Phraseological clauses in Constructional HPSG

Frank Richter ⓘD

University of Tübingen

Manfred Sailer ⓘD

University of Göttingen

Abstract

In this paper we investigate German idioms which contain phraseologically fixed clauses (PCl). To provide a comprehensive HPSG theory of PCls we extend the idiom theory of Soehn (2006) in such a way that it can distinguish different degrees of regularity in idiomatic expressions. An in-depth analysis of two characteristic PCls shows how our two-dimensional theory of idiomatic expressions can be applied and illustrates the scope of the theory.

1   Introduction

The literature on idioms often focuses on VP idioms such as *kick the bucket* or *spill the beans*, where a particular verbal lexeme combines with a particular NP or PP complement. These combinations show different degrees of flexibility. Hardly any attention has been paid to idioms which comprise complete clauses. Idioms with phraseological clauses are mentioned in passim in phraseological studies such as Fleischer (1997) but have never been in the focus of empirical studies, or of detailed theoretical discussions. As clausal parts of idioms are structurally more complex than NPs or PPs, they are ideally suited for investigating a greater range of structural and semantic variation in idiomatic expressions.

In this paper we will look at phraseologically fixed clauses (PCl) in German. The discussion of PCls is particularly interesting in light of attempts to combine aspects of Construction Grammar with HPSG. One of the important insights of Construction Grammar is that constructions may span more than a local tree. This contrasts with the lexical nature of HPSG and its historical ties to context-free phrase structure grammars. Another result of phraseological research is the insight that idioms are not all of the same kind. In the context of PCls we will want to distinguish between decomposable and non-decomposable idioms (Wasow et al., 1983), and between grammatical and extra-grammatical idioms (Fillmore et al., 1988).

We start with a presentation of the empirical properties of German PCls (Section 2). In order to get the necessary theoretical tools for their analysis, we extend an existing approach to idioms in HPSG to be able to distinguish in our theory between different types of idioms (Section 3). In Section 4 the theory is applied to the PCl data. Section 5 contains a short comparison to an alternative attempt to formalize basic ideas of Construction Grammar in an HPSG-inspired framework. A short summary and conclusion are given at the end of the paper.

2   Data

In (1) and (2) we list idioms with phraseological clauses (PCl). Each PCl in (1) combines with a particular verb or a small group of verbs. Their behavior resembles

298

the behavior of the word *headway* in the English expression *make headway*, i.e. they act as a complement in a VP idiom where both the verb and the complement are part of the idiom. The PCls are declarative clauses ((1-c), (1-f)), interrogative clauses ((1-a), (1-b), (1-d)), and a free relative clause in (1-e). The PCls in (2) are adjunct clauses.

(1)   PCl is a complement clause to one or a small group of verbs

    a.   wissen, wo   Barthel den Most   holt
        know   where Barthel the young wine gets
        ('know every trick in the book')

    b.   (nicht) wissen, wo   X_dat der Kopf steht
        not   know   where X   the head stands
        ('have a lot of stress')

    c.   glauben, X_acc tritt   ein Pferd
        believe X   kicks a   horse
        ('be very surprised')

    d.   wissen, wo   (X_acc/dat) der Schuh drückt
        know   where X   the shoe   presses
        ('know what is worrying X')

    e.   hingehen/ bleiben (sollen), wo   der Pfeffer wächst
        go/   stay   (should) where the pepper grows
        ('go/ stay away')

    f.   glauben, X's Schwein pfeift
        believe   X's pig   whistles
        ('be very surprised')

(2)   PCl is an adjunct

    a.   bis der Arzt   kommt
        until the doctor arrives
        ('ad nauseam')

    b.   wenn Ostern und Pfingsten auf einen/ denselben Tag fallen
        when Eastern and Pentecost on one/   the same   day fall
        ('never')

    c.   aussehen, als   hätten X_dat die Hühner das Brot weggefressen
        look   as if had   X   the chicken the bread eaten away
        ('look stupefied')

    d.   wie Gott X_acc geschaffen hat
        as god X   created   has
        ('naked')

Apart from their idiomatic semantics, the PCls have the structural properties of regular German sentences. On closer scrutiny, they display an interesting continuum of grammatical and lexical fixedness and flexibility.

In (1-b), (1-c), (1-f), (2-c), and (2-d) the constituent marked with X is anaphoric to the matrix subject. In (1-d) the constituent marked with X is optional and need not be anaphoric to the matrix subject.

(3)   Ich möchte wissen, wo (dich/dir) der Schuh drückt.
        (lit.: I want to know where the shoe presses you)

PCls permit a certain degree of grammatical variation. Speakers of some German dialects prefer to use proper nouns with definite articles. These speakers use a variant of (1-a) with a PCl subject of the form *der Barthel* (*the Barthel*). Similarly, *until*-clauses in German may optionally contain an overt complementizer *dass* (*that*). Indeed, a variant of (2-a) with an overt complementizer is attested, i.e. *bis dass der Arzt kommt* (*until that the doctor arrives*).

However, not just any grammatical variation is permitted. Let us consider the idiom in (1-b). Outside of idiomatic phrases a combination of a possessive dative NP and a definite NP can be freely replaced with a construction with the same dative NP and a definite NP that contains a possessive determiner. The possessor is then coreferential with the dative NP. The pattern is illustrated in (4-a). This otherwise systematic variation is not possible with the idiom. We use "#" to indicate the non-availability of an idiomatic interpretation. The same alternation is also excluded for (2-c).

(4)   a.   Ich habe Peter den/seinen Kopf verbunden.
              (lit: 'I bandaged Peter the/his head')
        b.   Peter weiß nicht, wo ihm der/#sein Kopf steht.

Another systematic variation is the active-passive alternation. None of the PCls with a transitive verb in (1) allow a passive in their idiomatic meaning.

(5)   a.   #wissen, wo vom Barthel der Most geholt wird (passive of (1-a))
        b.   #wissen, wo X vom Schuh gedrückt wird (passive of (1-d))
        c.   #glauben, X wird von einem Pferd getreten (passive of (1-c))

Finally, the PCl in (1-c) is a verb-second clause. In free uses, we can find two kinds of alternation. First, verb-second complement clauses alternate with verb-final complement clauses. Second, any constituent of the clause can occur as the first constituent (in the *Vorfeld*) in verb-second clauses without a change in meaning. Both types of grammatical alternation are excluded in (1-c).

(6)   a.   #Ich glaube, dass mich ein Pferd tritt. (*dass*-clause)
        b.   #Ich glaube, ein Pferd tritt mich. (different first constituent)

All alternations discussed in (4)–(6) are neutral with respect to the truth-conditional semantics of the literal reading of the PCls, but some of the alternations influence the information structure. Among the latter kind are valence alternations, clause type alternation, and constituent fronting. This subclass of alternations is impossible with PCls. The permitted alternations, viz. *dass* insertion and the insertion of a definite determiner in front of a proper name, do not affect information structure.

Let us, next, turn to variation at the level of changing or adding lexical material. Some lexical variation is clearly permitted. In (2-b) the holidays can be changed, the subject may be any combination of Easter, Pentecost, and Christmas. However, some form of the verb *(zusammen-) fallen* is obligatory.

