental property of languages and that nonlocal dependencies should be analyzed in parallel for those languages that have it.

Van Trijp points out that SBCG does not have a performance model and contrasts this with FCG. On page 11 he states:

> So parsing starts by segmenting the utterance into discrete forms, which are then categorized into words by morphological and lexical constructions, and which can then be grouped together as phrases (see Steels, 2011b, for a detailed account of lexico-phrasal processing in FCG). So the parser will find similar constituents for all four utterances, as shown in examples (21–24). Since auxiliary-*do* in example (24) falls outside the immediate domain of the VP, it is not yet recognized as a member of the VP.
>
> All of these phrases are disconnected, which means that the grammar still has to identify the relations between the phrases. (van Trijp 2014: 7)

Van Trijp provides several tree fragments that contain NPs for subject and object and states that these have to be combined in order to analyze the sentences he discusses. This is empirically inadequate: if FCG does not make the competence/performance distinction, then the way utterances are analyzed should reflect the way humans process language (and this is what is usually claimed about FCG). However, all we know about human language processing points towards an incremental processing, that is, we process information as soon as it is available. We start to process the first word taking into account all of the relevant aspects (phonology, stress, part of speech, semantics, information structure) and come up with an hypothesis about how the utterance could proceed. As soon as we have two words processed (in fact even earlier: integration already happens during the processing of words) we integrate the second word into what we know already and continue to follow our hypothesis, or revise it, or simply fail. See Section 15.2 for details on processing and the discussion of experiments that show that processing is incremental. So, we have to say that van Trijp's analysis fails on empirical grounds: his modeling of performance aspects is not adequate.

The parsing scheme that van Trijp describes is pretty much similar to those of HPSG parsers, but these usually come without any claims about performance. Modeling performance is rather complex since a lot of factors play a role. It is therefore reasonable to separate competence and performance and continue to work the way it is done in HPSG and FCG. This does not mean that performance aspects should not be modeled, in fact psycholinguistic models using HPSG have been developed in the past (Konieczny 1996), but developing both a grammar with large coverage and the performance model that combines with it demands a lot of resources.

I now turn to parsimony: van Trijp uses a subject-verb anchor construction that combines the subject and the main verb. Because of examples like (74) it must be possible to have discontinuous subject-verb constructions:[29]

(74) Peter often reads books.

But if such constructions can be discontinuous one has to make sure that (75b) cannot be an instantiation of the subject-verb construction:

(75) a. The boy I think left.
b. * I the boy think left.

Here it is required to have some adjacency between the subject and the verb it belongs to, modulo some intervening adverbials. This is modelled quite nicely in phrase structure grammars that have a VP node. Whatever the internal structure of such a VP node may be, it has to be adjacent to the subject in sentences like the ones above. The dislocated element has to be adjacent to the complex consisting of subject and VP. This is what the Filler-Head Schema does in HPSG and SBCG. Van Trijp criticizes SBCG for having to stipulate such a schema, but I cannot see how his grammar can be complete without a statement that ensures the right order of elements in sentences with fronted elements.

Van Trijp stated that FCG differs from what he calls generative approaches in that it does not want to characterize only the well-formed utterances of a language. According to him, the parsing direction is much more liberal in accepting input than other theories. So it could well be that he is happy to find a structure for (75b). Note though that this is incompatible with other claims made by van Trijp: he argued that FCG is superior to other theories in that it comes with a performance model (or rather in not separating competence from performance at all). But then (75b) should be rejected both on competence and performance grounds. It is just unacceptable and speakers reject it for whatever reasons. Any sufficiently worked out theory of language has to account for this.

