Chapter 17

Idioms

Manfred Sailer
Goethe-Unversität Frankfurt

This chapter first sketches basic empirical properties of idioms. The state of the art before the emergence of HPSG is presented, followed by a discussion of four types of HPSG approaches to idioms. A section on future research closes the discussion.

1 Introduction

In this chapter, I will use the term *idiom* interchangeably with the broader terms such as *phraseme*, *phraseologism*, *phraseological unit*, or *multiword expression*. This means, that I will subsume under this notion expressions such as prototypical idioms (*kick the bucket* 'die'), support verb constructions (*take advantage*), formulaic expressions (*Good morning!*), and many more.[1] The main focus of the discussion will, however, be on prototypical idioms, as these have been in the center of the theoretical development.

I will sketch some empirical aspects of idioms in Section 2. In Section 3, I will present the theoretical context within which idiom analyses arose in HPSG. An overview of the development within HPSG will be given in Section 4. Desiderata for future research are mentioned in Section 5, before I close with a short conclusion.

2 Empirical domain

In the context of the present handbook, the most useful characterization of idioms might be the definition of *multiword expression* from Baldwin & Kim (2010).

[1] I will provide a paraphrase for all idioms at their first mention. They are also listed in the appendix, together with their paraphrase and a remark on which aspects of the idiom are discussed in the text.

Manfred Sailer. 2020. Idioms. In Stefan Müller, Anne Abeillé, Robert D. Borsley & Jean-Pierre Koenig (eds.), *Head-Driven Phrase Structure Grammar: The handbook*, 733–764. Berlin: Language Science Press. DOI: ??

For them, any combination of words counts as a multiword expression if it is syntactically complex and shows some degree of *idiomaticity* (i.e., irregularity), be it lexical, syntactic, semantic, pragmatic, or statistical.[2] I speak of a "combination of words" in the sense of a *substantive* or *lexically filled idiom*, which contrasts with *formal* or *lexically open idioms* (Fillmore et al. 1988: 505).

Baldwin & Kim's criteria can help us structure the data presentation in this section, expanding their criteria where it seems suitable. My expansions concern the aspect known as *fixedness* in the phraseological tradition as in Fleischer (1997).[3]

For Baldwin & Kim (2010), *lexical idiosyncrasy* concerns expressions with words that only occur in an idiom, so-called *phraseologically bound words*, or *cranberry words* (Aronoff 1976). Examples include *make headway* 'make progress', *take umbrage* 'take offence', *in a trice* 'in a moment/very quickly'.[4] For such expressions, the grammar has to make sure that the bound word does not occur outside the idiom, i.e., we need to prevent combinations such as (1b).[5]

(1) a. They fixed the problem in a trice.
b. * It just took them a trice to fix the problem.

We can expand this type of idiosyncrasy to include a second important property of idioms. Most idioms have a fixed inventory of words. In their summary of this aspect of idioms, Gibbs & Colston (2007: 827–828) include the following examples: *kick the bucket* means 'die', but *kick the pail*, *punt the bucket*, or *punt the pail* do not have this meaning. However, some degree of lexical variation seems to be allowed, as the idiom *break the ice* 'relieve tension in a strained situation' can be varied into *shatter the ice*.[6] So, a challenge for idiom theories is to guarantee that the right lexical elements are used in the right constellation.

[2] In the phraseological tradition, the aspect of *lexicalization* is added (Fleischer 1997; Burger 1998). This means that an expression is stored in the lexicon. This criterion might have the same coverage as *conventionalization* used in Nunberg et al. (1994). These criteria are addressing the mental representation of idioms as a unit and are, thus, rather psycholinguistic in nature.

[3] Baldwin & Kim (2010) describe idioms in terms of syntactic fixedness, but they seem to consider fixedness a derived notion.

[4] See https://www.english-linguistics.de/codii/, accessed 2019-09-03, for a list of bound words in English and German (Trawiński et al. 2008).

[5] Tom Wasow (p.c.) points out that there are attested uses of many alleged bound words outside their canonical idiom, as in (i). Such uses are, however, rare and restricted.

(i) Not a trice later, the sounds of gunplay were to be heard echoing from Bad Man's Rock. (COCA)

[6] While Gibbs & Colston (2007), following Gibbs et al. (1989), present this example as a lexical

Syntactic idiomaticity is used in Baldwin & Kim (2010) to describe expressions that are not formed according to the productive rules of English syntax, following Fillmore et al. (1988), such as *by and large* 'on the whole/everything considered', *trip the light fantastic* 'dance'.

In my expanded use of this notion, this also subsumes irregularities/restrictions in the syntactic flexibility of an idiom, i.e., the question whether an idiom can occur in the same syntactic constructions as an analogous non-idiomatic combination. In Transformational Grammar, such as Weinreich (1969) and Fraser (1970), lists of different syntactic transformations were compiled and it was observed that some idioms allow for certain transformations but not for others. This method has been pursued systematically in the framework of *Lexicon-Grammar* (Gross 1982).[7] Sag et al. (2002) distinguish three levels of fixedness: *fixed*, *semi-fixed*, and *flexible*. Completely fixed idioms include *of course*, *ad hoc* and are often called *words with spaces*. Semi-fixed idioms allow for morphosyntactic variation such as inflection. These include some prototypical idioms (*trip the light fantastic*, *kick the bucket*) and complex proper names. In English, semi-fixed idioms show inflection, but cannot easily be passivized, nor do they allow for parts of the idiom being topicalized, see (2).

(2) a. Alex kicked / might kick the bucket.
b. * The bucket was kicked by Alex.
c. * The bucket, Alex kicked.

Flexible idioms pattern with free combinations. For them, we do not only find inflection, but also passivization, topicalization, pronominalization of parts, etc. Free combinations include some prototypical idioms (*spill the beans* 'reveal a secret', *pull strings* 'exert influence/use one's connections'), but also collocations (*brush one's teeth*) and light verbs (*make a mistake*).

The assumption of two flexibility classes is not uncontroversial: Horn (2003) distinguishes two types among what Sag et al. (2002) consider flexible idioms. Fraser (1970) assumes six flexibility classes, looking at a wide range of syntactic operations. Ruwet (1991) takes issue with the cross-linguistic applicability of the classification of syntactic operations. Similarly, Schenk (1995) claims that for languages such as Dutch and German, automatic/meaningless syntactic processes

variation, Glucksberg (2001: 85), from which it is taken, characterizes it as having a somewhat different aspect of an "abrupt change in the social climate". Clear cases of synonymy under lexical substitution are found with German *wie warme Semmeln/Brötchen/Schrippen weggehen* (lit.: like warm rolls vanish) 'sell like hotcakes' in which some regional terms for rolls can be used in the idiom.

[7] See Laporte (2018) for a recent discussion of applying this method for a classification of idioms.

other than just inflection are possible for semi-fixed idioms, such as verb-second movement and some types of fronting.

The analytic challenge of syntactic idiomaticity is to capture the difference in flexibility in a non-ad hoc way. It is this aspect of idioms that has received particular attention in Mainstream Generative Grammar (MGG),[8,9] but also in the HPSG approaches sketched in Section 4.

Semantic idiomaticity may sound pleonastic, as, traditionally, an expression is called idiomatic if it has a conventional meaning that is different from its literal meaning. Since I use the terms idiom and idiomaticity in their broad senses of phraseological unit and irregularity, respectively, the qualification *semantic* idiom(aticity) is needed.