(7)   #wenn Ostern und Pfingsten auf demselben Tag liegen/ zu liegen kommen/ am selben Tag sind.

We also find some variation with respect to the matrix predicate that the PCl occurs with. The expression in (2-c) may combine with any matrix predicate that describes someone's facial expression or someone's appearance. The same variation is also found with other PCls that express a similar content. They all are comments on the way the referent of the matrix subject looks.

(8)   aussehen/ ein Gesicht machen/ dastehen,
      look/   a   face   make/   appear   …
      a.   als hätten X die Hühner das Brot weggefressen. (=(2-c))
      b.   als hätte es X die Ernte verhagelt
           lit. look as if X's harvest was destroyed by hail.
      c.   als hätte X ein Lineal verschluckt.
           lit. as if X had swallowed a ruler
      d.   wie eine Kuh, wenn's donnert.
           lit. like a cow when it thunders

There is also some systematic variation in the matrix predicates that express the idea of "thinking". All PCls that occur with *glauben* (*believe*) in the present tense also marginally accept *denken* (*think*) in the present tense. In past tense, however, they systematically prefer *denken*.

(9)   Ich glaube/   ?denke/   ?*glaubte/   dachte   …
      I   believe.PRES/ think.PRES/ believed.PAST/ thought.PAST
      a.   mich tritt ein Pferd. (=(1-c))
      b.   mein Hamster bohnert. (lit.: my hamster is polishing the floor)
      c.   ich steh im Wald. (lit.: I am standing in the woods)

In addition to the variation of obligatory material, some PCls may host more lexical or semantic material. For example, there is variation in the tense form of some but not all PCls.

(10)   temporally flexible idioms
       a.   Ich hab damals Tetris gespielt, bis der Arzt gekommen ist.
            (pres. perfect of (2-a))
            ('I used to play Tetris ad nausam')
            (www.stern.de/digital/computer/scheibe/
            scheibes-kolumne-pc-nostalgie-619141.html, 14.10.2009.)
       b.   Er wusste nicht, wo ihm der Kopf stand. (simple past of (1-b))

(11) temporally fixed idioms

    a. #Sie hat nicht gewusst, wo Barthel den Most geholt hat. (pres. perf. of (1-a))

    b. #Ich glaube, mein Schwein hat gepfiffen. (pres. perf. of (1-f))

Similarly, modals are allowed in some but not all of the PCls.

(12) modally flexible idioms

    a. Hudezeck versteht sich auf die Kunst, die Lachmuskeln so zu stra-pazieren, bis der Arzt kommen muss. ((2-a) with *must*)
        (www.fnp.de/fnp/mobil/rmn01.c.5440589.de.htm, 14.10.2009.)

    b. Als Reiseleiter ist Terje ein Mann der Praxis und weiß, wann und wo auf Reisen der Schuh drücken könnte. ((1-d) with *could*)
        (www.skantur.de/allgemein/award.htm, 14.10.2009.)

(13) modally fixed idioms

    a. #Peter soll bleiben, wo der Pfeffer wachsen kann. ((1-e) with *should*)

    b. #Ich glaube, mein Schwein könnte pfeifen. ((1-f) with *could*)

PCls do not tolerate negation (see (14)), but non-truth-conditional modifiers such as *eigentlich* (*actually*), *sprichwörtlich* (*proverbial*) can usually be added (see (15)).

(14) a. #Peter weiß, wo ihn der Schuh nicht drückt. ((1-d) with negation)

    b. #Peter weiß, wo Barthel den Most nicht holt. ((1-a) with negation)

    c. #wenn Ostern und Pfingsten nicht auf einen Tag fallen ((2-b) with negation)

(15) a. Peter weiß nicht mehr, wo ihm eigentlich der Kopf steht. ((1-b) with *actually*)

    b. Martha weiß, wo Barthel den sprichwörtlichen Most holt. ((1-a) with *proverbial*)

Focus sensitive particles such as *auch* (*as well*) and *selbst* (*even*) are allowed if they combine with lexically free slots of the PCls but not when combining with lexically fixed constituents. This is illustrated in (16).

(16) a. Peter weiß, wo [selbst ihn]/ [auch ihn] der Schuh drückt.
       Peter knows where even him/ him as well him the shoe presses
       ('Peter knows what is worrying even him/ him as well.')

    b. Peter weiß, wo ihn #[auch/ selbst der Schuh] drückt.

It is possible to add a negation to the expression in (1-d) if the negation is inside the embedded free slot.

(17) Peter weiß, wo [nicht ihn sondern den Hans] der Schuh drückt.
       Peter knows where not him but the Hans the shoe presses

('Peter knows what is worrying not him but Hans.')

The following picture emerges from our inspection of the properties of PCls: First, the information structure of the literal meaning must not be changed. This implies that the application of valence alternating operations is excluded (passive, possessive dative shift) and so are changes with respect to the topicalized constituent. Second, the propositional semantic core of the literal meaning must remain constant. In some restricted cases it may be modified by modal and temporal expressions. Third, syntactic and semantic alternations are possible if they concern free slots in the PCl or if they do not affect the information structure or the core propositional semantics of the PCl. Fourth, obligatory anaphoric relations may hold between elements inside the PCl and matrix elements.

The properties of PCls show that they cannot be treated as big "words with spaces". Instead, they are inherently complex syntactic units with different degrees of flexibility. This is parallel to what was observed for other idioms in Wasow et al. (1983) and elsewhere, and clearly sets PCls apart from fully fixed forms such as proverbs.

3 The Two-Dimensional Theory of Idioms

In this section we propose an extension of the theory of irregularity developed in Richter and Sailer (2003), Sailer (2003), and Soehn (2006). After summarizing the most important parts of that theory in Section 3.1, we will significantly extend it in Section 3.2 to capture the different degrees of regularity found in idiomatic constructions in a straightforward way. As PCls are characterized by a high degree of syntactic regularity, this extension is particularly important for a systematic analysis of PCls.

3.1 Internal and External Idiosyncrasies

The two-dimensional theory of idioms builds on the distinction between *decomposable* and *non-decomposable* idioms in Wasow et al. (1983). In our terminology, decomposable idioms will be treated as combinations of words with *external* irregularities, non-decomposable idioms are analyzed as phrases whose structures are *internally* irregular.

Decomposable idioms comprise expressions such as *make waves* (*cause trouble*) and *spill the beans* (*divulge information*). They show a considerable degree of syntactic and semantic flexibility. Following Wasow et al. (1983) and Gazdar et al. (1985) we treat them as syntactically free combinations of words with an idiom-specific meaning: In its idiomatic use the word *spill* is synonymous to *divulge* and the word *beans* means *information*. Our theory forces the two idiom-specific meaning variants of *spill* and *beans* to co-occur obligatorily. Under this perspective the expression *spill the beans* is not idiosyncratic with respect to the way it is put together from its syntactic and semantic components. What is special about it is that

it contains two words with a highly restricted distribution, the idiomatic variants of *spill* and *beans*. For this reason we regard the idiomatic variants of these words as *distributionally* or *externally* idiosyncratic.