One of the success stories of non-transformational grammar is the SLASH-based analysis of nonlocal dependencies by Gazdar (1981b). This analysis made it possible for the

[29] Unless modals and tense auxiliaries are treated as main verbs (which they should not in English), constructions with modals seem to be another case where the subject and the main verb are not adjacent:

(i) a. Peter will read the book.
b. Peter has read the book.

first time to explain Ross's Across the Board Extraction (Ross 1967). The examples were already discussed on page 199 and are repeated here for convenience:

(76) a. The kennel which Mary made and Fido sleeps in has been stolen. (= S/NP & S/NP)

b. The kennel in which Mary keeps drugs and Fido sleeps has been stolen. (= S/PP & S/PP)

c. * The kennel (in) which Mary made and Fido sleeps has been stolen. (= S/NP & S/PP)

The generalization is that two (or more) constituents can be coordinated if they have identical syntactic categories and identical SLASH values. This explains why *which* and *in which* in (76a,b) can fill two positions in the respective clauses. Now, theories that do not use a SLASH feature for the percolation of information about missing elements have to find different ways to make sure that all argument slots are filled and that the correct correspondence between extracted elements and the respective argument role is established. Note that this is not straightforward in models like the one suggested by van Trijp, since he has to allow the preposition *in* to be combined with some material to the left of it that is simultaneously also the object of *made*. Usually an NP cannot simply be used by two different heads as their argument. As an example consider (77a):

(77) a. * John said about the cheese that I like.

b. John said about the cheese that I like it.

If it would be possible to use material several times, a structure for (77a) would be possible in which *the cheese* is the object of the preposition *about* and of the verb *like*. This sentence, however, is totally out: the pronoun *it* has to be used to fill the object slot.

There is a further problem related to discontinuity. If one does not restrict continuity, then constituent orders like (78b) are admitted by the grammar:

(78) a. Deshalb klärt, dass Peter kommt, ob Klaus spielt.
therefore resolves that Peter comes whether Klaus plays
'Therefore that Peter comes resolves whether Klaus plays.'

b. * Deshalb klärt dass ob Peter Klaus kommt spielt.
therefore resolves that whether Peter Klaus comes plays

The interesting thing about the word salad in (78b) is that the constituent order within the *dass* clause and within the *ob* clause is correct. That is, the complementizer precedes the subject, which in turn precedes the verb. The problem is that the constituents of the two clauses are mixed.

In a model that permits discontinuous constituents, one cannot require that all parts of an argument have to be arranged after all parts that belong to another argument since discontinuity is used to account for nonlocal dependencies. So, it must be possible to have *Klaus* before other arguments (or parts of other arguments) since *Klaus* can be extracted. An example of mixing parts of phrases is given in (79):

(79) Dieses Buch hat der Mann mir versprochen, seiner Frau zu geben, der gestern
this book has the man me promised his wife to give who yesterday
hier aufgetreten ist.
here performed is

'The man who performed here yesterday promised me to give this book to his wife.'

We see that material that refers to *der Mann* 'the man', namely the relative clause *der gestern hier aufgetreten ist* 'who performed here yesterday', appears to the right. And the object of *geben* 'to give', which would normally be part of the phrase *dieses Buch seiner Frau zu geben* 'this book his wife to give' appears to the left. So, in general it is possible to mix parts of phrases, but this is possible in a very restricted way only. Some dependencies extend all the way to the left of certain units (fronting) and others all the way to the right (extraposition). Extraposition is clause-bound, while extraction is not. In approaches like GPSG, HPSG and SBCG, the facts are covered by assuming that constituents for a complete clause are continuous apart from constituents that are fronted or extraposed. The fronted and extraposed constituents are represented in SLASH and EXTRA (Keller 1995; Müller 1999a), respectively, rather than in valence features, so that it is possible to require of constituents that have all their valents saturated to be continuous.

Summing up the discussion of parsimony, it has to be said that van Trijp has to provide the details on how continuity is ensured. The formalization of this is not trivial and only after this is done can FCG be compared with the SLASH-based approach.

In addition to all the points discussed so far, there is a logical flaw in van Trijp's argumentation: he states that:

> whereas the filler-gap analysis cannot explain why *do*-support does not occur in *wh*-questions where the subject is assigned questioning focus, this follows naturally from the interaction of different linguistic perspectives in this paper's approach. (van Trijp 2014: 22)

The issue here is whether a filler-gap analysis or an analysis with discontinuous constituents is suited better for explaining the data. A correct argumentation against the filler-gap analysis would require a proof that information structural or other functional constraints cannot be combined with this analysis. This proof was not provided and in fact I think it cannot be provided since there are approaches that integrate information structure. Simply pointing out that a theory is incomplete does not falsify a theory. This point was already made in my review of Boas (2003) and in a reply to Boas (2014). See Müller (2005a: 655–656), Müller (2007b: Chapter 20), and Müller & Wechsler (2014b: Footnote 15).

