

On Idiom Parts and their Contexts*

Jan-Philipp Soehn (Tübingen)

Abstract

This article examines idiomatic expressions as sources of both regularity and irregularity in language. Some morphological, lexical, syntactical, and semantical characteristics of idioms are discussed. It is shown how a lexical licensing mechanism, which is formulated within a formal grammar framework, can deal with the data. After that, this proposal is extended to the phenomenon of negative polarity.

KEYWORDS: Irregularity, Regularity, HPSG, Idioms, Bound Words, Polarity Items, German

1 Introduction

- (1) Finally, the wrangling neighbours have buried the ax.
- (2) The early bird catches the grub.
- (3) We will be required to ignite the midnight petroleum.¹

These expressions reveal some interesting properties of idioms. The fact that the reader automatically understands which idiom or proverb has been altered in the three examples (*to bury the hatchet*, *the early bird catches the worm*, *to burn the midnight oil*) shows that these expressions are not unparseable blocks of language. Instead, we understand idioms to be entities consisting of specific parts. However, although we recognise the English idioms underlying our introductory examples, nobody would say that these sentences are well-formed instances of these idioms. As a matter of fact, idioms must occur as whole units: It seems that the components of an idiomatic expression like to close ranks and never occur without each other. In other words, idiom parts must be licensed by each other.

Another phenomenon we would like to draw attention to is polarity. Our examples (4)² through (6) show that idioms are sensitive even to the contexts in which they occur. Thus, licensing not only affects lexical entities but also semantic properties of the context.

* The research for this paper was funded by the *Deutsche Forschungsgemeinschaft*. I am grateful to Doris Penka, Frank Richter, Christine Römer and Manfred Sailer for insightful comments and discussion and Janah Putnam for help with the challenges of English.

¹Data's misquoting of *to burn the midnight oil* in the television series "Star Trek: The Next Generation".

²The phrase in (4) without negation can only be understood literally.

- (4) *(Don't) put all your eggs in one basket.
 (5) Nobody/*someone lifted a finger to help her.
 (6) I can*(not) make head or tail of it.

This contribution is organized as follows: First, some properties and peculiarities of idioms will be explored. We refer to English and German data as evidence for our claims. Secondly, a lexical licensing mechanism is sketched, which has been implemented in a formal grammar framework. Lastly, this approach is extended to semantic licensers such as negation.

2 Regularity and Irregularity of Idioms

In this section, we consider some properties of idioms which violate the general criteria of regularity in language. We will state these criteria of regularity (CR)³, taken from Sailer (2003: Ch. 6.1) and Soehn (2006).

2.1 Morphological and Lexical Characteristics

CR 1 Every lexical item is morphologically of a regularly built shape.

CR 2 Every word belongs to a regular inflectional paradigm.

2.1.1 Fixed Properties

There are idioms which violate these morphological and lexical criteria. Number (singular in (7)), tense (present in (8)), or mode (passive in (9)) can be fixed. Thus, these idioms can be used only with a subset of (otherwise grammatical) inflectional forms.⁴

- (7) *ein blaues Wunder / (*zwei blaue Wunder) erleben*
 "a blue wonder / (*two blue wonders) experience"
 'to get a nasty surprise'
- (8) *etw. ist gehüpft wie gesprungen*
 "sth. is hopped as leaped"
 'it doesn't matter'
- (9) *mit allen Wassern gewaschen sein*
 "with all waters washed be"
 'to be up to every trick'

2.1.2 Morphological Anomalies

Besides idioms consisting of regular words, we encounter anomalies as an archaic form of the dative plural (*Wassern* in (9)) or dative singular (10). Another anomaly is the missing agreement in (11).

³Note that these criteria are generalizations which are violated not only by idioms but also by other irregular(!) phenomena.

⁴We are aware of the fact that there are specific literary or expressive contexts in which certain idioms can be used more freely or creatively.

- (10) *etw. schlägt zu **Buche***
 "sth. strikes to book"
 'sth. adds up'
- (11) *sich bei jdm. **lieb** Kind machen*
 "refl-pron at sb._(dat) dear child make"
 'to endear oneself to sb.'

These anomalies and frozen properties show that idioms need not consist of the (synchronically) normal lexical inventory of a language, but have idiosyncratic properties on the morphological level.

2.1.3 Bound Words

Idioms may also comprise bound words or "cranberry words" (Aronoff 1976) – lexical elements which are highly collocationally restricted. These can occur only in very specific environments. For German, the Collaborative Research Center 441 (Project A5) at the University of Tübingen compiled about 450 such instances⁵ from the literature, cf. Dobrovolskij (1988), Dobrovolskij/Piirainen (1994) and Fleischer (1997). Dobrovolskij (1988) provides quite a large number of examples for German, Dutch and English.

- (12) *to learn/do sth. by **rote*** 'automatically, by heart'
- (13) *to cock a **snook*** 'to thumb the nose'

The words printed in bold are restricted to the given contexts. Sometimes there is some variation, as in *to lie/go/lay **doggo*** (Brit. slang; 'to hide oneself'), but a free distribution is not possible. Some German examples:

- (14) *jdn. über den Löffel **balbieren***
 "sb._(acc) over the spoon barber"
 'to cheat on sb.'
- (15) ***Fersengeld** geben*
 "heel money give"
 'to turn tail and run'

Bound words (or: unique elements) are lexical units which have been "frozen" during language development over time. Dobrovolskij (1988: 87) calls them relics from an earlier stage of language development. Thus, the mere occurrence of a bound word is an unequivocal indication that the phrase must be idiomatic. This is because idioms with a possible non-idiomatic reading only consist of material which can be used unrestrictedly.