To implement the idea of distributional restrictions formally, two new attributes have been introduced into the grammar architecture. First, Soehn (2004) proposes the feature LISTEME.[1] Its value is a unique atomic identifier for each item that is listed in the lexicon. This feature allows us to distinguish between the word *spill* in the meaning *divulge*, used in the idiom *spill the beans*, and the non-idiomatic word *spill*. Second, the list-valued feature COLL (context of lexical licensing) is defined on the sort *sign*. For every lexical item, the value of COLL is a non-empty list of objects of sort *barrier*. These barrier objects specify two things: (i) a syntactic domain (in which they apply), and (ii) a requested licensing property. A general principle guarantees that in each structure, for each lexical item in this structure, and for each *barrier* object on the COLL list of each lexical item, the licensing property holds in the syntactic domain specified in the *barrier* object. To take a concrete example, the lexical specification of the idiomatic word *beans* introduces a COLL list with one *barrier* object on it. The barrier demands that the idiomatic word *beans* co-occur with a verb whose LISTEME value is *divulge-spill* and that the word *beans* act as the theme argument of that verb. Since the idiom *spill the beans* is very flexible syntactically, the licensing domain of our example is the entire utterance in which the idiomatic word *beans* occurs.[2]

The external dimension of idiosyncrasy is complemented by a dimension of *internal* idiosyncrasy. Non-decomposable idioms, such as *saw logs*, are syntactically and semantically frozen. This type of idioms may even have an otherwise unattested syntactic structure. Wasow et al. (1983) mention, among others, *kingdom come* and *trip the light fantastic* as extreme cases. An analysis in terms of fixed phrasal expressions appears to be most appropriate for these expressions. An expression such as *saw logs* is stored in the lexicon as a complex phrase with fixed syntactic structure and the idiosyncratic meaning *snore*.

The theory of idioms in Sailer (2003) and Soehn (2004, 2006) offers an encoding of such internally idiosyncratic phrases in HPSG that directly exploits the COLL feature. If a phrase has the specification [COLL *elist*], it is regular, or non-idiomatic. In this case the rules of regular syntactic and semantic combinatorics apply to it. An internally irregular phrase has the specification [COLL *nelist*] and is exempt. In this architecture, the lexicon contains the descriptions of words and of internally idiosyncratic phrases. The lexical entries of the latter are called *phrasal lexical entries* (PLE). The PLE for the expression *kingdom come* in (18) provides a simple example.

---

[1]For most purposes, this feature corresponds to the feature LEXICAL-ID (LID) in Sag (2007b). In contrast to Soehn (and Sag), we assume that LISTEME is not a head feature. See Section 3.2 for our motivation.

[2]The exact formalization of the collocational mechanism is not relevant in our present discussion. See Soehn (2004, 2006) for details.

(18)   The phrasal lexical entry of *kingdom come*:

$$
\begin{bmatrix}
phrase \\
\text{PHON } \boxed{1} \oplus \boxed{2} \\
\text{SS L}
\begin{bmatrix}
\text{CAT LISTEME } & kingdom\text{-}come \\
\text{CONT MAIN } & paradise
\end{bmatrix} \\
\text{DTRS }\Big[\text{N-DTRS } \big\langle \big[\text{PHON } \boxed{1}\ \langle kingdom\rangle\big], \big[\text{PHON } \boxed{2}\ \langle come\rangle\big]\big\rangle\Big] \\
\text{COLL } nelist
\end{bmatrix}
$$

The phrase in (18) cannot be a regular phrase of English since its semantics is not derived regularly from the semantics of its constituents. The irregularity is possible because the specification [COLL *nelist*] exempts the phrase from the principles of combinatorial semantics. With respect to the internal syntactic structure, there is no identifiable syntactic head.[3] For this reason we specify it as consisting of two non-head daughters, but leave the details of the syntactic combination underspecified. This accounts for the fact that it is not clear what exactly the syntactic relation is between the two words *kingdom* and *come*.

The analysis accounts for the irregularity of the expression, and it also captures the idea that the two words in it are regular members of the English lexicon. This is a direct consequence of merely mentioning the phonological values of the daughters without restricting their syntactic or semantic structure in any other way. The grammar must independently license signs with the required phonological properties. This requirement can be met by the regular words *kingdom* and *come*. This mechanism may seem trivial at first, but it illustrates a principled locality restriction on idiosyncratic phrases: A phrasal lexical entry can only locally license idiosyncratic properties of a phrasal node, but it cannot introduce idiosyncratic properties at its daughters or at any deeper level of syntactic embedding.[4]

Thus far it may seem that the central attribute, COLL, is used for two unrelated and independent purposes: It encodes idiosyncratic distributional requirements, and it specifies whether a phrase is internally irregular. However, these two dimensions of irregularity are in fact related. Inspired by the idea that all lexical elements may enter into collocational relations (Sinclair, 1991), Sailer (2003) formulates the *Predictability Hypothesis.*

(19)   Predictability Hypothesis (Sailer, 2003, p. 366):
        For every sign whose internal properties are fully predictable, the distributional behavior of this sign is fully predictable as well.

Since simple lexical items such as basic words or nonderived lexemes are internally idiosyncratic, they may also display idiosyncratic distributional properties.

---

[3]Historically the expression stems from the phrase *thy kingdom come* in the Lord's Prayer. There the noun *kingdom* is used with a determiner, and the entire expression is a clause, not a noun phrase.

[4]As discussed in Sailer (2003), this sets apart our HPSG treatment of internally idiosyncratic phrases from an analysis that relies on *en bloc* insertion or the analysis in Tree Adjoining Grammar in Abeillé (1995), which treats non-decomposable idioms as idiosyncratic trees of arbitrary depth.

Similarly, internally idiosyncratic phrases may also show distributional irregularities. Again, the phrase *kingdom come* is a good example. It is almost exclusively restricted to the combinations with one of the three prepositions in *until/till/to kingdom come*.[5] Based on this observation we may assume that a more complete description of the COLL list of the PLE (18) should mention a *barriers* object whose purpose it is to require that *kingdom come* be the complement of one of these prepositions.

3.2   Partial Regularity in Irregular Phrases

While the idiom *kingdom come* is syntactically irregular, the majority of idioms is not. Idiomatic expressions are characterized by a particular, 'idiomatic' meaning and by their frozen syntactic structure, but apart from their unusual fixedness they have regular internal syntactic properties. Typical examples are expressions such as *kick the bucket* (*die*) and *saw logs* (*snore*). In both cases, we assume that the syntactic structure is that of a transitive verb that combines with its direct object. In the version of our theory in Section 3.1 we had to encode all regular aspects of the structure of idioms explicitly in each phrasal lexical entry. This introduces an unwanted descriptive overhead in grammars which already contain all relevant syntactic well-formedness conditions on phrases. In the present section we propose an extension of the theory that captures the insight that many irregular phrases are merely irregular with respect to the syntax-semantics mapping but not with respect to their syntax.