The conclusion about the FCG analysis of nonlocal dependencies is that there are some empirical flaws that can be easily fixed or assumptions that can simply be dropped (role of *do* as object marker, claim that the initial position in English fronting construction is the topic), some empirical shortcomings (coordination, admittance of illformed utterances with discontinuous constituents), some empirical problems when the analysis is extended to other languages (scope of adjuncts in German), and the parsimony of the analyses is not really comparable since the restrictions on continuity are not really worked out (or at least not published). If the formalization of restrictions on continuity in FCG turns out to be even half as complex as the formalization that is necessary for accounts of nonlocal dependencies (extraction and extraposition) in linearization-based HPSG that Reape (2000) suggested,[30] the SLASH-based analysis would be favorable.

In any case, I do not see how nonlocal dependencies could be used to drive a wedge between SBCG and FCG. If there are functional considerations that have to be taken into account, they should be modeled in both frameworks. In general, FCG should be more restrictive than SBCG since FCG claims to integrate a performance model, so both competence and performance constraints should be operative. I will come back to the competence/performance distinction in the following section, which is a more general comparison of SBCG and FCG.

10.6.4.5 Comparison to Sign-Based Construction Grammar/HPSG

According to van Trijp (2013), there are the differences shown in Table 10.1. These differences will be discussed in the following subsections.

Table 10.1: Differences between SBCG and FCG according to van Trijp (2013: 112)

Scientific model	Theoretical physics (abstract calculus)	Evolutionary theory (complex adaptive system)
Linguistic approach	Generative (competence model)	Cognitive-functional (parsing and production)
Formalization	Mathematical (amenable for implementation)	Computational (implemented)
Constructions	Static type constraints	Dynamic mappings
Constructicon	Signature and grammar	Open-ended inventory
Processing	Assumption of processing-independence	Bidirectional processing model

10.6.4.5.1 Competence/performance distinction

As for the linguistic approach, the use of the term *generative* is confusing. What van Trijp means – and also explains in the paper – is the idea that one should separate

[30] See Kathol & Pollard (1995) for a linearization-based account of extraposition. This account is implemented in the Babel System (Müller 1996c). See (Müller 1999d) on restricting discontinuity. Linearization-based approaches were argued to not be able to account for apparent multiple frontings in German (Müller 2005c, 2015b) and hence linearization-based approaches were replaced by more traditional variants that allow for continuous constituents only.

competence and performance. We will deal with both the generative-enumerative vs. constraint-based view and with the competence/performance distinction in more detail in the Chapters 14 and 15, respectively. Concerning the cognitive-functional approach, van Trijp writes:

> The goal of a cognitive-functional grammar, on the other hand, is to explain how speakers express their conceptualizations of the world through language (= *production*) and how listeners analyze utterances into meanings (= *parsing*). Cognitive-functional grammars therefore implement both a competence and a processing model. (van Trijp 2013: 90)

It is true that HPSG and SBCG make a competence/performance distinction (Sag & Wasow 2011). HPSG theories are theories about the structure of utterances that are motivated by distributional evidence. These theories do not contain any hypothesis regarding brain activation, planning of utterance, processing of utterances (garden path effects) and similar things. In fact, none of the theories that are discussed in this book contains an explicit theory that explains all these things. I think that it is perfectly legitimate to work in this way: it is legitimate to study the structure of words without studying their semantics and pragmatics, it is legitimate to study phonology without caring about syntax, it is legitimate to deal with specific semantic problems without caring about phonology and so on, provided there are ways to integrate the results of such research into a bigger picture. So, it is wrong to develop models like those developed in current versions of Minimalism (called Biolinguistics), where it is assumed that utterances are derived in phases (NPs, CPs, depending on the variant of the theory) and then shipped to the interfaces (spell out and semantic interpretation). This is not what humans do (see Chapter 15). But if we are neutral with respect towards such issues, we are fine. In fact, there is psycholinguistic work that couples HPSG grammars to performance models (Konieczny 1996) and similar work exists for TAG (Shieber & Johnson 1993; Demberg & Keller 2008).