One challenge of the modelling of idioms is capturing the relation between the literal and the idiomatic meaning of an expression. Gibbs & Colston (2007) give an overview of psycholinguistic research on idioms. Whereas it was first assumed that speakers would compute the literal meaning of an expression and then derive the idiomatic meaning, evidence has been accumulated that the idiomatic meaning is accessed directly.

Wasow et al. (1984) and Nunberg et al. (1994) explore various semantic relations for idioms, in particular *decomposability* and *transparency*. An idiom is *decomposable* if its idiomatic meaning can be distributed over its component parts in such a way that we would arrive at the idiomatic meaning of the overall expression if we interpreted the syntactic structure on the basis of such a meaning assignment. The idiomatic meaning of the expression *pull strings* can be decomposed by interpreting *pull* as *exploit/use* and *strings* as *connections*. The expressions *kick the bucket* and *saw logs* 'snore' are not decomposable.

An idiom is *transparent* if there is a synchronically accessible relation between the literal and the idiomatic meaning of an idiom. For some speakers, *saw logs* is transparent in this sense, as the noise produced by this activity is similar to a snoring noise. For *pull strings*, there is an analogy to a puppeteer controlling the puppets' behavior by pulling strings. A non-transparent idiom is called *opaque*.

Some idioms do not show semantic idiomaticity at all, such as collocations (*brush one's teeth*) or support verb constructions (*take a shower*). Many body-part expressions such as *shake hands* 'greet' or *shake one's head* 'decline/negate' constitute a more complex case: They describe a conventionalized activity and

[8]I follow Culicover & Jackendoff (2005: 3) in using the term *Mainstream Generative Grammar* to refer to work in Minimalism and the earlier Government & Binding framework.

[9]See the references in Corver et al. (2019) for a brief up-to-date overview of MGG work.

denote the social meaning of this activity.[10]

In addition, we might need to assume a *figurative* interpretation. For some expressions, in particular proverbs or cases like *take the bull by the horns* 'approach a problem directly') we might get a figurative reading rather than an idiomatic reading. Glucksberg (2001) explicitly distinguishes between idiomatic and figurative interpretations. In his view, the above-mentioned case of *shatter the ice* would be a figurative use of the idiom *break the ice*. While there has been a considerable amount of work on figurativity in psycholinguistics, the integration of its results into formal linguistics is still a desideratum.

Pragmatic idiomaticity covers expressions that have a *pragmatic point* in the terminology of Fillmore et al. (1988). These include complex formulaic expressions (*Good morning!*). There has been little work on this aspect of idiomaticity in formal phraseology.

The final type of idiomaticity is *statistical idiomaticity*. Contrary to the other idiomaticity criteria, this is a usage-based aspect. If we find a high degree of co-occurrence of a particular combination of words that is idiosyncratic for this combination, we can speak of a statistical idiomaticity. This category includes *collocations*. Baldwin & Kim (2010) mention *immaculate performance* as an example. Collocations are important in computational linguistics and in foreign-language learning, but their status for theoretical linguistics and for a competence-oriented framework such as HPSG is unclear.

This discussion of the various types of idiomaticity shows that idioms do not form a homogeneous empirical domain but rather are defined negatively. This leads to the basic analytical challenge of idioms: while the empirical domain is defined by absence of regularity in at least one aspect, idioms largely obey the principles of grammar. In other words, there is a lot of regularity in the domain of idioms, while any approach still needs to be able to model the irregular properties.

3 Predecessors to HPSG analyses of idioms

In this section, I will sketch the theoretical environment within which HPSG and HPSG analyses of idioms have emerged.

The general assumption about idioms in MGG is that they must be represented as a complex phrasal form-meaning unit. Such units are inserted *en bloc* into the structure rather than built by syntactic operations. This view goes back to Chom-

[10] The basic reference for the phraseological properties of body-part expressions is Burger (1976).

sky (1965: 190). With this unquestioned assumption, arguments for or against particular analyses can be constructed. To give just one classical example, Chomsky (1981) uses the passivizabilty of some idioms as an argument for the existence of Deep Structure, i.e., a structure on which the idiom is inserted holistically. Ruwet (1991) and Nunberg et al. (1994) go through a number of such lines of argumentation showing their basic problems.

The holistic view on idioms is most plausible for idioms that show many types of idiomaticity at the same time, though it becomes more and more problematic if only one or a few types of idiomaticity are attested. HPSG is less driven by analytical pre-decisions than other frameworks; see Borsley & Müller (2020), Chapter 29 of this volume. Nonetheless, idioms have been used to motivate assumptions about the architecture of linguistic signs in HPSG as well.

Wasow et al. (1984) and Nunberg et al. (1994) are probably the two most influential papers in formal phraseology in the last decades. While there are many aspects of Nunberg et al. (1994) that have not been integrated into the formal modelling of idioms, there are at least two insights that have been widely adapted in HPSG. First, not all idioms should be represented holistically. Second, the syntactic flexibility of an idiom is related to its semantic decomposability. In fact, Nunberg et al. (1994) state this last insight even more generally:[11]

> We predict that the syntactic flexibility of a particular idiom will ultimately be explained in terms of the compatibility of its semantics with the semantics and pragmatics of various constructions. (Nunberg et al. 1994: 531)

Wasow et al. (1984) and Nunberg et al. (1994) propose a simplified first approach to a theory that would be in line with this quote. They argue that, for English, there is a correlation between syntactic flexibility and semantic decomposability in that non-decomposable idioms are only semi-fixed, whereas decomposable idioms are flexible, to use our terminology from Section 2. This idea has been directly encoded formally in the idiom theory of Gazdar et al. (1985: Chapter 7), who define the framework of *Generalized Phrase Structure Grammar* (GPSG).

Gazdar et al. (1985) assume that non-decomposable idioms are inserted into sentences *en bloc*, i.e., as fully specified syntactic trees which are assigned the idiomatic meaning holistically. This means that the otherwise strictly context-free grammar of GPSG needs to be expanded by adding a (small) set of larger trees. Since non-decomposable idioms are inserted as units, their parts cannot

[11] Aspects of this approach are already present in Higgins (1974) and Newmeyer (1974).

be accessed for syntactic operations such as passivization or movement. Consequently, the generalization about semantic non-decomposability and syntactic fixedness of English idioms from Wasow et al. (1984) is implemented directly.

Decomposable idioms are analyzed as free combinations in syntax. The idiomaticity of such expressions is achieved by two assumptions: First, there is lexical ambiguity, i.e., for an idiom like *pull strings*, the verb *pull* has both a literal meaning and an idiomatic meaning. Similarly for *strings*. Second, Gazdar et al. (1985) assume that lexical items are not necessarily translated into total functions but can be partial functions. Whereas the literal meaning of *pull* might be a total function, the idiomatic meaning of the word would be a partial function that is only defined on elements that are in the denotation of the idiomatic meaning of *strings*. This analysis predicts syntactic flexibility for decomposable idioms, just as proposed in Wasow et al. (1984).

Nunberg et al. (1994: 511–514) show that the connection between semantic decomposability and syntactic flexibility is not as straightforward as suggested. They say that, in German and Dutch, "noncompositional idioms are syntactically versatile" (Nunberg et al. 1994: 514). Similar observations have been brought forward for French in Ruwet (1991). Bargmann & Sailer (2018) and Fellbaum (2019) argue that even for English, passive examples are attested for non-decomposable idioms such as (3).