The core innovation is a distinction between constraints that apply to different modules of the grammar. We tentatively assume that there are phonological, syntactic, and semantic principles. The SUBCAT PRINCIPLE is a syntactic constraint, the SEMANTICS PRINCIPLE is one of the semantic ones, and linearization principles count as phonological constraints. Since the COLL value of signs is the place where we mark their idiosyncrasies, this is also where we specify to which degree a sign exhibits idiosyncratic behavior. For that purpose, we enrich the structure of COLL values. From now on they are of sort *coll*. The subsorts of *coll* specify the degree of regularity of an expression. For irregular items an attribute REQ(UIREMENT) is defined, whose value is a list of *barrier* objects. This list corresponds exactly to the earlier COLL value. The details are provided in FIGURE 1.

The names of the maximally specific subsorts of *coll* are inspired by the corresponding classification of idioms in Fillmore et al. (1988). In that system the idiom *kingdom come* is classified as extra-grammatical because it only shows regularity with respect to its phonological properties. In our hierarchy, it is classified as an irregular construction which obeys the restrictions of phonologically regular constructions but not those of syntactically or semantically regular constructions. It receives a phrasal lexical entry with the COLL value *extra-grammatical-idiom*.

---

[5]The British National Corpus contains 35 occurrences of *kingdom come* out of which 5 are irrelevant (band name or coincidental co-occurrence of the two words), 14 reflect the biblical usage, 13 are with one of the above-mentioned prepositions, 3 others are uses of *kingdom come* as a noun.

Figure 1: Sort hierarchy below the sort *coll*
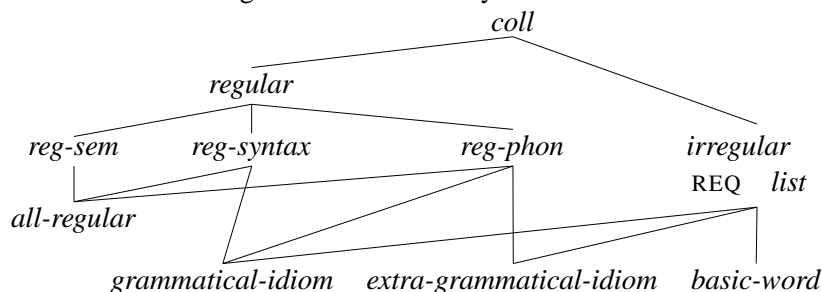
The idiom *saw logs* is a grammatical idiom according to Fillmore et al. (1988), and is therefore specified as [COLL *grammatical-idiom*]. The maximally specific sort *grammatical-idiom* is a subsort of both *regular-phonology* and *regular-syntax*. For that reason it is subject to all syntactic and phonological principles, but not to the principles of regular semantic composition. Finally, regular phrases have the COLL value *all-regular* and obey all principles of syntax, semantics and phonology.

In Section 3.1 we said that the principles of grammar only apply to signs with an empty COLL list. This theory must now be revised and made sensitive to the subsorts of *coll*. Syntactic principles such as the IMMEDIATE DOMINANCE PRINCIPLE (ID PRINCIPLE), the HEAD FEATURE PRINCIPLE, the NONLOCAL FEATURE PRINCIPLE etc. must be modified. All syntactic principles defined on the sort *sign* are relativized to apply only to signs with the COLL specification *regular-syntax*. This can be achieved by simply adding a further antecedent to the original formulation of the principles. The relativized version of the ID PRINCIPLE in (20) illustrates this technique.

(20)   Relativized ID PRINCIPLE:

$$\begin{bmatrix} \text{COLL} & \textit{regular-syntax} \end{bmatrix}$$
$$\Rightarrow \big(\textit{phrase} \Rightarrow (\text{HEAD-SPECIFIER-SCHEMA } \textbf{or } \text{HEAD-COMPLEMENT-SCHEMA } \textbf{or } \dots)\big)$$

Idioms that are not extra-grammatical are still subject to all syntactic well-formedness conditions according to the *coll* hierarchy, because *grammatical-idiom* is subsort of *regular-syntax*. It follows that each grammatical idiom must obey one of the ID SCHEMATA. For example, the VP *saw logs* is specified as a regular head-complement construction, and as such it is licensed by the HEAD-COMPLEMENT SCHEMA. No special steps need to be taken to guarantee this result.

Let us finally turn to the principles of semantic composition. While many idioms are syntactically regular, they all show semantic idiosyncrasy. To capture this behavior, the principles of semantic composition need to be relativized parallel to what we did in (20). The relativized SEMANTICS PRINCIPLE is given in (21), where SP is the description of the original SEMANTICS PRINCIPLE.

(21)   Relativized SEMANTICS PRINCIPLE:

$$\left[\text{COLL}\ \textit{regular-semantics}\right] \Rightarrow \text{SP}$$

The sort hierarchy in FIGURE 1 is constructed in such a way that we exclude the existence of irregular phrases that are semantically regular but syntactically or phonologically irregular. On the other hand, as soon as a phrase is syntactically irregular, we expect it to be semantically irregular as well. In this respect we agree with the assumptions in Fillmore et al. (1988).[6]

So far we have focussed on the principles of grammar. Let us now consider the question of what constitutes the lexicon. All properties of signs with the COLL specification *regular* follow from the properties of their constituents and from the general combinatorial principles of the grammar. Signs with the COLL value *irregular*, however, require a further specification of those of their properties that are not predictable from general grammar rules. This specification is given in the lexicon. In our sort hierarchy below *coll* we distinguish three subsorts of *irregular*. The sorts *grammatical-idiom* and *extra-grammatical-idiom* are confined to irregular signs that have non-trivial internal syntactic structure. In the context of the present discussion, these are phrasal signs. The sort *basic-word* is reserved for signs without internal structure. In the present discussion, this means that it is the COLL value of words. Words (which we view here as non-recursive signs) always display an unpredictable form-meaning combination, which qualifies them as irregular. The idea that basic words are necessarily irregular and that phrases cannot have the COLL value *basic-word* is captured in the principle in (22).

(22)   BASE-LEXICON PRINCIPLE:

$$\left[\textit{word}\ \right] \Leftrightarrow \left[\text{COLL}\ \textit{basic-word}\right]$$

In the preceding discussion we deliberately ignored the fact that words may have internal structure as well. As soon as a more elaborate view on morphological structure (such as the one presented in Sag et al. (2003)) is adopted, the BASE-LEXICON PRINCIPLE needs to be refined in such a way that the most basic, non-recursive subsort of *sign* replaces *word* on the lefthand side of the principle. Furthermore, the type hierarchy below *coll* will need some extension as well in response to additional principles of the morphological combinatorics.

The lexicon is defined by means of a WORD PRINCIPLE. This principle provides lexical entries for all irregular signs. In (23) LE refers to lexical entries of basic words, PLE refers to phrasal lexical entries of grammatical or extra-grammatical idioms.

---

[6]All idioms considered in this paper are phonologically regular. Nonetheless we include the type *regular-phon* to allow for a relativization of the principles of phonological combinatorics such as the CONSTITUENT ORDER PRINCIPLE.