⁵See <http://www.sfb441.uni-tuebingen.de/a5/codii/index.html>.

2.2 Syntactic Characteristics

CR 3 Every phrase is syntactically of a regularly built shape.

CR 4 Every element in a phrase occurs in the same form in some other combination.

2.2.1 Anomalies

In a relatively small number of idioms, we can find syntactic anomalies in the surface structure. Consider the following use of a count noun without a determiner:

(16) *to follow suit* 'to do the same thing as the person preceding you'

This kind of construction does not follow standard syntactic rules and thus immediately identifies a phrase as idiomatic. Consider a few German examples taken from Keil (1997: 21):

(17) *mit jdm. ist nicht gut Kirschen essen*
 "with sb._(dat) is not good cherries eat" (anomalous construction)
 'it's best not to tangle with sb.'

(18) *jdn. Lügen strafen*
 "sb._(acc) lies punish" (different subcategorization pattern of *strafen*)
 'to disprove sb.'

(19) *etw. ist nicht ganz ohne*
 "sth. is not completely without" (preposition without complement)
 'sth. is more difficult than it seemed at first glance'

(20) *einen an der Waffel haben*
 "one on the waffle have" (pronoun without antecedent)
 'not to be right in one's head'

2.2.2 Valence Structure

A syntactic feature of a verb is its valence structure. Following Keil (1997) and Burger (2003) we distinguish between internal and external arguments. An internal argument is an integral part of the idiom. Altering it would entail the loss of the idiomatic meaning. In contrast, external arguments are subcategorized for by the verb but can vary according to the context. For example:

(21) *He entirely lost his head.*

Here *head* is an internal argument of *lose*. A different direct object leads to a totally different phrase with a non-idiomatic reading:

(22) *He (*entirely) lost his wallet.*

The subject of *lose* in (21) remains external and can vary according to which person lost control of their actions.

Often an idiomatic verb has the same number of arguments as its non-idiomatic counterpart, such as in (21) and (22). However, sometimes we encounter an increase or decrease in this number. As an example of an increase, consider the German idiom

- (23) *Bauklötze staunen über etw.*
 "building-bricks goggle about sth."
 'to be flabbergasted about sth.'

The direct object (*Bauklötze*) is not part of the valence structure of *staunen* in its non-idiomatic use, so the verb has one additional argument. Another example is the following:

- (24) *jdm. mit etw. in den Ohren liegen*
 "sb._(dat) with sth. in the ears lie"
 'to solicit sb.'

The verb *liegen* ('lie') normally subcategorizes for a subject and a prepositional phrase indicating a location. In its idiomatic use this PP is internal and furthermore the verb takes a dative object and a second PP indicating the theme of solicitation.

On the basis of a different valence structure (and different theta-grids) one can see in these examples that the verb is used idiomatically. Thus, valence structure is another formal marker of idiomaticity.

Torzova (1983) reports some empirical evidence for changes in valence structure. She based her findings on an examination of 20th century German belletristic literature and compared verbs in idiomatic and free readings. Torzova found the following results:

	% of the idiomatic verbs
reduction of 1 external argument	62
reduction of 2 external arguments	6
increase in the no. of arguments	10
identical valence structure	22

2.2.3 Argument Modification

In the course of idiomatization, when a "normal" or external argument becomes an intrinsic part of an idiom (thus, an internal argument), we observe that some changes in morphosyntactic properties may occur. Consider the following example where the accusative case can be substituted by the dative.

- (25) *etw. kostete jdn./*jdm. einen Geldbetrag*
 "sth. cost sb._(acc/*dat) a amount-of-money"
 'sth. cost sb. a certain amount of money' (non-idiomatic)

- (26) *etw. kostete jdn./jdm. das Leben*
 "sth. cost sb._(acc/dat) the life"
 'sb. lost his life' (idiomatic)

Thus, we find a different case for one object, according to the lexical content of the other object. Normally, this does not happen.

2.2.4 Grammatical Properties of Arguments

CR 5 Every (phraseologically) internal argument has the same properties as an external argument.

Information about the categorial status of arguments comprises the part-of-speech and whether the argument has to be a word, a phrase or a clause. A part-of-speech has specific grammatical properties which are general and which do not have to be encoded in the lexical entries for each instance of a given part-of-speech. In German, e.g. nouns are declined and can occur together with a definite determiner, adjectives can be compared, etc. However, we can find idioms for which the normal grammatical properties seem to have been altered. For example, in (16) there is no determiner. Such irregularities have to be encoded explicitly in the lexical entry of the idiom. Otherwise, there would be a contradiction between the idiomatic and the regular behavior. The following internal arguments do not obey general grammatical regularities:

- Lexicalized nominal pairs:

(27) *(auf) Stein und Bein schwören*
 "on stone and bone swear"
 'to swear insistently'

- Prepositional phrases (in many cases the nominal complement cannot be modified):

(28) *einen Streit vom Zaun brechen*
 "a quarrel from-the fence break"
 'to start an argument'

- Complements consisting of an adjective and a prepositional phrase (idiomatized comparisons, cf. Agricola 1992: 29):

(29) *dumm wie Bohnenstroh sein*
 "stupid as bean-straw be"
 'to be very stupid'

2.2.5 Syntactic Stability

Some idioms, in contrast to non-idiomatic phrases, reveal a certain inflexibility regarding different syntactic transformations. The general picture to be drawn from the literature (e. g. Fraser 1970; Fleischer 1997) is that all idioms behave idiosyncratically. In this section we will show that at least a few properties follow from independent regular principles of grammar. We briefly mention some of these phenomena, which were also discussed by Keil (1997).