(3) Live life to the fullest, you never know when the bucket will be kicked.
(Fellbaum 2019: 756)

The current state of our knowledge of the relation between syntactic and semantic idiosyncrasy is that the semantic idiomaticity of an idiom does have an effect on its syntactic flexibility, though the relation is less direct than assumed in the literature based on Wasow et al. (1984) and Nunberg et al. (1994).

4 HPSG analyses of idioms

HPSG does not make a core-periphery distinction; see Müller (2014). Consequently, idioms belong to the empirical domain to be covered by an HPSG grammar. Nonetheless, idioms are not discussed in Pollard & Sag (1994) and their architecture of grammar does not have a direct place for an analysis of idioms.[12] They situate all idiosyncrasy in the lexicon, which consists of lexical entries for

[12] This section follows the basic structure and argument of Sailer (2012) and Richter & Sailer (2014).

basic words. Every word has to satisfy a lexical entry and all principles of grammar; see Davis & Koenig (2020), Chapter 4 of this volume.[13] All properties of a phrase can be inferred from the properties of the lexical items occurring in the phrase and the constraints of grammar.

In their grammar, Pollard & Sag (1994) adhere to the *Strong Locality Hypothesis* (SLH), i.e., all lexical entries describe leaf nodes in a syntactic structure and all phrases are constrained by principles that only refer to local (i.e., *synsem*) properties of the phrase and to local properties of its immediate daughters. This hypothesis is summarized in (4).

(4) Strong Locality Hyphothesis (SLH)
The rules and principles of grammar are statements on a single node of a linguistic structure or on nodes that are immediately dominated by that node.

This precludes any purely phrasal approaches to idioms. Following the heritage of GPSG, we would assume that all regular aspects of linguistic expressions can be handled by mechanisms that follow the SLH, whereas idiomaticity would be a range of phenomena that may violate it. It is, therefore, remarkable that a grammar framework that denies a core-periphery distinction would start with a strong assumption of locality, and, consequently, of regularity.

This is in sharp contrast to the basic motivation of Construction Grammar, which assumes that constructions can be of arbitrary depth and of an arbitrary degree of idiosyncrasy. Fillmore et al. (1988) use idiom data and the various types of idiosyncrasy discussed in Section 2 as an important motivation for this assumption. To contrast this position clearly with the one taken in Pollard & Sag (1994), I will state the *Strong Non-locality Hypothesis* (SNH) in (5).

(5) Strong Non-locality Hypothesis (SNH)
The internal structure of a construction can be arbitrarily deep and show an arbitrary degree of irregularity at any substructure.

The actual formalism used in Pollard & Sag (1994) and King (1989) – see Richter (2020), Chapter 3 of this volume – does not require the strong versions of the

[13] I refer to the lexicon in the technical sense as the collection of lexical entries, i.e., as *descriptions*, rather than as a collection of lexical items, i.e., linguistic signs. Since Pollard & Sag (1994) do not discuss morphological processes, their lexical entries describe full forms. If there is a finite number of such lexical entries, the lexicon can be expressed by a *Word Principle*, a constraint on words that contains a disjunction of all such lexical entries. Once we include morphology, lexical rules, and phrasal lexical entries in the picture, we need to refine this simplified view.

locality and the non-locality hypotheses, but is compatible with weaker versions. I will call these the *Weak Locality Hypothesis* (WLH), and the *Weak Non-locality Hypothesis* (WNH); see (6) and (7) respectively.

(6) Weak Locality Hypothesis (WLH)
At most the highest node in a structure is licensed by a rule of grammar or a lexical entry.

According to the WLH, just as in the SLH, each sign needs to be licensed by the lexicon and/or the grammar. This precludes any *en bloc*-insertion analyses, which would be compatible with the SNH. According to the WNH, in line with the SLH, a sign can, however, impose further constraints on its component parts, that may go beyond local (i.e., *synsem*) properties of its immediate daughters.

(7) Weak Non-locality Hypothesis (WNH)
The rules and principles of grammar can constrain – though not license – the internal structure of a linguistic sign at arbitrary depth.

This means that all substructures of a syntactic node need to be licensed by the grammar, but the node may impose idiosyncratic constraints on which particular well-formed substructures it may contain.

In this section, I will review four types of analyses developed within HPSG in a mildly chronological order: First, I will discuss a conservative extension of Pollard & Sag (1994) for idioms (Krenn & Erbach 1994) that sticks to the SLH. Then, I will look at attempts to incorporate constructional ideas more directly, i.e., ways to include a version of the SNH. The third type of approach will exploit the WLH. Finally, I will summarize recent approaches, which are, again, emphasizing the locality of idioms.

4.1 Early lexical approaches

Krenn & Erbach (1994), based on Erbach (1992), present the first comprehensive HPSG account of idioms. They look at a wide variety of different types of German idioms, including support verb constructions. They only modify the architecture of Pollard & Sag (1994) marginally and stick to the Strong Locality Hypothesis. They base their analysis on the apparent correlation between syntactic flexibility and semantic decomposability from Wasow et al. (1984) and Nunberg et al. (1994). Their analysis is a representational variant of the analysis in Gazdar et al. (1985).

To maintain the SLH, Krenn & Erbach (1994) assume that the information available in syntactic selection is slightly richer than what has been assumed in Pollard & Sag (1994): First, they use a lexeme-identification feature, LEXEME, which is located inside the INDEX value and whose value is the semantic constant associated with a lexeme. Second, they include a feature THETA-ROLE, whose value indicates which thematic role a sign is assigned in a structure. In addition to standard thematic roles, they include a dummy value *nil*. Third, as the paper was written in the transition phase between Pollard & Sag (1987) and Pollard & Sag (1994), they assume that the selectional attributes contain complete *sign* objects rather than just *synsem* objects. Consequently, selection for phonological properties and internal constituent structure is possible, which we could consider a violation of the SLH.

The effect of these changes in the analysis of idioms can be seen in (8) and (9). In (8), I sketch the analysis of the syntactically flexible, decomposable idiom *spill the beans*. There are individual lexical items for the idiomatic words.[14]

(8) a.
$$\begin{bmatrix} \textsc{phon} & \langle \textit{spill} \rangle \\ \textsc{synsem} & \begin{bmatrix} \textsc{cat} & \begin{bmatrix} \textsc{subcat} & \left\langle \text{NP}, \text{NP}\begin{bmatrix} \textsc{lexeme} & \textit{beans\_i} \end{bmatrix} \right\rangle \end{bmatrix} \\ \textsc{cont} & \begin{bmatrix} \textsc{rel} & \textit{spill\_i} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

b.
$$\begin{bmatrix} \textsc{phon} & \langle \textit{beans} \rangle \\ \textsc{synsem} & \begin{bmatrix} \textsc{content} & \begin{bmatrix} \textsc{index} & \begin{bmatrix} \textsc{lexeme} & \textit{beans\_i} \end{bmatrix} \end{bmatrix} \end{bmatrix} \end{bmatrix}$$

The LEXEME values of these words can be used to distinguish them from their ordinary, non-idiomatic homonyms. Each idiomatic word comes with its idiomatic meaning, which models the decomposability of the expression. The lexical items satisfying the entries in (8) can undergo lexical rules such as passivization.