Finally, there is also work in Construction Grammar that abstracts away from performance considerations. For instance, Adele Goldberg's book from 1995 does not contain a worked out theory of performance facts. It contains boxes in which grammatical functions are related to semantic roles. So this basically is a competence theory as well. Of course there are statements about how this is connected to psycholinguistic findings, but this is also true for theories like HPSG, SBCG and Simpler Syntax (Jackendoff 2011: 600) that explicitly make the competence/performance distinction.

10.6.4.5.2 Mathematical formalization vs. implementation

The difference between mathematical and computational formalization is a rather strange distinction to make. I think that a formal and precise description is a prerequisite for implementation (see the discussion in Section 3.6.2 and Section 4.7.2). Apart from this, a computer implementation of SBCG is trivial, given the systems that we have for processing HPSG grammars. In order to show this, I want to address one issue that van Trijp discusses. He claims that SBCG cannot be directly implemented. On issues of complexity of constraint solving systems he quotes (Levine & Meurers 2006: Section 4.2.2):

> Actual implementations of HPSG typically handle the problem by guiding the linguistic processor using a (rule-based) phrase structure backbone, but the disadvantage of this approach is that the "organization and formulation of the grammar is different from that of the linguistic theory" (Levine & Meurers 2006: Section 4.2.2). (van Trijp 2013)

He concludes:

> Applying all these observations to the operationalization of SBCG, we can conclude that an SBCG grammar is certainly amenable for computational implementation because of its formal explicitness. There are at least two computational platforms available, mostly used for implementing HPSG-based grammars, whose basic tenets are compatible with the foundations of SBCG: LKB (Copestake 2002) and TRALE (Richter 2006). However, none of these platforms supports a 'direct' implementation of an SBCG grammar as a general constraint system, so SBCG's performance-independence hypothesis remains conjecture until proven otherwise.

There are two issues that should be kept apart here: efficiency and faithfulness to the theory. First, as Levine and Meurers point out, there were many constraint solving systems at the beginning of the 90's. So there are computer systems that can and have been used to implement and process HPSG grammars. This is very valuable since they can be used for direct verification of specific theoretical proposals. As was discussed by Levine and Meurers, trying to solve constraints without any guidance is not the most efficient way to deal with the parsing/generation problem. Therefore, additional control-structure was added. This control structure is used for instance in a parser to determine the syntactic structure of a phrase and other constraints will apply as soon as there is sufficient information available for them to apply. For instance, the assignment of structural case happens once the arguments of a head are realized. Now, is it bad to have a phrase structure backbone? One can write down phrase structure grammars that use phrase structure rules that have nothing to do with what HPSG grammars usually do. The systems TRALE (Meurers, Penn & Richter 2002; Penn 2004) and LKB will process them. But one is not forced to do this. For instance, the grammars that I developed for the CoreGram project (Müller 2013a, 2015a) are very close to the linguistic theory. To see that this is really the case, let us look at the Head-Argument Schema. The Head-Argument Schema is basically the type *head-argument-phrase* with certain type constraints that are partly inherited from its supertypes. The type with all the constraints was given on page 274 and is repeated here as (80):

(80) (syntactic) constraints on *head-argument-phrase*:

$$\begin{bmatrix} \textsc{synsem}|\textsc{loc}|\textsc{cat} \begin{bmatrix} \textsc{head} & \boxed{1} \\ \textsc{subcat} & \boxed{2} \end{bmatrix} \\ \textsc{head-dtr}|\textsc{synsem}|\textsc{loc}|\textsc{cat} \begin{bmatrix} \textsc{head} & \boxed{1} \\ \textsc{subcat} & \boxed{2} \oplus \langle \boxed{3} \rangle \end{bmatrix} \\ \textsc{non-head-dtrs} \ \langle \ [\ \textsc{synsem}\ \boxed{3}\ ] \ \rangle \\ \textit{head-argument-phrase} \end{bmatrix}$$

This can be translated into a phrase structure grammar in a straight-forward way:

(81) a.
$$\begin{bmatrix} \textsc{synsem}|\textsc{loc}|\textsc{cat} \begin{bmatrix} \textsc{head} & \boxed{1} \\ \textsc{subcat} & \boxed{2} \end{bmatrix} \\ \textsc{head-dtr}\ \boxed{4}\ |\textsc{synsem}|\textsc{loc}|\textsc{cat} \begin{bmatrix} \textsc{head} & \boxed{1} \\ \textsc{subcat} & \boxed{2} \oplus \langle \boxed{3} \rangle \end{bmatrix} \\ \textsc{non-head-dtrs} \ \langle \ \boxed{5}\ [\ \textsc{synsem}\ \boxed{3}\ ] \ \rangle \\ \textit{head-argument-phrase} \end{bmatrix} \rightarrow \boxed{4}, \boxed{5}$$

b.
$$\begin{bmatrix} \textsc{synsem}|\textsc{loc}|\textsc{cat} \begin{bmatrix} \textsc{head} & \boxed{1} \\ \textsc{subcat} & \boxed{2} \end{bmatrix} \\ \textsc{head-dtr}\ \boxed{4}\ |\textsc{synsem}|\textsc{loc}|\textsc{cat} \begin{bmatrix} \textsc{head} & \boxed{1} \\ \textsc{subcat} & \boxed{2} \oplus \langle \boxed{3} \rangle \end{bmatrix} \\ \textsc{non-head-dtrs} \ \langle \ \boxed{5}\ [\ \textsc{synsem}\ \boxed{3}\ ] \ \rangle \\ \textit{head-argument-phrase} \end{bmatrix} \rightarrow \boxed{5}, \boxed{4}$$

The left hand side of the rule is the mother node of the tree, that is, the sign that is licensed by the schema provided that the daughters are present. The right hand side in (81a) consists of the head daughter $\boxed{4}$ followed by the non-head daughter $\boxed{5}$. We have the opposite order in (81b), that is, the head daughter follows the non-head daughter. The two orders correspond to the two orders that are permitted by LP-rules: the head precedes its argument if it is marked INITIAL+ and it follows it if it is marked INITIAL−.

The following code shows how (82b) is implemented in TRALE:

```
arg_h rule (head_argument_phrase,
          synsem:loc:cat:head:initial:minus,
          head_dtr:HeadDtr,
          non_head_dtrs:[NonHeadDtr]
         )
  ===>
cat> NonHeadDtr,
cat> HeadDtr.
```

A rule starts with an identifier that is needed for technical reasons like displaying intermediate structures in the parsing process in debugging tools. A description of the mother node follows and after the arrow we find a list of daughters, each introduced by the operator `cat>`.[31] Structure sharing is indicated by values with capital letters. The above TRALE rule is a computer-readable variant of (81b), but includes the explicit specification of the value of INITIAL.

Now, the translation of a parallel schema using a MOTHER feature like (82a) into a phrase structure rule is almost as simple:

(82) a.
$$\begin{bmatrix} \text{MOTHER}|\text{SYNSEM}|\text{LOC}|\text{CAT} \begin{bmatrix}\text{HEAD} & \boxed{1}\\ \text{SUBCAT} & \boxed{2}\end{bmatrix} \\ \text{HEAD-DTR}\ \boxed{4}\ |\text{SYNSEM}|\text{LOC}|\text{CAT} \begin{bmatrix}\text{HEAD} & \boxed{1}\\ \text{SUBCAT} & \boxed{2} \oplus \langle \boxed{3} \rangle\end{bmatrix} \\ \text{NON-HEAD-DTRS}\ \langle \boxed{5}\ [\ \text{SYNSEM}\ \boxed{3}\ ]\ \rangle \\ \textit{head-argument-cx}\end{bmatrix}$$

b.
$$\boxed{6} \rightarrow \boxed{4}, \boxed{5} \text{ where } \begin{bmatrix} \text{MOTHER}\ \boxed{6}\ |\text{SYNSEM}|\text{LOC}|\text{CAT} \begin{bmatrix}\text{HEAD} & \boxed{1}\\ \text{SUBCAT} & \boxed{2}\end{bmatrix} \\ \text{HEAD-DTR}\ \boxed{4}\ |\text{SYNSEM}|\text{LOC}|\text{CAT} \begin{bmatrix}\text{HEAD} & \boxed{1}\\ \text{SUBCAT} & \boxed{2} \oplus \langle \boxed{3} \rangle\end{bmatrix} \\ \text{NON-HEAD-DTRS}\ \langle \boxed{5}\ [\ \text{SYNSEM}\ \boxed{3}\ ]\ \rangle \\ \textit{head-argument-cx}\end{bmatrix}$$

(82b) is only one of the two phrase structure rules that correspond to (82a), but since the other one only differs from (82b) in the ordering of $\boxed{4}$ and $\boxed{5}$, it is not given here.