Passivization

CR 6 Every transitive VP can occur in the passive voice.

There are two dimensions one has to take into account when testing whether an idiom passivizes or not: the morphosyntactic dimension and the semantic dimension. As for the first, the verbal part of the idiom must be able to undergo passivization under non-idiomatic circumstances. For German, most transitive verbs can occur in the passive, but there are a few exceptions such as *haben* ('have/possess') and most verbs of sensory perception ('smell', 'taste', etc.) which cannot passivize. If and when these verbs occur in an idiom, this idiom is not able to undergo passivization either. Concerning the semantic dimension, the idiom must be interpreted as a transitive VP (Dobrovolskij 1999). Take the idiom *to bite the dust*, meaning 'to die'. One cannot say "*The dust was bitten by him*", because "*He died*" cannot be passivized. Sometimes, the passive is possible, but the idiomatic reading is lost (cf. "*The head was lost by him*"). There are cases of idioms with no free reading, such as the German *das Gesicht verlieren* ('to lose face'), of which a passivization would not only lead to the loss of the idiomatic reading, but to the complete loss of any sensible reading.

Nominalization

CR 7 Every verb can be nominalized (for German: along with some of its arguments).

Here we will focus on conversion – the change of part-of-speech without any change in word form. In German, this is generally possible (*laufen* ⇒ *Laufen*). In English this is possible in some cases (*to run* ⇒ *the run*). For German, some of the verb's arguments can be incorporated. Consider the following examples from Gallmann (1985, 1990):

- (30) *das So-Tun-als-ob*
 "the the so-do-as-if"
 'pretending'
- (31) *Das ist zum An-die-Decke-Gehen!*
 "that is to to-the-ceiling-go"
 'I feel like hitting the roof'

This is possible with idioms, as well. Compare (32)⁶ :

- (32) *das Handtuchwerfen*
 "the towel-throwing"
 'the giving up'

Meibauer (2003) investigated this phenomenon empirically and states the rule that the "first part" (the words that precede the verb) has to form a single constituent. This holds both for idioms and for non-idiomatic utterances.

⁶"*Das Handtuchwerfen ist seine Sache nicht, nie gewesen.*" (Die Presse, 17.10.1998).

Imperatives

CR 8 Every declarative sentence can occur in the imperative mood.

For imperatives, the same restrictions hold as for non-idiomatic utterances: Forming an imperative must be syntactically possible⁷ and an imperative must be pragmatically well-formed. Sometimes, a positive imperative does not make sense; the examples without *never* or *not* are awkward.

(33) *Gerate *(niemals) in Verruf!*
 "get never into discredit"
 '*(never) lose your good name'

(34) *Lege *(nicht) jedes Wort auf die Goldwaage!*
 "lay not each word on the assay-balance"
 'do *(not) mince words'

Idiosyncratic behavior can be found for some idioms whose paraphrase is felicitous in the imperative mode, whereas the idiom is not.

(35) ? *Schneide ihr den Lebensfaden ab*
 "cut her the life-thread away"
 'kill her'

Questions

CR 9 Every argument of a verb can be replaced by a *wh*-expression.

Exceptions to this criterion of regularity are idioms whose parts are not referential.

(36) *jdn. übers Ohr hauen*
 "sb._(acc) over-the ear hit"
 'to cheat on sb.'

The NP *Ohr* is not referential. Questions such as "*Above what did she hit him?*" (– "*The knee.*") or "*Above which ear did she hit him?*" (– "*The left ear.*") are grammatical but do not have an idiomatic reading.

⁷Syntactically anomalous idioms cannot occur in the imperative mode (Keil 1997: 26).

Negation

CR 10 Every declarative sentence can be negated.

In German, the negation of a proposition can be achieved in two ways: using *nicht* ('not') or *kein* (negative indefinite article). Due to space limitations we do not go into details here but merely state that the behavior of these negatives is the same in both idiomatic and non-idiomatic contexts. However, if an idiom already contains negation, adding a negative to it results in a double-negated or unacceptable utterance. In the German idiom *einen Teufel tun* (*to do a devil* – 'not to do sth.')

there is an intrinsic negation in the meaning. An "additional" overt negation with *nicht* or *nie* ('never') is ungrammatical. Interestingly though, there is a negative concord effect with *niemand* ('nobody'): "*Niemand tut einen Teufel ihr zu helfen.*" 'nobody does anything to help her'.

The occurrence of the negative indefinite article *kein* is always possible wherever an indefinite article may occur. If this is not the case, *kein* is not felicitous: Either the idiomatic reading is lost or the utterance is ungrammatical.

(37) *Sie hat die Flinte nicht ins Korn geworfen.*
 "she has the gun not into-the crop thrown"
 'she has not given up'

(38) *Sie hat keine Flinte ins Korn geworfen.*
 "she has not-a gun into-the crop thrown"
 'she has not thrown a gun into the crop' (only non-idiomatic reading)

(39) *Er hat noch nicht das Zeitliche gesegnet.*
 "he has yet not the here and now blessed"
 'he has not died yet'

(40) **Er hat noch kein Zeitliches gesegnet.*

An idiosyncratic property of most idioms in which there is no article at all is that *kein* cannot occur.