The idiomatic verb *spill* selects an NP complement with the LEXEME value *beans_i*. The lexicon is built in such a way that no other word selects for this LEXEME value. This models the lexical fixedness of the idiom.

The choice of putting the lexical identifier into the INDEX guarantees that it is shared between a lexical head and its phrase, which allows for syntactic flexibility inside the NP. Similarly, the information shared between a trace and its

[14] We do not need to specify the REL value for the noun *beans*, as the LISTEME and the REL value are usually identical.

antecedent contains the INDEX value. Consequently, participation in unbounded dependency constructions is equally accounted for. Finally, since a pronoun has the same INDEX value as its antecedent, pronominalization is also possible.

I sketch the analysis of a non-decomposable, fixed idiom, *kick the bucket*, in (9). In this case, there is only a lexical entry of the syntactic head of the idiom, the verb *kick*. It selects the full phonology of its complement. This blocks any syntactic processes inside this NP. It also follows that the complement cannot be realized as a trace, which blocks extraction.[15] The special THETA-ROLE value *nil* will be used to restrict the lexical rules that can be applied. The passive lexical rule, for example, would be specified in such a way that it cannot apply if the NP complement in its input has this theta-role.

(9)
$$\left[\begin{array}{ll} \textsc{phon} & \langle \textit{kick} \rangle \\ \textsc{synsem} & \left[\begin{array}{ll} \textsc{cat} & \left[\textsc{subcat} \left\langle \text{NP}, \text{NP}\left[\begin{array}{ll} \textsc{phon} & \langle \textit{the, bucket} \rangle \\ \textsc{theta-role} & \textit{nil} \end{array}\right] \right\rangle\right] \\ \textsc{cont} & \left[\textsc{rel} \ \textit{die}\right] \end{array}\right] \end{array}\right]$$

With this analysis, Krenn & Erbach (1994) capture both the idiosyncratic aspects and the regularity of idioms. They show how it generalizes to a wide range of idiom types. I will briefly mention some problems of the approach, though.

There are two problems for the analysis of non-decomposable idioms. First, the approach is too restrictive with respect to the syntactic flexibility of *kick the bucket*, as it excludes cases such as *kick the social/figurative bucket*, which are discussed in Ernst (1981). Second, it is built on equating the class of non-decomposable idioms with that of semi-fixed idioms. As shown in my discussion around example (3), this cannot be maintained.

There are also some undesired properties of the LEXEME value selection. The index identity between a pronoun and its antecedent would require that the subject of the relative clause in (10) has the same INDEX value as the head noun *strings*. However, the account of the lexical fixedness of idioms is built on the assumption that no verb except for the idiomatic *pull* selects for an argument with LEXEME value *strings_i*.[16]

[15] See Borsley & Crysmann (2020), Chapter 13 of this volume for details on the treatment of extraction in HPSG.

[16] Pulman (1993) discusses the analogous problem for the denotational theory of Gazdar et al. (1985).

(10) Parky pulled the strings that got me the job. (McCawley 1981: 137)

Notwithstanding these problems, the analytic ingredients of Krenn & Erbach (1994) constitute the basis of later HPSG analyses. In particular, a mechanism for lexeme-specific selection has been widely assumed in most approaches. The attribute THETA-ROLE can be seen as a simple form of an *inside-out* mechanism, i.e., as a mechanism of encoding information about the larger structure within which a sign appears.

4.2 Phrasal approach

With the advent of constructional analyses within HPSG, starting with Sag (1997), it is natural to expect phrasal accounts of idioms to emerge as well, as idiomaticity is a central empirical domain for Construction Grammar; see Müller (2020), Chapter 33 of this volume. In this version of HPSG, there is an elaborate type hierarchy below *phrase*. Sag (1997) also introduces *defaults* into HPSG, which play an important role in the treatment of idioms in constructional HPSG. The clearest phrasal approach to idioms can be found in Riehemann (2001), which incorporates insights from earlier publications such as Riehemann (1997) and Riehemann & Bender (1999). The overall framework of Riehemann (2001) is constructional HPSG with *Minimal Recursion Semantics* (Copestake et al. 1995; 2005); see also Koenig & Richter (2020), Chapter 22 of this volume.

For Riehemann, idioms are phrasal units. Consequently, she assumes a subtype of *phrase* for each idiom, such as *spill-beans-idiomatic-phrase* or *kick-bucket-idiomatic-phrase*. The proposal in Riehemann (2001) simultaneously is phrasal and obeys the SLH. To achieve this, Riehemann (2001) assumes an attribute WORDS, whose value contains all words dominated by a phrase. This makes it possible to say that a phrase of type *spill-beans-idiomatic-phrase* dominates the words *spill* and *beans*. This is shown in the relevant type constraint for the idiom *spill the beans* in (11).[17]

[17] The percolation mechanism for the feature WORDS is rather complex. In fact, in Riehemann (2001: Section 5.2.1) the idiom-specific words appear within a C-WORDS value and the other words are dominated by the idiomatic phrase in the value of an attribute OTHER-WORDS, both of which together form the value of WORDS. While all the values of these features are subject to local percolation principles, the fact that entire words are percolated undermines the locality intuition behind the SLH.

(11) Constraint on the type *spill-beans-idiomatic-phrase* from Riehemann (2001: 185):

$$\left[\begin{array}{l} \textit{spill-beans-ip} \\ \textsc{words} \left\{ \begin{array}{l} \left[\begin{array}{l} \textit{i\_spill} \\ \ldots\ \textsc{liszt} \left\langle \left[\begin{array}{l} \textit{i\_spill\_rel} \\ \textsc{undergoer}\ \boxed{1} \end{array}\right] \right\rangle \end{array}\right] \overset{\leq}{\sqcap} \left[\ldots\ \textsc{liszt} \left\langle \textit{\_spill\_rel} \right\rangle\right], \\ \left[\begin{array}{l} \textit{i\_beans} \\ \ldots\ \textsc{liszt} \left\langle \left[\begin{array}{l} \textit{i\_beans\_rel} \\ \textsc{inst}\ \boxed{1} \end{array}\right] \right\rangle \end{array}\right] \overset{\leq}{\sqcap} \left[\ldots\ \textsc{liszt} \left\langle \textit{\_beans\_rel} \right\rangle\right], \ldots \end{array} \right\} \end{array}\right]$$

The WORDS value of the idiomatic phrase contains at least two elements, the idiomatic words of type *i_spill* and *i_beans*. The special symbol $\overset{\leq}{\sqcap}$ used in this constraint expresses a default. It says that the idiomatic version of the word *spill* is just like its non-idiomatic homonym, except for the parts specified in the left-hand side of the default. In this case, the type of the words and the type of the semantic predicate contributed by the words are changed. Riehemann (2001) only has to introduce the types for the idiomatic words in the type hierarchy but need not specify type constraints on the individual idiomatic words, as these are constrained by the default statement within the constraints on the idioms containing them.

As in the account of Krenn & Erbach (1994), the syntactic flexibilty of the idiom follows from its free syntactic combination and the fact that all parts of the idiom are assigned an independent semantic contribution. The lexical fixedness is a consequence of the requirement that particular words are dominated by the phrase, namely the idiomatic versions of *spill* and *beans*.