For grammars in which the order of the elements corresponds to the observable order of the daughters in a DTRS list, the connection to phrase structure rules is even simpler:

(83)
$$\boxed{1} \rightarrow \boxed{2} \text{ where } \begin{bmatrix} \text{MOTHER} & \boxed{1} \\ \text{DTRS} & \boxed{2} \\ \textit{construction} & \end{bmatrix}$$

The value of DTRS is a list and hence $\boxed{2}$ stands for the list of daughters on the right hand side of the phrase structure rule as well. The type *construction* is a supertype of all constructions and hence (83) can be used to analyze all phrases that are licensed by the grammar. In fact, (83) is one way to put the meta constraint in (31).

So, this shows that the version of SBCG that has been developed by Sag (2012) has a straightforward implementation in TRALE.[32] The question remains whether *SBCG's performance-independence hypothesis remains conjecture until proven otherwise* as van Trijp sees it. The answer is: it is not a conjecture since any of the old constraint-solving

[31] Other operators are possible in TRALE. For instance, `sem_head` can be used to guide the generator. This is control information that has nothing to do with linguistic theory and not necessarily with the way humans process natural language. There is also a `cats` operator, which precedes lists of daughters. This can be used to implement flat phrase structures.

[32] A toy fragment of English using a MOTHER feature and phrase structure rules with specifications of the kind given above can be downloaded at http://hpsg.fu-berlin.de/Fragments/SBCG-TRALE/.

systems of the nineties could be used to process SBCG. The question whether this is efficient is an engineering problem that is entirely irrelevant for theoretical linguistics. Theoretical linguistics is concerned with human languages and how they are processed by humans. So whether some processing system that does not make any claims about human language processing is efficient or not is absolutely irrelevant. Phrase structure-based backbones are therefore irrelevant as well, provided they refer to the grammar as described in theoretical work.

Now, this begs the question whether there is a contradiction in my claims. On page 329 I pointed out that SBCG is lacking a formalization in Richter's framework (Richter 2004). Richter and also Levine & Meurers (2006) pointed out that there are problems with certain theoretically possible expressions and it is these expressions that mathematical linguists care about. So the goal is to be sure that any HPSG grammar has a meaning and that it is clear what it is. Therefore, this goal is much more foundational than writing a single grammar for a particular fragment of a language. There is no such foundational work for FCG since FCG is a specific toolkit that has been used to implement a set of grammars.

10.6.4.5.3 Static constraints vs. dynamic mappings and signature and grammar vs. open-endedness

The cool thing about Fluid Construction Grammar is its fluidity, that is there are certain constraints that can be adapted if there is pressure, the inventory of the theory is open-ended, so categories and features can be added if need be.

Again, this is not a fundamental difference between HPSG/SBCG and FCG. An HPSG grammar fragment of a specific language is a declarative representation of linguistic knowledge and as such it of course just represents a certain fragment and does not contain any information how this set of constraints evolved or how it is acquired by speakers. For this we need specific theories about language evolution/language change/language acquisition. This is parallel to what we said about the competence/performance distinction, in order to account for language evolution we would have to have several HPSG grammars and say something about how one developed from the other. This will involve weighted constraints, it will involve recategorization of linguistic items and lots more.[33] So basically HPSG has to be extended, has to be paired with a model about language evolution in the very same way as FCG is.