(41) *jdn. in (*keinen) Misskredit bringen*
 "sb._(acc) in *not-a discredit bring"
 'to discredit sb.'

Modification

CR 11 Every NP can be semantically modified.

Some idioms do not behave according to this criterion:⁸

⁸We differentiate between modification which is possible within the language system, and occasional modification (cf. Burger 2003; Wotjak 1992; Sabban 1998). Speakers often alter or modify idioms in some way or another in order to achieve a special effect which attracts the attention of the listener. We ignore this kind of modification here.

- (42) *am* (*großen/*dünnen/...) *Hungertuch nagen*
 "on-the (*big/*thin/...) hunger-cloth gnaw"
 'to be impoverished'

Semantic interpretation plays a role, as well. For example, one can *quickly* or *unexpectedly bite the dust*, but it is impossible to *firmly bite the dust* or to *bite the settled dust* understood idiomatically. Thus, the modification of idioms is quite predictable on the basis of semantics but is still very idiosyncratic.

2.3 Summary

Summing up this section, we have presented several criteria of regularity and examined to what extent these are violated by idioms. The conclusion to be drawn is that idioms exhibit a great deal of idiosyncrasies which justifies an analysis within a lexicalist framework (see next section). However, the claim that idioms do not behave subject to CRs at all must be rejected. In fact, the immunity of idioms relative to certain constructions is in part due to independent factors.

3 An Analysis

An analysis which can cope with the data has to meet two demands. On the one hand, it must guarantee the co-occurrence of all idiom parts. On the other hand, it must be flexible enough to include all possible changes (modification, transformation, etc.) that follow from independent factors. Thus, it would not be a good idea to encode an idiom such as *spill the beans* as an unalterable string whereas one might arguably encode *The early bird catches the worm* as a fixed⁹ phrase. Such an analysis was presented in Soehn (2004a, b, 2006) within the framework of "Head-Driven Phrase Structure Grammar" (HPSG, cf. Pollard/Sag 1994), a constraint-based, lexicalist approach to grammatical theory which models human languages as systems of constraints. Due to space limitations we will not explain the analysis in formal detail but will attempt to convey the idea behind the analysis.

3.1 Listemes

There is a well-established relationship between the verb and its complements, which is called selection or subcategorization. Normally a verb's argument structure contains quite general information about the part-of-speech of a complement or some sortal and selectional restrictions. For HPSG, Krenn/Erbach (1994) refine the selection mechanism to subcategorize for specific lexemes that can handle idiomatic expressions. Technically and conceptually enhanced, this approach was adopted in Soehn (2006) where a feature¹⁰ LISTEME was introduced, following the idea of Di Sciullo/Williams (1988).

⁹Irrespective of the great variety of ways in which speakers use this proverb.

¹⁰HPSG grammars use feature structures, often written as attribute-value-matrices (AVMs), to represent grammar principles, grammar rules and lexical entries. Every feature structure is of a certain type and contains attributes (features) which have values, which are, in turn, feature structures. A constituent is licensed if it is described by a feature structure and this feature structure conforms to each grammatical principle.

CR 12 Every element in a phrase occurs in the same meaning in some other combination.

The idiomatic meaning of words or phrases, which are parts of decomposable idioms¹¹, is encoded in a separate lexical entry existing in addition to the one with the literal meaning (thus, the latter meets CR 12). Each lexical item has a unique value for its feature LISTEME and thus the selection mechanism can be applied in a more fine-grained manner. For example, the idiomatic verb *spill* (meaning 'divulge') can select its complement *the beans* via its LISTEME value.

Being able to exactly distinguish between lexical elements serves another important purpose. Thereby, we can exclude idiomatic verbs from being subject to certain transformations such as passivization. No matter if passivization is achieved by a lexical rule (cf. Sailer 2003: 96) or by specific subcategorization frames of the auxiliary (cf. Müller 2002, 2003), one can explicitly exclude certain listemes.

3.2 The Feature COLL

Having explored one direction of listeme-listeme-co-occurrence, we still need a way to specify distributional idiosyncrasies of non-heads. We have shown that a verb can subcategorize for its idiomatic complements, but we must also restrict the occurrence of the latter. For example, how can a nominal idiom part "select" a certain verb? It somehow must impose a certain LISTEME value on the verb by which it is selected. There are several possibilities, such as external selection by a new selection feature (explored in Soehn 2003) or a new collocation mechanism as described by Richter/Sailer (1999a,b)¹². We prefer to adopt the second option and propose a collocation module using the feature COLL (Context Of Lexical Licensing).

In the value of COLL, the required licensing context of a lexical element is specified. Technically, the value of COLL is a list which can contain certain elements, the so-called barriers. Barriers are phrases which dominate the lexical element in question. They form the minimal context (a PP or VP containing the element) on which the element imposes restrictions. Certain (i. e. local) properties of a barrier can be specified via the lexical entry.

Take, for example, the entry of *the beans* ('the secret'), a part of the idiom already mentioned. The COLL list¹³ comprises one element, a VP-barrier. This VP-barrier is required to have the LISTEME value *spill*. Thereby we define *the beans* as only being allowed to occur within a VP whose head is an instance of a *spill*-listeme. Additionally, we introduce a principle of grammar (Licensing Principle, LIP) that licenses linguistic signs only if they occur in the specified context. More precisely, if for a lexical element there is a barrier specified in COLL,

¹¹ I. e. the overall meaning of such an idiom can be computed from its parts.

¹²This theory was further developed in Sailer (2003) but our version of the mechanism (Soehn 2006) is more restrictive.