(23) WORD PRINCIPLE:

$$\begin{bmatrix} sign \\ \text{COLL} \;\; irregular \end{bmatrix} \Rightarrow (\; \text{LE}_1 \vee \ldots \text{LE}_n \vee \text{PLE}_1 \vee \ldots \vee \text{PLE}_{n'} \;)$$

In Section 3.1 we emphasized the importance of the LISTEME attribute for our theory of idioms. The name "listeme" is chosen very deliberately in Soehn (2006) because Soehn assumes that all listed expressions contribute their own unique LISTEME value. This means that every lexical entry, phrasal or not, may have its own LISTEME value. An internally irregular phrase such as *kick the bucket* has a LISTEME value, say *kick-the-bucket-idiom*, which differs from the LISTEME values of all of its daughters. While the HEAD FEATURE PRINCIPLE guarantees that all head features such as VFORM, AUX are shared between the phrasal mother and the head daughter, the idiomatic phrase *kick the bucket* and its head daughter do not share the LISTEME value. It follows that Soehn's assumption that LISTEME is a head feature is not compatible with the present architecture. For this reason we treat LISTEME as a *category* feature instead. With the new position in the feature geometry, it is necessary to introduce a principle for the percolation of LISTEME values in regular phrases. The principle that takes care of that is given in (24). It is among those principles that apply only to non-idiomatic phrases.

(24) The LISTEME PRINCIPLE:

$$\begin{bmatrix} \text{DTRS} \;\; headed\text{-}phrase \\ \text{COLL} \;\; all\text{-}regular \end{bmatrix} \Rightarrow \begin{bmatrix} \text{SYNS LOC CAT LISTEME} \;\boxed{1} \\ \text{DTRS} \; \begin{bmatrix} \text{H-DTR} \; \begin{bmatrix} \text{SYNS LOC CAT LISTEME} \;\boxed{1} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

As a consequence of our classification of grammar principles into syntactic, semantic and phonological, the structure of the *coll* hierarchy, and the fundamental lexical principles (22) and (23), there are four possibilities for a well-formed sign: Basic words always exhibit some degree of idiosyncrasy and are singled out as having their own irregular collocation type, *basic-word*. Phrases come in three flavors. A phrase may be completely regular, in which case it has the COLL *all-regular* and is subject to all principles of grammar. If it is irregular, it must be licensed by one of the phrasal lexical entries in the WORD PRINCIPLE. If a phrase is of COLL TYPE *grammatical-idiom*, it is subject to the regular syntax principles. If a phrase is an *extra-grammatical-idiom*, it is irregular to a degree that it is exempt from the principles of syntax.

Before closing this section let us briefly return to the role of the Predictability Hypothesis (19). This hypothesis establishes a link between internal irregularity and the potential of specifying external idiosyncrasy. In the version of the theory in Section 3.1, all and only regular signs have an empty COLL list. In the modified architecture of the present section, only signs with a COLL value of sort *irregular* have an REQ attribute, and only they can be specified for externally idiosyncratic behavior. The idea of the Predictability Hypothesis is directly encoded in the signature of the grammar module that handles irregularity.

Note that our new architecture foresees cases in which an internally irregular sign is not distributionally constrained. While this possibility is denied in most of the work on collocations, it seems to be the standard assumption in formal approaches to grammar. It is also compatible with the original formulation of the Predictability Hypothesis. So far, we do not see compelling reasons for claiming that idioms such as *kick the bucket* are distributionally constrained. In previous versions of the theory we were forced by the architecture to assume that there was some *barrier* object inside the COLL value, although no such object was explicitly specified. With the new version of the theory, it is possible to combine the COLL value *grammatical-idiom* of a phrase with an empty COLL REQ list.
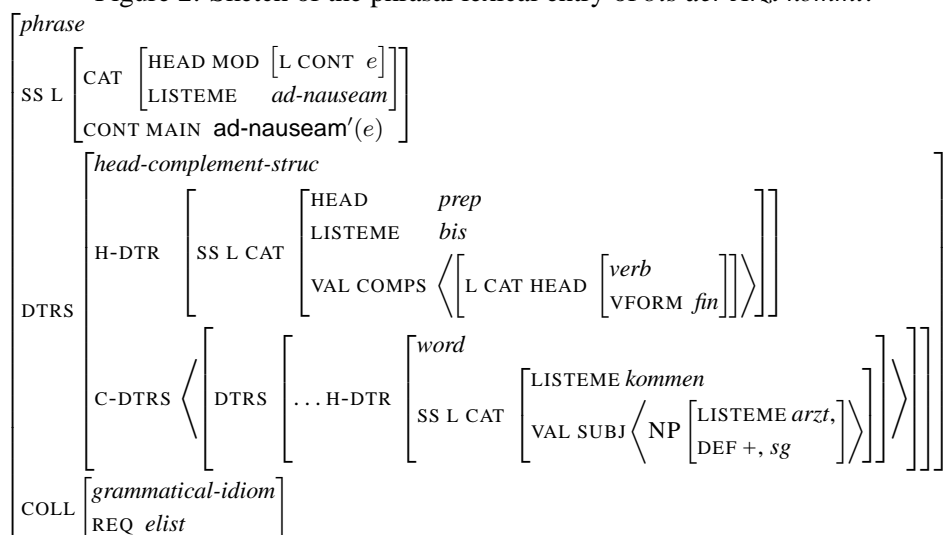
4   Modeling Phraseological Clauses as Phrasal Lexical Entries

In FIGURE 2 we sketch the PLE for the idiom in (2-a). The PLE specifies that the overall clause is a modifier with the semantics *ad nauseam*. The phrase is a head-complement combination, where the head daughter is the preposition *bis* (*until*). The non-head daughter is a finite clause. Inside the complement, there must be a verbal word with the LISTEME value *kommen* whose subject is a definite singular NP with the word *Arzt* as its lexical head. The PLE specifies the COLL value as *grammatical-idiom*. Consequently, all principles of syntax apply, which means that we do not need to specify the HEAD value of the clause nor the effect of the SUBCAT PRINCIPLE or of the HEAD-COMPLEMENT SCHEMA. The REQ value of the clause is empty, which expresses the observation that there are no further constraints on the distribution of the PCl.

The data section showed that there are some restrictions on the structure of this PCl: While tense and modality may vary, negation is not permitted. This can be expressed by requiring that there be no negation in the content of the PCl. For other PCls we must also ban modal operators from the semantic representations. Since modalities can be contributed by modal verbs and by adverbials, the restriction must be imposed on operators in the CONTENT value of the PCls.

In Section 2 we saw that all PCls we considered disallow alternations that change the information structure of their literal meaning. Since the constituents of PCls are non-idiomatic in our theory, the literal meaning of their combination is in principle available. As there are various proposals to model information structure in HPSG, it should in principle be possible to formulate an appropriate constraint on information structure. For reasons of space, we will not pursue this direction here. Instead, we exclude valence alternations by other types of restrictions in the PLEs. To exclude the passive and the dative-possessive alternation in (1-a) and (1-b), we impose syntactic restrictions on the ARG-ST or the VAL value of the words in the expressions, which are all available in PLEs. To keep the analysis simple and as complete as possible, we will stick with this strategy for the rest of the paper.