10.6.4.5.4 Theoretical physics vs. Darwinian evolutionary theory

Van Trijp compares SBCG and FCG and claims that SBCG follows the model of theoretical physics – like Chomsky does, while FCG adopts a Darwinian model of science –

[33] There are systems that use weighted constraints. We had a simple version of this in the German HPSG grammar that was developed in Verb*mobil* project (Müller & Kasper 2000) already. Further theoretical approaches to integrate weighted constraints are Brew (1995) and more recently Guzmán Naranjo (2015). Usually such weighted constraints are not part of theoretical papers, but there are exceptions as for instance the paper by Briscoe and Copestake about lexical rules (Briscoe & Copestake 1999).

like Croft does, the difference being that SBCG makes certain assumptions that are true of all languages, while FCG does not make any a priori assumptions. The fundamental assumptions made in both theories are that the objects that we model are best described by feature value pairs (a triviality). FCG assumes that there is always a syntactic and a semantic pole (fundamental assumption in the system) and researchers working in HPSG/SBCG assume that if languages have certain phenomena, they will be analyzed in similar ways. For instance, if a language has nonlocal dependencies, these will be analyzed via the SLASH mechanism. However, this does not entail that one believes that grammars of all languages have a SLASH feature. And in fact, there may even be languages that do not have valence features (Koenig & Michelson 2010), which may be a problem for FCG since it relies on the SYN-pole for the matching phase. So as far as SBCG is concerned, there is considerable freedom to choose features that are relevant in an analysis, and of course additional features and types can be assumed in case a language is found that provides evidence for this. The only example of a constraint provided by van Trijp that is possibly too strong is the locality constraint imposed by the MOTHER feature. The idea about this feature is that everything that is of relevance in a more nonlocal context has to be passed up explicitly. This is done for nonlocal dependencies (via SLASH) and for instance also for information concerning the form of a preposition inside of a PP (via PFORM). Certain verbs require prepositional objects and restrict the form of the preposition. For instance, *wait* has to make sure that its prepositional object has the preposition *for* in it. Since this information is usually available only at the preposition, it has to be passed up to the PP level in order to be directly selectable by the governing verb.

(84) I am waiting for my man.

So, assuming strict locality of selection requires that all phenomena that cannot be treated locally have to be analyzed by passing information up. Assuming strict locality is a design decision that does not have any empirical consequences, as far as it does not rule out any language or construction in principle. It just requires that information has to be passed up that needs to be accessed at higher nodes. As I have shown in Section 10.6.2, the locality constraint is easily circumvented even within SBCG and it makes the analysis of idioms unnecessarily complicated and unintuitive, so I suggest dropping the MOTHER feature. But even if MOTHER is kept, it is not justified to draw a distinction between SBCG and FCG along the lines suggested by van Trijp.

Independent of the MOTHER issue, the work done in the CoreGram project (Müller 2013a, 2015a) shows that one can derive generalizations in a bottom-up fashion rather than imposing constraints on grammars in a top-down way. The latter paper discusses Croft's methodological considerations and shows how methodological pitfalls are circumvented in the project. HPSG/SBCG research differs from work in Chomskyan frameworks in not trying to show that all languages are like English or Romance or German or whatever, rather languages are treated on their own as it is common in the Construction Grammar community. This does not imply that there is no interest in generalizations and universals or near universals or tendencies, but again the style of working and the rhetoric in HPSG/SBCG is usually different from the ones in Mainstream Generative

Grammar. Therefore, I think that the purported difference between SBCG and FCG does not exist.

10.6.4.5.5 Permissiveness of the theories

Van Trijp claims that HPSG/SBCG is a "generative grammar" since its aim is to account for and admit only grammatical sentences. FCG on the other hand is more permissive and tries to get the most out of the input even if it is fragmentary or ungrammatical (see also Steels 2013: 166). While it is an engineering decision to be able to parse ungrammatical input – and there are most certainly systems for the robust processing of HPSG grammars (Kiefer, Krieger & Nederhof 2000; Copestake 2007), it is also clear that humans cannot parse everything. There are strong constraints whose violations cause measurable effects in the brain. This is something that a model of language (that includes competence and performance factors or does not make the difference at all) has to explain. The question is what the cause of deviance is: is it processing complexity? Is it a category mismatch? A clash in information structure? So, if FCG permits structures that are not accepted by human native speakers and that do not make any sense whatsoever, additional constraints have to be added. If they are not added, the respective FCG theory is not an adequate theory of the language under consideration. Again, there is no difference between HPSG/SBCG and FCG.