¹³Short for "the list which is the value of COLL".

there must be a phrase in the actual utterance which has all of the properties defined for the barrier.

Why do we need a list of barriers as the value of COLL? Sometimes it does not suffice to define only one barrier, because the licensing context is more complex. Consider the German idiom *zu Potte kommen* ("to pot come" – 'to get going' / 'to get through'). The noun *Potte* needs to occur within a PP headed by *zu*. This PP is in turn the complement of the verb *kommen*. Because *Potte* can be regarded as the only idiosyncratic item in this idiom, it encodes both criteria on its COLL list: one PP-barrier with the LISTEME value *zu* and one VP-barrier with the LISTEME value *kommen*. See Figure 1 for an illustration.

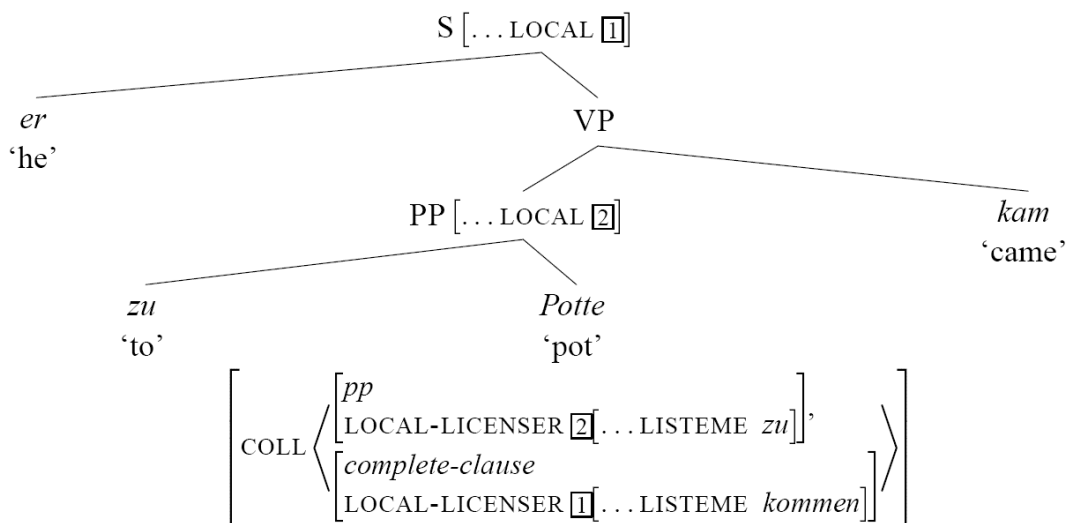


Figure 1: Example of COLL — The LIP guarantees the identities in [1] and [2]

Unaffected by our collocation module are the other parts of grammar, such as verb placement (cf. Müller 2005), word-order, or the mechanism for selection. The COLL module is flexible enough to be able to handle syntactic changes, in particular to allow for idiom parts to topicalize or to be modified. Passivization or nominalization can be prohibited via the LEXEME value (see above).

3.3 Phrasal Lexical Entries

CR 13 The meaning of an entire phrase is arrived at by combining the meanings of its parts in a regular way.

So far, we have implicitly dealt only with decomposable idioms, where each part has a separate lexical entry. Non-decomposable idioms such as *kick the bucket* or *das Handtuch werfen* ('to give up' – *throw in the towel*) have to be handled differently (for the concept of compositionality, cf. Rajchštejn 1980; Burger/Buhofer/Sialm 1982; Gibbs et al. 1989; Nunberg/Sag/Wasow 1994; Geeraerts/Bakema 1993). As their meanings cannot be computed from the parts, the whole idiomatic phrase is encoded in a lexical entry. These phrasal lexical

entries (PLEs, cf. Sailer 2003) license phrases which are not subject to general rules of grammar. For example, since an idiomatic phrase with its own meaning has different semantics, a different LISTEME value and other idiosyncratic properties, neither the HEAD-FEATURE-PRINCIPLE¹⁴ nor the SEMANTICS PRINCIPLE¹⁵ holds. Thus, every single value must be specified directly, providing a lot of freedom to encode all idiosyncratic behavior. In the example "*The coach threw in the towel.*", the leaves of the structure-tree are licensed by the non-idiomatic lexical entries of the words. At the level of the VP, the meaning of the non-idiomatic phrase can be substituted by an idiomatic one, which is licensed by the phrasal lexical entry. Thus, the difference between the analysis of decomposable idioms and the analysis of non-decomposable idioms is the following: In the former case we define new lexical entries for the idiom parts (word level) and combine the idiomatic reading in a compositional way, whereas in the latter case, we define a whole phrase (with its idiomatic meaning) in the lexicon (phrase level) and for the words we resort to the lexical entries which already exist.

We define PLEs in such a way that they also have a non-empty list as their value of COLL. This allows PLEs to be excluded from regular principles of grammar. Furthermore, there are several cases of non-decomposable idioms which reveal a distributional idiosyncrasy, thus a specific barrier is required. Consider:

- (43) *sich freuen* [*wie ein Schneekönig*]
 "refl-pron rejoice as a snow-king"
 'to be very glad'

The prepositional phrase is non-decomposable (meaning 'very') and is restricted to the verb *sich freuen*. Another example of a non-decomposable and highly idiosyncratic idiom is *wissen, wo Bartel den Most holt* (corresponds to *to know which side one's bread is buttered on*), discussed by Sailer (2004).