The appeal of the account is particularly clear in its application to non-decomposable, semi-fixed idioms such as *kick the bucket* (Riehemann 2001: 212). For such expressions, the idiomatic words that constitute them are assumed to have an empty semantics and the meaning of the idiom is contributed as a constructional semantic contribution only by the idiomatic phrase. Since the WORDS list contains entire words, it is also possible to require that the idiomatic word *kick* be in active voice and/or that it take a complement compatible with the description of the idiomatic word *bucket*. This analysis captures the syntactically regular internal structure of this type of idioms and is compatible with the occurrence

of modifiers such as *proverbial*. At the same time, it prevents passivization and excludes extraction of the complement.

Riehemann's approach clearly captures the intuition of idioms as phrasal units much better than any other approach in HPSG. However, it faces a number of problems. First, the integration of the approach with constructional HPSG is done in such a way that the phrasal types for idioms are cross-classified in complex type hierarchies with the various syntactic constructions in which the idiom can appear. This allows Riehemann to account for idiosyncratic differences in the syntactic flexibility of idioms, but the question is whether such an explicit encoding misses generalizations that should follow from independent properties of the components of an idiom and/or of the syntactic construction – in line with the quote from Nunberg et al. (1994) on page 744.

Second, the mechanism of percolating dominated words to each phrase is not compatible with the intuitions of most HPSG researchers. Since no empirical motivation for such a mechanism aside from idioms is provided in Riehemann (2001), this idea has not been pursued in other papers.

Third, the question of how to block the free occurrence of idiomatic words, i.e., the occurrence of an idiomatic word without the rest of the idiom, is not solved in Riehemann (2001). While the idiom requires the presence of particular idiomatic words, the occurrence of these words is not restricted.[18] Note that idiomatic words may sometimes be found without the other elements of the idiom – evidenced by expressions such as in *bucket list* 'list of things to do before one dies'. Such data may be considered as support of Riehemann's approach; however, the extent to which we find such free occurrences of idiomatic words is extremely small.[19]

Before closing this subsection, I would like to point out that Riehemann (2001) and Riehemann & Bender (1999) are the only HPSG papers on idioms that address the question of statistical idiomaticity, based on the variationist study in Bender (2001). In particular, Riehemann (2001: 297–301) proposes phrasal constructions for collocations even if these do not show any lexical, syntactic, semantic, or prag-

[18] Since the problem of free occurrences of idiomatic words is not an issue for parsing, versions of Riehemann's approach have been integrated into practical parsing systems (Villavicencio & Copestake 2002); see Bender & Emerson (2020), Chapter 25 of this volume. Similarly, the approach to idioms sketched in Flickinger (2015) is part of a system for parsing and machine translation. Idioms in the source language are identified by bits of semantic representation – analogous to the elements in the WORDS set. This approach, however, does not constitute a theoretical modelling of idioms; it does not exclude ill-formed uses of idioms but identifies potential occurrences of an idiom in the output of a parser.

[19] See the discussion around (1) for a parallel situation with bound words.

matic idiosyncrasy but just a statistical co-occurrence preference. She extends this into a larger plea for an *experience-based HPSG*. Bender (2001) discusses the same idea under the notions of *minimal* versus *maximal* grammars, i.e., grammars that are as free of redundancy as possible to capture the grammatical sentences of a language with their correct meaning versus grammars that might be open to a connection with usage-based approaches to language modelling. Bender (2001: 292) sketches a version of HPSG with frequencies/probabilities attached to lexical and phrasal types.[20]

4.3 Mixed lexical and phrasal approaches

While Riehemann (2001) proposes a parallel treatment of decomposable and non-decomposable idioms – and of flexible and semi-fixed idioms – the division between fixed and non-fixed expressions is at the core of another approach, the *two-dimensional theory of idioms*. This approach was first outlined in Sailer (2000) and referred to under this label in Richter & Sailer (2009; 2014). It is intended to combine constructional and collocational approaches to grammar.

The basic intuition behind this approach is that signs have internal and external properties. All properties that are part of the feature structure of a sign are called *internal*. Properties that relate to larger feature structures containing this sign are called its *external* properties. The approach assumes that there is a notion of *regularity* and that anything diverging from it is *idiosyncratic* – or idiomatic, in the terminology of this chapter.

This approach is another attempt to reify the GPSG analysis within HPSG. Sailer (2000) follows the distinction of Nunberg et al. (1994) into non-decomposable and non-flexible idioms on the one hand and decomposable and flexible idioms on the other. The first group is considered internally irregular and receives a constructional analysis in terms of a *phrasal lexical entry*. The second group is considered to consist of independent, smaller lexical units that show an external irregularity in being constrained to co-occur within a larger structure. Idioms of the second group receive a collocational analysis. The two types of irregularity are connected by the *Predictability Hypothesis*, given in (12).

(12) Predictability Hypothesis (Sailer 2000: 366):

[20] An as-yet unexplored solution to the problem of free occurrence of idiomatic words within an experience-based version of HPSG could be to assign the type *idiomatic-word* an extremely low probability of occurring. This might have the effect that such a word can only be used if it is explicitly required in a construction. However, note that neither defaults nor probabilities are well-defined part of the formal foundations of theoretical work on HPSG; see Richter (2020), Chapter 3 of this volume.

> For every sign whose internal properties are fully predictable, the distributional behavior of this sign is fully predictable as well.

In the most recent version of this approach, Richter & Sailer (2009; 2014), there is a feature COLL defined on all signs. The value of this feature specifies the type of internal irregularity. The authors assume a cross-classification of regularity and irregularity with respect to syntax, semantics, and phonology – ignoring pragmatic and statistical (ir)regularity in their paper. Every basic lexical entry is defined as completely irregular, as its properties are not predictable. Fully regular phrases such as *read a book* have a trivial value of COLL. A syntactically internally regular but fixed idiom such as *kick the bucket* is classified as having only semantic irregularity, whereas a syntactically irregular expression such as *trip the light fantastic* is of an irregularity type that is a subsort of syntactic and semantic irregularity, but not of phonological irregularity. Following the terminology of Fillmore et al. (1988), this type is called *extra-grammatical-idiom*. The phrasal lexical entry for *trip the light fantastic* is sketched in (13), adjusted to the feature geometry of Sag (1997).

Sag97 does not have VAL

(13) Phrasal lexical entry for the idioms *trip the light fantastic*:

$$\begin{bmatrix}
\textit{headed-phrase} \\
\textsc{phon} & \boxed{1} \oplus \langle \textit{the, light, fantastic} \rangle \\
\textsc{synsem} & \begin{bmatrix}
\textsc{head} & \boxed{4} \\
\textsc{listeme} & \textit{trip-the-light-fantastic} \\
\textsc{val} & \boxed{5}\left[\textsc{subj}\left\langle \boxed{2}\left[\ldots\textsc{index}\ \boxed{3}\right]\right\rangle\right] \\
\textsc{cont} & \begin{bmatrix} \textit{trip-light-fant} \\ \textsc{dancer}\ \boxed{3} \end{bmatrix}
\end{bmatrix} \\
\textsc{head-dtr} & \begin{bmatrix}
\textsc{phon} & \boxed{1} \\
\textsc{synsem} & \begin{bmatrix}
\textsc{head} & \boxed{4}\ \textit{verb} \\
\textsc{listeme} & \textit{trip} \\
\textsc{val} & \boxed{5}\begin{bmatrix} \textsc{subj} & \langle \boxed{2} \rangle \\ \textsc{comps} & \langle\rangle \end{bmatrix}
\end{bmatrix}
\end{bmatrix} \\
\textsc{coll} & \textit{extra-grammatical-idiom}
\end{bmatrix}$$

In (13), the constituent structure of the phrase is not specified, but the phonology is fixed, with the exception of the head daughter's phonological contribution. This accounts for the syntactic irregularity of the idiom. The semantics of the idiom is not related to the semantic contributions of its components, which accounts for the semantic idiomaticity.