310

Figure 2: Sketch of the phrasal lexical entry of *bis der Arzt kommt*:

$$
\begin{bmatrix}
\textit{phrase} \\
\text{SS L}
\begin{bmatrix}
\text{CAT}
\begin{bmatrix}
\text{HEAD MOD } \begin{bmatrix} \text{L CONT } e \end{bmatrix} \\
\text{LISTEME } \textit{ad-nauseam}
\end{bmatrix} \\
\text{CONT MAIN ad-nauseam}'(e)
\end{bmatrix} \\
\text{DTRS}
\begin{bmatrix}
\textit{head-complement-struc} \\
\text{H-DTR} \begin{bmatrix} \text{SS L CAT} \begin{bmatrix} \text{HEAD } \textit{prep} \\ \text{LISTEME } \textit{bis} \\ \text{VAL COMPS} \left\langle \begin{bmatrix} \text{L CAT HEAD} \begin{bmatrix} \textit{verb} \\ \text{VFORM } \textit{fin} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix} \\
\text{C-DTRS} \left\langle \begin{bmatrix} \text{DTRS} \begin{bmatrix} \ldots \text{H-DTR} \begin{bmatrix} \textit{word} \\ \text{SS L CAT} \begin{bmatrix} \text{LISTEME } \textit{kommen} \\ \text{VAL SUBJ} \left\langle \text{NP} \begin{bmatrix} \text{LISTEME } \textit{arzt,} \\ \text{DEF +, } \textit{sg} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle
\end{bmatrix} \\
\text{COLL} \begin{bmatrix} \textit{grammatical-idiom} \\ \text{REQ } \textit{elist} \end{bmatrix}
\end{bmatrix}
$$

Let us now turn to a more intricate example. The data (25)–(28) reveal details about the frozenness of the PCl in (1-c). We compare the PCl in the (a) sentences with a parallel non-idiomatic construction in the (b) sentences. (25) shows that the PCl requires an anaphoric relation between the matrix subject and the accusative argument in the PCl. As we remarked earlier, neither an overt complementizer nor a change in the constituent that occupies the *vorfeld* are permitted (see (26) and (27)). The PCl may not occur in the *vorfeld* of the matrix sentence (28).

(25)   a.   Ich glaub, mich/#dich tritt ein Pferd.
           I   believe me/you   kicks a   horse
           'I am very surprised.'
      b.   Ich glaub, mich/dich jagt eine Kuh.
           I   believe me/you   chases a   cow
           'I believe a cow is chasing me/you.'

(26)   a.   #Ich glaub, dass mich ein Pferd tritt.
      b.   Ich glaub, dass dich eine Kuh jagt.

(27)   a.   #Ich glaub, ein Pferd tritt mich.
      b.   Ich glaub, eine Kuh jagt dich.

(28)   a.   #[Mich tritt ein Pferd], glaub ich.
      b.   [Dich jagt eine Kuh], glaub ich.

In FIGURE 3 we sketch the relevant PLE. The COLL value *grammatical-idiom* accounts for the syntactically regular internal structure. The PCl is specified as a verb-second clause whose lexical head is the verb *treten*. This verb must take two arguments. The first one is an indefinite NP headed by *Pferd*. The second one is an

accusative NP which is fronted. This condition follows from the LOC value identity, [1], between the element on the ARG-ST list and the highest non-head daughter, which is a filler daughter. The VP might be modified by adjuncts, which is accounted for by only requiring that *treten* be the syntactic head of the construction. The appeal to the regular expression notation ((DTRS HDTR)$^+$) is only meant as a more readable abbreviation of a (technically more accurate) relational expression that relates the head daughter of the head-filler structure to its head daughters.

Next we turn to the COLL REQ value. As in the other sentences in (1), the combination of the matrix verb and the PCl behaves like in decomposable idioms of the type *spill the beans*, except that the complement is now a clause instead of an NP. According to the theory in Soehn (2004), this means that the matrix verb selects a complement with a particular LISTEME value. The complement clause, in turn, has a non-empty REQ list. The element on its REQ specifies that the PCl must co-occur with a particular matrix verb, the listeme *surprise-glauben*. The PCl must be the complement clause of this matrix verb. Furthermore, the sort specification indicates the syntactic domain within which the co-occurrence must hold. In Soehn (2004) the sort *vp_ne* is used to specify that the relevant domain is the LOCAL value of the smallest projection of the matrix verb that dominates both the matrix verb and the complement clause. In other words, the PCl must occur as a sister to (the trace of) the matrix verb. What is most important for our purposes is that information about the matrix verb is available in the formulation of the PLE of the clausal complement. This is necessary to encode that the INDEX value of the embedded direct object, [2], is identical with that of the matrix subject.
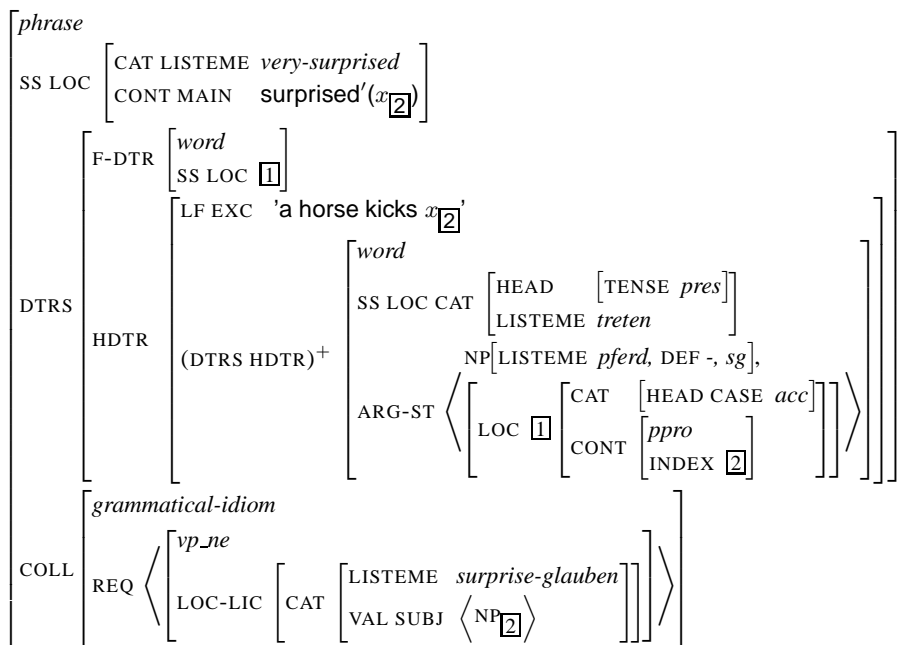
To sum up, the PLE in FIGURE 3 excludes passive alternation (see (5-c)), because it specifies that the verb *treten* occurs with a transitive argument structure. It also requires that the PCl be a verb-second clause (see (6-a)) by specifying that it is a head-filler structure, and it determines the first constituent (see (6-b)) by specifying it in the PLE. The anaphoric relationship between the embedded accusative and the matrix subject is also encoded directly.