Thus, the purpose of COLL can be characterized as twofold. Firstly, the COLL list is the locus in which distributional idiosyncrasies can be encoded. Secondly, a non-empty COLL list is an indication that the respective sign is lexical and that one cannot trace all properties back to general rules. Hence, the lexicon only contains descriptions of elements with non-empty COLL lists. All regularly built phrases, as well as the output of lexical rules, have predictable properties and as a consequence, their COLL value is the empty list.

The grammar formalism allows us to create PLEs which are "flexible" enough to accommodate topicalization or the verb placement mechanism (cf. Soehn 2006 for details). Such complex conditions are expressed by recursive relations¹⁶ over feature structures.

¹⁴This principle enforces the identity of the HEAD values along the syntactic projection line.

¹⁵This principle is responsible for the correct combination of each part of a regular phrase to obtain the overall meaning.

¹⁶For relations in another context cf. Richter.

4 Idioms as Negative Polarity Items

Polarity items are lexical or phrasal units which can only occur in either negative contexts (negative PIs, NPIs) or in non-negative contexts (positive PIs, PPIs), whereby the notion of negative context must be more precisely defined. Thus, polarity items are licensed (or triggered, cf. van der Wouden 1997: 60) by their contexts, or conversely, their distribution has to be restricted.

According to the most elaborate theory of distribution that currently exists, NPEs are licensed when they are in the scope of a monotone decreasing operator (cf. Ladusaw 1980; Zwarts 1997). These operators can be classified according to their inference behaviour. Zwarts (1995, 1997) developed a hierarchy of different degrees of "negativity", whereby various subclasses of NPIs can be distinguished. Besides this inference theory there are also syntactic approaches (Progovac 1994) and attempts to trace polarity back to pragmatic factors (Krifka 1995). All of these theories cover only a subset of the NPIs of a given language – at least, it is often unclear how to generalize the approach to include all polarity items. In the literature there is also a discussion of whether negativity or non-veridicality (Giannakidou 1998) is the more suitable term for characterizing NPI-licensing contexts. Our aim here is only to depict a way to implement polarity into a lexicalist theory and to show how to guarantee the occurrence of an NPI only in the right contexts.

In the introduction we gave some examples of idioms which are NPIs (repeated below).

- (44) **(Don't) put all your eggs in one basket.*
 (45) *Nobody/*someone lifted a finger to help her.*
 (46) *I can*(not) make head or tail of it.*

For German, we find the following idioms (as well as many others):

- (47) *Er macht aus seiner Meinung keinen/*einen Hehl.*
 "he makes out-of his opinion not-a/*a secret"
 'he makes no secret of his opinion'
 (48) *Sie hat kein/*ein Blatt vor den Mund genommen.*
 "she has not-a/*a leaf in front of the mouth taken"
 'she was very outspoken about it'
 (49) *Mit dieser Technik kann niemand/*jemand einen Blumentopf gewinnen.*
 "with this method can nobody/*somebody a flower pot win"
 'with this method nobody will win any prizes'

Here, we must cope not only with the co-occurrence of different listemes but also with phenomena on the level of logical form. In Soehn (2006) the semantic module LRS ("Lexical Resource Semantics", cf. Richter/Sailer 2004) has been adopted, in which the logical form of a phrase including scope relations is available on phrasal level via the feature LF EXTERNAL-CONTENT. It is not difficult to extend our collocation module to cope with restrictions on logical form: We introduce another feature, LF-LICENSER (in addition to LOCAL-LICENSER), which is appropriate for all barriers. To take example (47), the noun *Hehl* would constrain its context in the following way: the head of the VP in which it occurs must have *machen* as its

LISTEME value and the external content of the clause in which it occurs must contain a negative operator¹⁷ having scope over it. The COLL value can be described as in Figure 2, whereas $\boxed{1}$ refers to the CONTENT MAIN value of *Hehl*, the genuine semantic contribution of that word.

The value of EXTERNAL-CONTENT has to be this general because it must be compatible with the following variants of (47).

- (50) *Wenige der Beteiligten machen aus ihrer Meinung einen Hehl.*
 "few of-the involved make out-of their opinion a secret"
 'Few of the persons involved make a secret of their opinion.'
- (51) *Niemand hier macht aus seiner Meinung einen Hehl.*
 "nobody here makes out-of his opinion a secret"
 'None of us makes a secret of his opinion.'
- (52) *Er macht aus seiner Meinung nie einen Hehl.*
 "he makes out-of his opinion never a secret"
 'He never makes a secret of his opinion.'

Examples such as (45) and (49) are so-called minimizers. Further instances are *(not) to budge an inch* or *kein Sterbenswörtchen sagen* (*not-a dying-word say* – 'not to say a word'). One might say that it follows from their semantics that minimizers are always constructed with a negation and that there is no need to encode this separately in the lexicon. However, as Vallduví (1994) points out for minimizers in Catalan and Spanish, there are different types of minimizers. Some of them go without a negation but still have some idiosyncratic demands on their contexts.

To conclude, we can model the distributional idiosyncrasy on the semantic level. Certainly, the exact specification of the value of EXTERNAL-CONTENT has to be explored thoroughly for each lexical item in question.