Soehn (2006) applies this theory to German. He solves the problem of the relatively large degree of flexibility of non-decomposable idioms in German by using underspecified descriptions of the constituent structure dominated by the idiomatic phrase.

For decomposable idioms, the two-dimensional theory assumes a collocational component. This component is integrated into the value of an attribute REQ, which is only defined on *coll* objects of one of the irregularity types. This encodes the Predictability Hypothesis. The most comprehensive version of this collocational theory is given in Soehn (2009), summarizing and extending ideas from Soehn (2006) and Richter & Soehn (2006). Soehn assumes that collocational requirements can be of various types: a lexical item can be constrained to co-occur with particular *licensers* (or collocates). These can be other lexemes, semantic operators, or phonological units. In addition, the domain within which this licensing has to be satisfied is specified in terms of syntactic barriers, i.e., syntactic nodes dominating the externally irregular item.

To give an example, the idiom *spill the beans* would be analyzed as consisting of two idiomatic words *spill* and *beans* with special LISTEME values *spill-i* and *beans-i*. The idiomatic verb *spill* imposes a lexeme selection on its complement. The idiomatic noun *beans* has a non-empty REQ value, which specifies that it must be selected by a word with LISTEME value *spill-i* within the smallest complete clause dominating it.

The two-dimensional approach suffers from a number of weaknesses. First, it presupposes a notion of regularity. This assumption is not shared by all linguists. Second, the criteria for whether an expression should be treated constructionally or collocationally are not always clear. Idioms with irregular syntactic structure need to be analyzed constructionally, but this is less clear for non-decomposable idioms with regular syntactic structure such as *kick the bucket*.

4.4 Recent lexical approaches

Kay et al. (2015) marks an important re-orientation in the analysis of idioms: the lexical analysis is extended to all syntactically regular idioms, i.e., to both decom-

posable (*spill the beans*) and non-decomposable idioms (*kick the bucket*).[21] Kay et al. (2015) achieve a lexical analysis of non-decomposable idioms by two means: (i), an extension of the HPSG selection mechanism, and (ii), the assumption of semantically empty idiomatic words.

As in previous accounts, the relation among idiom parts is established through lexeme-specific selection, using a feature LID (for *lexical identifier*). The authors assume that there is a difference between idiomatic and non-idiomatic LID values. Only heads that are part of idioms themselves can select for idiomatic words.

For the idiom *kick the bucket*, Kay et al. (2015) assume that all meaning is carried by the lexical head, an idiomatic version of *kick*, whereas the other two words, *the* and *bucket* are meaningless. This meaninglessness allows Kay et al. to block the idiom from occurring in constructions which require meaningful constituents, such as questions, *it*-clefts, middle voice, and others. To exclude passivization, the authors assume that the English passive cannot apply to verbs selecting a semantically empty direct object.

The approach in Kay et al. (2015) is a recent attempt to maintain the SLH as much as possible. Since the SLH has been a major conceptual motivation for Sign-based Construction Grammar, Kay et al.'s paper is an important contribution showing the empirical robustness of this assumption.

Bargmann & Sailer (2018) propose a similar lexical approach to non-decomposable idioms. They take as their starting point the syntactic flexibility of semantically non-decomposable idioms in English and, in particular, in German. There are two main differences between Kay et al.'s paper and Bargmann & Sailer's: (i), Bargmann & Sailer assume a collocational rather than a purely selectional mechanism to capture lexeme restrictions of idioms, and (ii), they propose a redundant semantics rather than an empty semantics for idiom parts in non-decomposable idioms. In other words, Bargmann & Sailer (2018) propose that both *kick* and *bucket* contribute the semantics of the idiom *kick the bucket*. Bargmann & Sailer argue that the semantic contributions of parts of non-decomposable, syntactically regular idioms are the same across languages, whereas the differences in syntactic flexibility are related to the different syntactic, semantic, and pragmatic constraints imposed on various constructions. To give just one example, while there are barely any restrictions on passive subjects in German, there are strong discourse-structural constraints on passive subjects in English.

Both Kay et al. (2015) and Bargmann & Sailer (2018) attempt to derive the (partial) syntactic inflexibility of non-decomposable idioms from independent properties of the relevant constructions. As such, they subscribe to the programmatic

[21] This idea has been previously expressed within a Minimalist perspective in Everaert (2010).

statement of Nunberg et al. (1994) quoted on page 744. In this respect, the extension of the lexical approach from decomposable idioms to all syntactically regular expressions has been a clear step forward.

Findlay (2017) provides a recent discussion and criticism of lexical approaches to idioms in general, which applies in particular to non-decomposable expressions. His reservations comprise the following points. First, there is a massive proliferation of lexical entries for otherwise homophonous words. Second, the lexical analysis does not represent idioms as units, which might make it difficult to connect their theoretical treatment with processing evidence. Findlay refers to psycholinguistic studies, such as Swinney & Cutler (1979), that point to a faster processing of idioms than of free combinations. While the relevance of processing arguments for an HPSG analysis are not clear, I share the basic intuition that idioms, decomposable or not, are a unit and that this should be part of their linguistic representation.

5 Where to go from here?

The final section of this article contains short overviews of research that has been done in areas of phraseology that are outside the main thread of this chapter. I will also identify desiderata.

5.1 Neglected phenomena

Not all types of idioms or idiomaticity mentioned in Section 2 have received an adequate treatment in the (HPSG) literature. I will briefly look at three empirical areas that deserve more attention: neglected types of idiom variation, phraseological patterns, and the literal and non-literal meaning components of idioms.

Most studies on idiom variation have looked at verb- and sentence-related syntactic constructions, such as passive and topicalization. However, not much attention has been paid to lexical variation in idioms. This variation is illustrated by the following examples from Richards (2001: 184, 191).

(14) a. The Count gives everyone the creeps.
b. You get the creeps (just looking at him).
c. I have the creeps.

In (14), the alternation of the verb seems to be very systematic – and has been used by Richards (2001) to motivate a lexical decomposition of these verbs. A similar argument has been made in Mateu & Espinal (2007) for similar idioms

in Catalan. We are lacking systematic, larger empirical studies of this type of substitution, and it would be important to see how it can be modelled in HPSG. One option would be to capture the *give–get–have* alternation(s) with lexical rules. Such lexical rules would be different from the standard cases, however, as they would change the lexeme itself rather than just alternating its morpho-syntactic properties or its semantic contribution.

In the case mentioned in footnote 6, the alternation consists of substituting a word with a (near) synonym and keeping the meaning of the idiom intact. Again, HPSG seems to have all the required tools to model this phemonenon – for example, by means of hierarchies of LEXICAL-ID values. However, the extent of this phenomenon across the set of idioms is not known empirically.