There is in fact further evidence that a special relationship holds between the embedded PCl and the matrix predicate, and that the matrix predicate is not the free form of the verb *glauben*. We already saw in (9) that, depending on the tense form, the matrix verb is either a form of *glauben* or of *denken*. The relevant judgments are shown in (29).

(29) a. Ich dachte, mich tritt ein Pferd. (past of *denken*, present in the PCl)
  b. ??Ich glaubte, mich tritt ein Pferd. (past of *glauben*, present in the PCl)
  c. #Ich glaubte, mich trat ein Pferd. (past of *glauben*, past in the PCl)
  d. ?Ich denke, mich tritt ein Pferd. (present of *denken*, present in the PCl)

The combination of *glauben* and the PCl is a decomposable idiom, because the very same variation of the matrix predicate can be observed with other PCls (see (9)). German has a special listeme (which we call *surprise-glauben*) which comprises forms of *glauben* and *denken* in its paradigm and combines with complement

Figure 3: Sketch of the PLE for the idiom *glauben, X_acc tritt ein Pferd*:

$$
\begin{bmatrix}
\textit{phrase} \\[4pt]
\text{SS LOC} \begin{bmatrix} \text{CAT LISTEME} & \textit{very-surprised} \\ \text{CONT MAIN} & \text{surprised}'(x_{\boxed{2}}) \end{bmatrix} \\[10pt]
\text{DTRS} \begin{bmatrix} \text{F-DTR} \begin{bmatrix} \textit{word} \\ \text{SS LOC } \boxed{1} \end{bmatrix} \\[6pt]
\text{HDTR} \begin{bmatrix} \text{LF EXC} & \text{'a horse kicks } x_{\boxed{2}}\text{'} \\[4pt]
(\text{DTRS HDTR})^{+} \begin{bmatrix} \textit{word} \\ \text{SS LOC CAT} \begin{bmatrix} \text{HEAD} & [\text{TENSE } \textit{pres}] \\ \text{LISTEME } \textit{treten} \end{bmatrix} \\[6pt]
\text{ARG-ST} \left\langle \begin{bmatrix} \text{NP}[\text{LISTEME } \textit{pferd}, \text{ DEF -, } sg], \\ \text{LOC } \boxed{1} \begin{bmatrix} \text{CAT} & [\text{HEAD CASE } acc] \\ \text{CONT} & \begin{bmatrix} \textit{ppro} \\ \text{INDEX } \boxed{2} \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix} \end{bmatrix} \end{bmatrix} \\[10pt]
\text{COLL} \begin{bmatrix} \textit{grammatical-idiom} \\[4pt]
\text{REQ} \left\langle \begin{bmatrix} \textit{vp\_ne} \\ \text{LOC-LIC} \begin{bmatrix} \text{CAT} \begin{bmatrix} \text{LISTEME} & \textit{surprise-glauben} \\ \text{VAL SUBJ} & \langle \text{NP}_{\boxed{2}} \rangle \end{bmatrix} \end{bmatrix} \end{bmatrix} \right\rangle \end{bmatrix}
\end{bmatrix}
$$

clauses that express (negative) surprises, astonishment, or annoyance. (30) shows that this listeme can be found with non-idiomatic complement clauses as well.

(30)   Ich glaub/ ?denk(e)/ ?*glaubte/ dacht(e)

    a.   der hat 'nen Vogel.
       he has a bird   ('...he is crazy')

    b.   das muss jetzt echt alles nochmal neu gemacht werden.
       this must now really all again new made be
       ('...this must all be redone [annoyed]')

We conclude that even though the matrix predicate is not the free form of *glauben*, it is an instance of a (special) attitude predicate that also occurs outside of idioms. For this reason, the matrix predicate need not be restricted to a particular PCl. However, the PCl in FIGURE 3 must be collocationally bound to this special matrix predicate, and the PCl must impose its context requirement in the lowest dominating VP to exclude its own topicalization.

5   Modelability under Strict Locality Assumptions?

The two-dimensional theory of idioms is capable of capturing the properties of PCls. Being able to refer to deeply embedded parts of a phrase in a PLE is an

important ingredient of this theory. It makes HPSG especially well-suited to integrate a fundamental insight of Construction Grammar: Constructions can span more than a local tree (Fillmore et al., 1988; Jackendoff, 1995).

In this section we briefly consider a few interesting aspects of a second approach to construction-like phenomena in HPSG, which offers a possible alternative to our analysis of PCls. However, we do not intend a thorough comparison of the two approaches and only point out a few interesting similarities and differences. In a recent series of papers (Sag (2007a,b) and others) it was shown that various phenomena of apparent non-locality can be encoded using an extension of HPSG's feature geometry and a restructuring of signs. In the framework proposed there, *Sign-Based Construction Grammar* (SBCG), phrasal signs no longer contain their daughters. Instead, *construct* objects are introduced that correspond to local trees. Signs only occur as nodes in these constructions. A sentence consists of a set of constructions, each of which represents a local tree, but these trees do not form a single joined feature structure. With this architectural change the formulation of PLEs like the ones in FIGURE 2 and FIGURE 3 is not possible.

To account for non-locality SBCG uses two head features: the listeme attribute LEXICAL-ID and the attribute XARG whose value is the subject of the sentence. These two attributes are sufficient to describe the construction in (2-a), because the obligatory elements in the embedded clause are the lexical head *kommen* and the subject, *Arzt*, i.e. exactly those parts that are locally available for the overall construction.

(31)   A SBCG description of *bis der Arzt kommt*:

$$
\begin{bmatrix}
\textit{bis-der-arzt-kommt-cxt} \\
\text{MOTHER} \begin{bmatrix} \text{MOD} & [\text{SEM} \ \boxed{1}] \\ \text{SEM} & \textit{ad-nauseam}(\boxed{1}) \end{bmatrix} \\
\text{DTRS} \quad \left\langle [\text{LID} \ \textit{bis}], \ \text{S} \begin{bmatrix} \text{XARG} & [\text{LID} \ \textit{arzt}] \\ \text{LID} & \textit{kommen} \end{bmatrix} \right\rangle
\end{bmatrix}
$$

To allow modal verbs and temporal auxiliaries we can simply assume that the LID value of a verbal complex is identical with that of the most deeply embedded lexical verb in the verbal complex. To exclude modal and temporal variation in other idioms, we could impose the same kind of restrictions as in Section 4, i.e. we could describe which operators may not occur in the content values of the daughters.

Recall that truth-conditionally neutral, grammatical variation occurs in some but not all PCls. In the two-dimensional account we refer to the ARG-ST value of an embedded verb to exclude passive and other valence alternations. Since SBCG allows reference to the highest subject in a PCl, active-passive alternations can similarly be excluded by requiring a particular LID value inside the XARG value. Alternations that do not involve the subject are harder to treat because it is only the subject information that percolates up the tree.