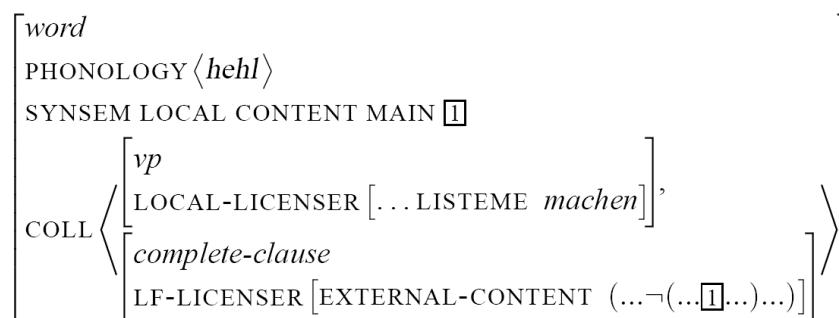


Figure 2: Part of the lexical entry of *Hehl*

¹⁷At first glance this seems to be a crude generalization, but note that the representations of most NPI licensing operators can be decomposed in such a way that (a) a negation symbol is introduced, and (b) the different degree of 'licensing strength' can be related to the kinds of operators which occur in the decomposed representation. For example: *not* = $\neg(\dots\phi\dots)$; *nobody* = $\neg \exists x(\dots\phi\dots)$; *few* = $\neg \textit{many}'(\dots\phi\dots)$ (cf. Sailer/Richter 2002) or *never* = $\neg \exists t(\dots\phi\dots)$.

5 Summary

In this article we presented some properties of idioms. Most of them can be seen as "idiom markers" because they mark a given phrase as idiomatic. The occurrence of bound words, morphological or syntactic anomalies and a differing valence structure stand out in this respect. The lexical licensing mechanism which was developed in Soehn has been sketched and extended to a different kind of co-occurrence which lies on the level of logical form. Further research will be carried out in order to gather and classify German NPIs and to more precisely specify their licensing context.

References

- Agricola, Erhard (ed.) (1992): *Wörter und Wendungen. Wörterbuch zum deutschen Sprachgebrauch*. Revision of 14th ed. Mannheim.
- Aronoff, Mark (1976/1985): *Word Formation in Generative Grammar*. 3rd print. Cambridge, MA. (=Linguistic Inquiry Monographs; 1).
- Burger, Harald (2003): *Phraseologie. Eine Einführung am Beispiel des Deutschen*. 2nd, revised ed. Berlin. (=Grundlagen der Germanistik 36).
- Burger, Harald/Buhofer, Annelies/Sialm, Ambros (1982): *Handbuch der Phraseologie*. Berlin.
- Di Sciullo, Anna-Maria/Williams, Edwin (1988): *On the Definition of Word*. 2nd ed. Cambridge, MA. (=Linguistic Inquiry Monographs 14).
- Dobrovolskij, Dmitrij (1988): *Phraseologie als Objekt der Universalienlinguistik*. Leipzig. (=Linguistische Studien).
- Dobrovolskij, Dmitrij (1999): "Gibt es Regeln für die Passivierung deutscher Idiome?" In: Bäcker, Iris (ed.): *Das Wort. Germanistisches Jahrbuch*. Bonn.
- Dobrovolskij, Dmitrij/Piirainen, Elisabeth (1994): "Sprachliche Unikalia im Deutschen: Zum Phänomen phraseologisch gebundener Formative". *Folia Linguistica* XXVIII/3-4: 449–473.
- Fleischer, Wolfgang (1997): *Phraseologie der deutschen Gegenwartssprache*. 2nd, revised ed. Tübingen.
- Fraser, Bruce (1970): "Idioms within a Transformational Grammar". *Foundations of Language* 6: 22–42.
- Gallmann, Peter (1985): *Graphische Elemente der geschriebenen Sprache*. Tübingen. (=Germanistische Linguistik 60).
- Gallmann, Peter (1990): *Kategoriell komplexe Wortformen. Das Zusammenwirken von Morphologie und Syntax bei der Flexion von Nomen und Adjektiv*. Tübingen: Max Niemeyer Verlag. (=Germanistische Linguistik; 108).
- Geeraerts, Dirk/Bakema, Peter (1993): "De prismatische semantiek van idiomen en composita". *Leuvense Bijdragen* 82: 185–226.
- Giannakidou, Anastasia (1998): *Polarity Sensitivity as (Non)veridical Dependency*. Amsterdam/Philadelphia.
- Gibbs, Raymond W. JR. et al. (1989): "Speakers' assumptions about the lexical flexibility of idioms". *Memory & Cognition* 17(1): 58–68.