Concerning syntactic variation, the nominal domain has not yet received the attention it might deserve. There is a well-known variation with respect to the marking of possession within idioms. This has been documented for English in Ho (2015), for Modern Hebrew in Almog (2012), and for Modern Greek and German in Markantonatou & Sailer (2016). In German, we find a relatively free alternation between a plain definite and a possessive; see (15a). This is, however, not possible with all idioms; see (15b).

(15) a. Alex hat den / seinen Verstand verloren.
Alex has the his mind lost
'Alex lost his mind.'

b. Alex hat *den / ihren Frieden mit der Situation gemacht.
Alex has the her peace with the situation made
'Alex made her peace with the situation.'

We can also find a free dative in some cases, expressing the possessor. In (16a), a dative possessor may co-occur with a plain definite or a coreferential possessive determiner; in (16b), only the definite article but not the possessive determiner is possible.

(16) a. Alex hat mir das / mein Herz gebrochen.
Alex has me.DAT the my heart broken
'Alex broke my heart.'

b. Alex sollte mir lieber aus den / *meinen Augen gehen.
Alex should me.DAT rather out of the my eyes go
'Alex should rather disappear from my sight.'

While they do not offer a formal encoding, Markantonatou & Sailer (2016) observe that a particular encoding of possession in idioms is only possible if it

would also be possible in a free combination. However, an idiom may be idiosyncratically restricted to a subset of the realizations that would be possible in a corresponding free combination. A formalization in HPSG might consist of a treatment of possessively used definite determiners, combined with an analysis of free datives as an extension of a verb's argument structure.

Related to the question of lexical variation are *phraseological patterns*, i.e., very schematic idioms in which the lexical material is largely free. Some examples of phraseological patterns are the *Incredulity Response Construction* as in *What, me worry?* (Akmajian 1984; Lambrecht 1990), or the *What's X doing Y?* construction (Kay & Fillmore 1999). Such patterns are of theoretical importance as they typically involve a non-canonical syntactic pattern. The different locality and non-locality hypotheses introduced above make different predictions. Fillmore et al. (1988) have presented such constructions as a motivation for the non-locality of constructions, i.e., as support of a SNH. However, Kay & Fillmore (1999) show that a lexical analysis might be possible for some cases at least, which they illustrate with the *What's X doing Y?* construction.

Borsley (2004) looks at another phraseological pattern, the *the X-er the Y-er* construction, or *comparative correlative construction*. Borsley analyzes this construction by means of two special (local) phrase structure types: one for the comparative *the*-clauses, and one for the overall construction. He shows that (i), the idiosyncrasy of the construction concerns two levels of embedding and is, therefore, non-local; however, (ii), a local analysis is still possible. This approach raises the question of whether the WNH is empirically vacuous since we can always encode a non-local construction in terms of a series of idiosyncratic local constructions. Clearly, work on more phraseological patterns is needed to assess the various analytical options and their consequences for the architecture of grammar.

A major charge for the conceptual and semantic analysis of idioms is the interaction between the literal and the idiomatic meaning. I presented the basic empirical facts in Section 2. All HPSG approaches to idioms so far basically ignore the literal meaning. This position might be justified, as an HPSG grammar should just model the structure and meaning of an utterance and need not worry about the meta-linguistic relations among different lexical items or among different readings of the same (or a homophonous) expression. Nonetheless, this issue touches on an important conceptual point. Addressing it might immediately provide possibilities to connect HPSG research to other disciplines and/or frameworks like cognitive linguistics, such as in Dobrovol'skij & Piirainen (2005), and psycholinguistics.

5.2 Challenges from other languages

The majority of work on idioms in HPSG has been done on English and German. As discussed in Section 4.4, the recent trend in HPSG idiom research necessitates a detailed study of individual syntactic structures. Consequently, the restriction to two closely related languages limits the possible phenomena that can be studied concerning idioms. It would be essential to expand the empirical coverage of idiom analyses in HPSG to as many different languages as possible. The larger degree of syntactic flexibility of French, German, and Dutch idioms (Ruwet 1991; Nunberg et al. 1994; Schenk 1995) has led to important refinements of the analysis in Nunberg et al. (1994) and, ultimately, to the lexical analyses of all syntactically regular idioms.

Similarly, the above-mentioned data on possessive alternations only become prominent when languages beyond English are taken into account. Modern Greek, German, and many others show the type of external possessor classified as a European areal phenomenon in Haspelmath (1999). It would be important to look at idioms in languages with other types of external possessors.

In a recent paper, Sheinfux et al. (2019) provide data from Modern Hebrew that show that opacity and figurativity of an idiom are decisive for its syntactic flexibility, rather than decomposability. This result stresses the importance of the literal reading for an adequate account of the syntactic behavior of idioms. It shows that the inclusion of other languages can cause a shift of focus to other types of idioms or other types of idiomaticity.

To add just one more example, HPSG(-related) work on Persian such as Müller (2010) and Samvelian & Faghiri (2016) establishes a clear connection between complex predicates and idioms. Their insights might also lead to a reconsideration of the similarities between light verbs and idioms, as already set out in Krenn & Erbach (1994).

As far as I can see, the following empirical phenomena have not been addressed in HPSG approaches to idioms, as they do not occur in the main object languages for which we have idiom analyses, i.e., English and German. They are, however, common in other languages: the occurrence of clitics in idioms (found in Romance and Greek); aspectual alternations in verbs (Slavic and Greek); argument alternations other than passive and dative alternation (such as anti-passive, causative, inchoative, etc. (in part found in Hebrew and addressed in Sheinfux et al. 2019); and displacement of idiom parts into special syntactic positions (focus position in Hungarian).

Finally, so far, idioms have usually been considered as either offering irregular structures or as being more restricted in their structures than free combinations.

In some languages, however, we find archaic syntactic structures and function words in idioms that do not easily fit these two analytic options. To name just a few, Lødrup (2009) argues that Norwegian used to have an external possessor construction similar to that of other European languages, which is only conserved in some idioms. Similarly, Dutch has a number of archaic case inflections in multiword expressions (Kuiper 2018: 129), and there are archaic forms in Modern Greek multiword expressions. It is far from clear what the best way would be to integrate such cases into an HPSG grammar.

6 Conclusion

Idioms are among the topics in linguistics for which HPSG-related publications have had a clear impact on the field and have been widely quoted across frameworks. This handbook article aimed at providing an overview over the development of idiom analyses in HPSG. There seems to be a development towards ever more lexical analyses, starting from the holistic approach for all idioms in Chomsky's work, to a lexical account for all syntactically regular expressions. Notwithstanding the advantages of the lexical analyses, I consider it a basic problem of such approaches that the unit status of idioms is lost. Consequently, I think that the right balance between phrasal and lexical aspects in the analysis of idioms has not yet been fully achieved.

The sign-based character of HPSG seems to be particularly suited for a theory of idioms as it allows one to take into consideration syntactic, semantic, and pragmatic aspects and to use them to constrain the occurrence of idioms appropriately.

Appendix: List of used idioms

Some idioms do not show semantic idiomaticity at all, such as collocations (*brush one's teeth*) or support verb constructions (*take a shower*). Many body-part expressions such as *shake hands* or *shake one's head* constitute a more complex case. They describe a conventionalized activity and denote the social meaning of this activity (Burger 1976).