10.6.4.5.6 A note on engineering

My biggest problem with FCG is that linguistic and engineering aspects are mixed.[34] Certain bookkeeping features that are needed only for technical reasons appear in linguistic papers, technical assumptions that are made to get a parser running are mixed with linguistic constraints. Bit vector encodings that are used to represent case information are part of papers about interesting case systems. There is certainly nothing wrong with bit vector encodings. They are used in HPSG implementations as well (Reape 1991: 55; Müller 1996c: 269), but this is not mixed into the theoretical papers.

It was a big breakthrough in the 80's when theoretical linguists and computational linguists started working together and developed declarative formalisms that were independent of specific parsers and processing systems. This made it possible to take over insights from a lot of linguists who were not concerned with the actual implementation but took care of finding linguistic generalizations and specifying constraints. Since this separation is given up in FCG, it will remain an engineering project without much appeal to the general linguist.

[34] This is not a problem if all FCG papers are read as papers documenting the FCG-system (see Footnote 26 on page 339) since then it would be necessary to include these technical details. If the FCG papers are to be read as theoretical linguistics papers that document a certain Construction Grammar analysis, the Lisp statements and the implementational details are simply an obstacle.

10.7 Summary and classification

There are currently three formalized variants of Construction Grammar: Sign-Based Construction Grammar, Embodied Construction Grammar, and Fluid Construction Grammar. The first two variants can be viewed as notational variants of (Constructional) HPSG (for SBCG with regard to this point, see Sag (2007: 411) and Sag (2010: 486)), or put differently, sister theories of HPSG. This is also true to a large extend for FCG, although van Trijp (2013) spends 25 pages working out the alleged differences. As I have shown in Section 10.6.4, HPSG and FCG are rather similar and I would say that these theories are sister theories as well.

Due to the origins of all three theories, respective analyses can differ quite considerably: HPSG is a strongly lexicalized theory, where phrasal dominance schemata have only been increasingly more used in the last ten years under the influence of Ivan Sag. The phrasal dominance schemata that Ivan Sag uses in his work are basically refinements of schemata that were present in earlier versions of HPSG. Crucially, all phenomena that interact with valence receive a lexical analysis (Sag, Boas & Kay 2012: Section 2.3). In CxG, on the other hand, predominantly phrasal analyses are adopted due to the influence of Adele Goldberg.

As already emphasized in Chapter 9, these are only tendencies that do not apply to all researchers working in the theories in question.

Exercises

1. Find three examples of utterances whose meaning cannot be derived from the meaning of the individual words. Consider how one could analyze these examples in Categorial Grammar (yes, Categorial Grammar).

Further reading

There are two volumes on Construction Grammar in German: Fischer & Stefanowitsch (2006) and Stefanowitsch & Fischer (2008). Deppermann (2006) discusses Construction Grammar from the point of view of conversational analysis. The 37(3) volume of the *Zeitschrift für germanistische Linguistik* from 2009 was also devoted to Construction Grammar. Goldberg (2003a) and Michaelis (2006) are overview articles in English. Goldberg's books constitute important contributions to Construction Grammar (1995; 2006; 2009). Goldberg (1995) has argued against lexical analyses such as those common in GB, LFG, CG, HPSG, and DG. These arguments can be invalidated, however, as will be shown in Section 21.7.1. Sag (1997), Borsley (2006), Jacobs (2008) and Müller & Lipenkova (2009) give examples of constructions that require a phrasal analysis if one wishes to avoid postulating empty elements. Jackendoff (2008) discusses the noun-preposition-noun construction that can only be properly analyzed as a phrasal construction (see Section 21.10). The discussion on whether argument structure constructions should be analyzed phrasally or lexically (Goldberg 1995, 2006; Müller 2006) culminated in a series of papers (Goldberg 2013a) and a target article by Müller & Wechsler (2014a) with several responses in the same volume.

Tomasello's publications on language acquisition (Tomasello 2000, 2003, 2005, 2006c) constitute a Construction Grammar alternative to the Principle & Parameters theory of acquisition as it does not have many of the problems that P&P analyses have (for more on language acquisition, see Chapter 16). For more on language acquisition and Construction Grammar, see Behrens (2009).

Dąbrowska (2004) looks at psycholinguistic constraints for possible grammatical theories.