- Keil, Martina (1997): *Wort für Wort. Repräsentation und Verarbeitung verbaler Phraseologismen*. Tübingen. (=Sprache und Information; 35).
- Krenn, Brigitte/Erbach, Gregor (1994): "Idioms and Support Verb Constructions". In: Nerbonne, John/Netter, Klaus/Pollard, Carl (eds.): *German in Head-Driven Phrase Structure Grammar*. Stanford: 365–396. (=Lecture Notes 46).
- Krifka, Manfred (1995): "The Semantics and Pragmatics of Weak and Strong Polarity Items". *Linguistic Analysis* 25: 209–257.
- Ladusaw, William (1980): *Polarity Sensitivity as Inherent Scope Relations*. New York.
- Meibauer, Jörg (2003): "Phrasenkomposita zwischen Wortsyntax und Lexikon". *Zeitschrift für Sprachwissenschaft* 22.2: 153–188.
- Müller, Stefan (2002): *Complex Predicates: Verbal Complexes, Resultative Constructions, and Particle Verbs in German*. Stanford. (=Studies in Constraint-Based Lexicalism 13). <http://www.cl.uni-bremen.de/stefan/Pub/complex.html.de>.
- Müller, Stefan (2003): "Object-To-Subject-Raising and Lexical Rule. An Analysis of the German Passive". In: Müller, Stefan (ed.): *Proceedings of the HPSG-2003 Conference, Michigan State University, East Lansing*. Stanford: 278–297. <http://csli-publications.stanford.edu/HPSG/4/>.
- Müller, Stefan (2005): "Zur Analyse der scheinbar mehrfachen Vorfeldbesetzung". *Linguistische Berichte* 203: 29–62.
- Nunberg, Geoffrey/Sag, Ivan A./Wasow, Thomas (1994): "Idioms". *Language* 70: 491–538.
- Pollard, Carl/Sag, Ivan A. (1994): *Head-Driven Phrase Structure Grammar*. Chicago. (=Studies in Contemporary Linguistics; 68).
- Progovac, Ljiljana (1994): *Negative and Positive Polarity. A binding approach*. Cambridge. (=Cambridge Studies in Linguistics).
- Rajchštejn, Aleksandr D. (1980): *Sopostavitel'nyj analiz nemeckoj i ruskoj frazeologii*. Moskow.
- Richter, Frank (1997): "Die Satzstruktur des Deutschen und die Behandlung langer Abhängigkeiten in einer Linearisierungsgrammatik. Formale Grundlagen und Implementierung in einem HPSG-Fragment". In: Hinrichs, Erhard W. et al. (eds.): *Ein HPSG-Fragment des Deutschen, Teil 1: Theorie*. Tübingen: 13–187. (=Arbeitspapiere des SFB 34095).
- Richter, Frank/Sailer, Manfred (1999a): "LF Conditions on Expressions of Ty2: An HPSG Analysis of Negative Concord in Polish." In: Borsley, Robert D./Przepiórkowski, Adam (eds.): *Slavic in Head-Driven Phrase Structure Grammar*. Stanford: 247–282.
- Richter, Frank/Sailer, Manfred (1999b): "A lexicalist collocation analysis of sentential negation and negative concord in French". In: Kordoni, Valia (ed.): *Tübingen Studies in Head-Driven Phrase Structure Grammar*. Tübingen: 231–300. (=Arbeitspapiere des SFB 340, Nr. 1321).
- Richter, Frank/Sailer, Manfred (2004): "Basic Concepts of Lexical Resource Semantics". In: Beckmann, Arnold/Preining, Norbert (eds.): *ESSLLI 2003 – Course Material I, Collegium Logicum 5*. Vienna: 87–143.

- Sabban, Annette (1998): *Okkasionelle Variationen sprachlicher Schematismen. Eine Analyse französischer und deutscher Presse- und Werbetexte*. Tübingen. (=Romanica Monacensia 53).
- Sailer, Manfred (2003): *Combinatorial Semantics and Idiomatic Expressions in Head-Driven Phrase Structure Grammar*. Phil. Dissertation (2000). Tübingen. (=Arbeitspapiere des SFB 340161).
- Sailer, Manfred (2004): "Distributionsidiosynkrasien. Korpuslinguistische Erfassung und grammatiktheoretische Deutung". In: Steyer, Kathrin (ed.), *Wortverbindungen – mehr oder weniger fest*. Berlin/New York: 194–221.
- Sailer, Manfred/Richter, Frank (2002): "Not for Love or Money: Collocations!" In: Jäger, Gerhard et al. (eds.): *Proceedings of Formal Grammar 2002*. Stanford: 149–160.
- Soehn, Jan-Philipp (2003): *Von Geisterhand zu Potte gekommen. Eine HPSG-Analyse von PPs mit unikaler Komponente*. Masters Thesis. Tübingen.
<http://www.sfs.uni-tuebingen.de/hpsg/archive/bibliography/papers/majp.ps>.
- Soehn, Jan-Philipp (2004a): "About Spilled Beans and Shot Breezes. A New Wordlevel Approach to Idioms." In: Jäger, Gerhard et al. (eds.): *Proceedings of Formal Grammar 2004*. Stanford: 125–140.
- Soehn, Jan-Philipp (2004b): "License to COLL". In: Müller, Stefan (ed.): *Proceedings of the HPSG-2004 Conference, Center for Computational Linguistics, Katholieke Universiteit Leuven*. Stanford: 261–273. <http://csli-publications.stanford.edu/HPSG/5/>.
- Soehn, Jan-Philipp (2006): *Über Bärendienste und erstaunte Bauklötze – Idiome ohne freie Lesart in der HPSG*. Phil. Dissertation (2005). Friedrich-Schiller-Universität Jena. Frankfurt a. M. etc.. (=Deutsche Sprache und Literatur 1930).
- Torzova, M. V. (1983): "Zur Valenz der Phraseologismen". *Deutsch als Fremdsprache* 5: 283–287.
- Vallduví, Enric (1994): "Polarity Items, n-words, and minimizers in Catalan and Spanish". *Probus* 6: 263–294.
- van der Wouden, Ton (1997): *Negative Contexts. Collocation, polarity and multiple negation*. London/New York.
- Wotjak, Barbara (1992): *Verbale Phraseolexeme in System und Text*. Tübingen. (=Germanistische Linguistik 125).
- Zwarts, Frans (1995): "Nonveridical Contexts". *Linguistic Analysis* 25: 286–312.
- Zwarts, Frans (1997): "Three Types of Polarity". In: Hamm, Fritz/Hinrichs, Erhard W. (eds.): *Plurality and Quantification*. Dordrecht: 177–237.