This brings us to an interesting problem. Information about the arguments in-

side a PCl is not only necessary to restrict valence alternation, it is also important to express the coreference constraints attested with many PCls. The PCl in (1-c) is a good example. The accusative NP inside the PCl must be a pronoun that is anaphorically related to the matrix subject. The accusative object is on the ARG-ST list of the embedded verb. The matrix subject is on the ARG-ST list of the matrix verb, the matrix verb has access to the LID value of the embedded verb and to its XARG value. However, neither of them can be used to establish a link between the embedded accusative NP and the matrix subject. The same problem occurs in other cases where the PCl contains an embedded open slot that must be anaphorically related to the matrix subject: The PCls in (1-b) and (2-c) require such a relation to an embedded dative object. A potential way out within SBCG is the introduction of a percolation mechanism for the entire ARG-ST values instead of the more restricted subject percolation mechanism. While this solution works for the cases of German PCls that we have found so far, the English example in (32) might still be a problem. In this expression the element X must be coreferential with the matrix subject. However, X is embedded in a locative modifier. Unless locative modifiers are on the ARG-ST list, the locality assumptions of SBCG do not seem to leave the necessary kind of structure accessible to enforce coreference between X and the matrix subject.

(32)   look as if butter wouldn't melt [in X's mouth] ('look completely innocent')

At the moment, we do not see which kind of solution is most appropriate for the general locality assumptions that underly SBCG. We thus leave this issue to future research.[7]

6   Conclusion

In this paper we drew attention to a largely neglected subclass of idioms: Idioms that contain full clauses, phraseologically fixed clauses (PCl). We investigated properties of German PCls and arrived at new generalizations about their potential fixedness and flexibility. While there is a certain range of syntactic and lexical variation, all PCls we investigated forbid the application of syntactic processes that change the information structure of their literal meaning.

To account for the frozenness of PCls together with their regular internal syntactic structure we substantively modified the two-dimensional theory of idioms developed in Richter and Sailer (2003), Sailer (2003), and Soehn (2006). These earlier versions of the theory had already incorporated the distinction between decomposable and non-decomposable idioms, but only the modified theory lets us express the systematic differences between grammatical and extra-grammatical idioms. The theory captures the empirical properties of German PCls.

---

[7]See (Müller, 2007, Chapter 12.3) for further critical remarks on SBCG's locality assumptions and fundamental open questions about its architecture.

We very briefly compared our account with a possible alternative analysis in the framework of Sign-Based Construction Grammar (SBCG). Some properties of PCls may be problematic for SBCG's strict locality assumptions. Our theory can be seen as taking a middle position between the SBCG view, which demands that constructions only span local trees, and the traditional Construction Grammar perspective, which holds that constructions can be of arbitrary structural complexity. In our system a construction is licensed by a phrasal lexical entry (PLE). A PLE does two important things: First, an idiosyncratic semantic and/or syntactic combination is licensed in a local tree. Second, restrictions can be imposed on constituents that are embedded inside this combination. The first property is a weak version of a locality assumption: A PLE can only license an idiosyncrasy in an immediate mother-daughter relation. The second property, however, is a weak version of a complexity assumption: We can refer to properties of elements that are deeply embedded in the structure of the phrasal sign licensed by the PLE. In this setting it is crucial to realize that the embedded constituents must be independently well-formed. This means that we can restrict which ones of the well-formed signs may occur inside the overall expression, but a PLE cannot license embedded, idiosyncratically structured signs. In this sense, our approach incorporates the idea of arbitrary depth of constructions, but it also inherits the insight of phrase structure grammars that complex structures are built from local trees.

The two-dimensional theory of idioms that we developed in this paper helps us to reduce the amount of individually specified idiosyncrasy in the description of idiomatic constructions even further than its predecessor. The principles of the regular syntactic combinatorics apply to grammatical but non-decomposable idioms. We obtain a very flexible grammar architecture which covers two apparently contradicting tendencies in the domain of idioms at the same time: The need to allow for irregularity at all levels; and the observation that most idioms are not completely arbitrary in their structure but largely obey regular principles of grammar.

References

Abeillé, Anne. 1995. The Flexibility of French Idioms: A Representation with Lexical Tree Adjoining Grammar. In Martin Everaert, Eric-Jan van der Linden, André Schenk and Rob Schreuder (eds.), *Idioms. Structural and Psychological Perspectives*, pages 15–42, Lawrence Erlbaum Associates, Hillsdale.

Fillmore, Charles, Kay, Paul and O'Connor, M. 1988. Regularity and Idiomaticity in Grammatical Constructions: The Case of *Let Alone*. *Language* 64, 501–538.

Fleischer, Wolfgang. 1997. *Phraseologie der deutschen Gegenwartssprache*. Niemeyer, Tübingen, second edition.

Gazdar, Gerald, Klein, Ewan, Pullum, Geoffrey and Sag, Ivan. 1985. *Generalized Phrase Structure Grammar*. Harvard University Press, Cambridge, Mass.

Jackendoff, Ray. 1995. The Boundaries of the Lexicon. In Martin Everaert, Eric-Jan van der Linden, André Schenk and Rob Schreuder (eds.), *Idioms. Structural and Psychological Perspectives*, pages 133–165, Lawrence Erlbaum Associates, Hillsdale.

Müller, Stefan. 2007. *Head-Driven Phrase Structure Grammar: Eine Einführung*. Tübingen: Stauffenburg Verlag.

Richter, Frank and Sailer, Manfred. 2003. Cranberry Words in Formal Grammar. In Claire Beyssade, Olivier Bonami, Patricia Cabredo Hofherr and Francis Corblin (eds.), *Empirical Issues in Formal Syntax and Semantics*, volume 4, pages 155–171, Paris: Presses Universitaires de Paris-Sorbonne.

Sag, Ivan A. 2007a. Remarks on Locality. In Stefan Müller (ed.), *Proceedings of the 14th International Conference on Head-Driven Phrase Structure Grammar, Stanford, 2007*, pages 394–414, Stanford: CSLI Publications.

Sag, Ivan A. 2007b. Sign-Based Construction Grammar. An informal synopsis, manuscript, Stanford.

Sag, Ivan A., Wasow, Thomas and Bender, Emily M. 2003. *Syntactic Theory: A Formal Introduction*. Stanford: CSLI, second edition.

Sailer, Manfred. 2003. Combinatorial Semantics and Idiomatic Expressions in Head-Driven Phrase Structure Grammar. Phil. Dissertation (2000). Arbeitspapiere des SFB 340. 161, Universität Tübingen.

Sinclair, John. 1991. *Corpus, Concordance, Collocation*. Oxford: Oxford University Press.

Soehn, Jan-Philipp. 2004. License to COLL. How to bind bound words and readings to their contexts. In Stefan Müller (ed.), *Proceedings of the 11th International Conference on Head-Driven Phrase Structure Grammar, Center for Computational Linguistics, Katholieke Universiteit Leuven*, pages 261–273, Stanford: CSLI Publications.

Soehn, Jan-Philipp. 2006. *Über Bärendienste und erstaunte Bauklötze. Idiome ohne freie Lesart in der HPSG*. Frankfurt am Main: Peter Lang, Ph.D. thesis, Friedrich-Schiller-Universität Jena.

Wasow, Thomas, Sag, Ivan A. and Nunberg, Geoffrey. 1983. Idioms: An Interim Report. In S. Hattori and K. Inoue (eds.), *Proceedings of the XIIIth International Congress of Linguists*, pages 102–115.