English

idiom	paraphrase	comment
break the ice	relieve tension in a strained situation	non-decomposable
brush one's teeth	clean one's teeth with a tooth brush	collocation, no idiomaticity
give s.o. the creeps	make s.o. feel uncomfortable	systematic lexical variation
Good morning!	(morning greeting)	formulaic expression
immaculate performance	perfect performance	statistical idiomaticity
in a trice	in a moment	bound word: *trice*
kick the bucket	die	non-decomposable
make headway	make progress	bound word: *headway*
pull strings	exert influence/use one's connections	flexible
saw logs	snore	transparent, non-decomposable, semi-flexible
shake hands	greet	body-part expression
shake one's head	decline/negate	body-part expression, possessive idiom
shit hit the fan	there is trouble	subject as idiom component, transparent/figurative, non-decomposable
shoot the breeze	chat	non-decomposable
spill the beans	reveal a secret	flexible
take a shower	clean oneself using a shower	collocation, light verb construction
take the bull by the horns	approach a problem directly	figurative expression
take umbrage	take offence	bound word: *umbrage*
trip the light fantastic	dance	syntactically irregular

German

idiom	gloss	translation	comment
den/seinen Verstand verlieren	the/one's mind lose	lose one's mind	alternation of possessor marking
jdm. das Herz brechen	s.o. the heart break	break s.o.'s heart	dative possessor and possessor alternation
jdm. aus den Augen gehen	s.o. out of the eyes go	disappear from s.o.'s sight	dative possessor, restricted possessor alternation
seinen Frieden machen mit	one's peace make with	make one's peace with	no possessor alternation possible
wie warme Semmeln/ Brötchen/Schrippen weggehen	like warm rolls vanish	sell like hotcakes	parts can be exchanged by synonyms

Abbreviations

GPSG	Generalized Phrase Structure Grammar (Gazdar et al. 1985)
MGG	Mainstream Generative Grammar
SLH	Strong Locality Hypothesis, see page 746
SNH	Strong Non-locality Hypothesis, see page 746
WLH	Weak Locality Hypothesis, see page 747
WNH	Weak Non-locality Hypothesis, see page 747

Acknowledgements

I have perceived Ivan A. Sag and his work with various colleagues as a major inspiration for a lot of my own work on idioms and multiword expressions. This is clearly reflected in the structure of this paper, too. I apologize for this bias, but I think it is legitimate within an HPSG handbook. I am grateful to Stefan Müller and Tom Wasow for comments on the outline and the first version of this chapter. I would not have been able to format this chapter without the support of Sebastian Nordhoff. I would like to thank Elizabeth Pankratz for comments and proofreading.

References

Akmajian, Adrian. 1984. Sentence types and the form-function fit. *NLLT* 2(1). 1–23.

Almog, Lior. 2012. *The formation of idioms: Evidence from possessive datives in Hebrew*. MA thesis, Tel Aviv University. http://humanities1.tau.ac.il/linguistics_eng/images/stories/Lior_Ordentlich_MA_2012.pdf (13 October, 2018).

Aronoff, Mark. 1976. *Word formation in Generative Grammar*. Cambridge, MA: MIT Press.

Baldwin, Timothy & Su Nam Kim. 2010. Multiword expressions. In Nitin Indurkhya & Fred J. Damerau (eds.), *Handbook of natural language processing*, 2nd edn., 267–292. Boca Raton: CRC Press.

Bargmann, Sascha & Manfred Sailer. 2018. The syntactic flexibility of semantically non-decomposable idioms. In Manfred Sailer & Stella Markantonatou (eds.), *Multiword expressions: Insights from a multi-lingual perspective* (Phraseology and Multiword Expressions 1), 1–29. Berlin: Language Science Press. DOI: 10.5281/zenodo.1182587.

Bender, Emily M. 2001. *Syntactic variation and linguistic competence: The case of AAVE copula absence*. Stanford University. (Doctoral dissertation). http://faculty.washington.edu/ebender/dissertation/ (18 August, 2020).

Bender, Emily M. & Guy Emerson. 2020. Computational linguistics and grammar engineering. In Stefan Müller, Anne Abeillé, Robert D. Borsley & Jean-Pierre Koenig (eds.), *Head-Driven Phrase Structure Grammar: The handbook*, 1009–1052. Berlin: Language Science Press. DOI: ??.

Borsley, Robert D. 2004. An approach to English comparative correlatives. In Stefan Müller (ed.), *Proceedings of the 11th International Conference on Head-Driven Phrase Structure Grammar, Center for Computational Linguistics, Katholieke Universiteit Leuven*, 70–92. Stanford, CA: CSLI Publications. http://csli-publications.stanford.edu/HPSG/2004/ (18 August, 2020).

Borsley, Robert D. & Berthold Crysmann. 2020. Unbounded dependencies. In Stefan Müller, Anne Abeillé, Robert D. Borsley & Jean-Pierre Koenig (eds.), *Head-Driven Phrase Structure Grammar: The handbook*, 507–561. Berlin: Language Science Press. DOI: ??.

Borsley, Robert D. & Stefan Müller. 2020. HPSG and Minimalism. In Stefan Müller, Anne Abeillé, Robert D. Borsley & Jean-Pierre Koenig (eds.), *Head-Driven Phrase Structure Grammar: The handbook*, 1151–1223. Berlin: Language Science Press. DOI: ??.

Burger, Harald. 1976. *Die Achseln zucken*: Zur sprachlichen Kodierung nichtsprachlicher Kommunikation. *Wirkendes Wort* 26. 311–339.

Burger, Harald. 1998. *Phraseologie: Eine Einführung am Beispiel des Deutschen*. Bielefeld: E. Schmidt.

Chomsky, Noam. 1965. *Aspects of the theory of syntax*. Cambridge, MA: MIT Press.

Chomsky, Noam. 1981. *Lectures on government and binding* (Studies in Generative Grammar 9). Dordrecht: Foris Publications. DOI: 10.1515/9783110884166.

Copestake, Ann, Dan Flickinger, Robert Malouf, Susanne Riehemann & Ivan A.Sag. 1995. Translation using Minimal Recursion Semantics. In *Proceedings of the Sixth International Conference on Theoretical and Methodological Issues in Machine Translation (TMI95), Leuven, Belgium*.

Copestake, Ann, Dan Flickinger, Carl J. Pollard & Ivan A. Sag. 2005. Minimal Recursion Semantics: An introduction. *Research on Language and Computation* 3(2–3). 281–332. DOI: 10.1007/s11168-006-6327-9.

Corver, Norbert, Jeroen van Craenenbroeck, William Harwood, Marko Hladnik, Sterre Leufkens & Tanja Temmermann. 2019. Introduction: The compositionality and syntactic flexibilty of verbal idioms. *Journal of Linguistics* 57(4). 725–733. DOI: https://doi.org/10.1515/ling-2019-0021.

Culicover, Peter W. & Ray S. Jackendoff. 2005. *Simpler Syntax*. Oxford: Oxford University Press. DOI: 10.1093/acprof:oso/9780199271092.001.0001.

Davis, Anthony & Jean-Pierre Koenig. 2020. The nature and role of the lexicon in HPSG. In Stefan Müller, Anne Abeillé, Robert D. Borsley & Jean-Pierre Koenig (eds.), *Head-Driven Phrase Structure Grammar: The handbook*, 117–165. Berlin: Language Science Press. DOI: ??.

Dobrovol'skij, Dmitrij & Elisabeth Piirainen. 2005. *Figurative language: Cross-cultural and cross-linguistic perspectives* (Current Research in the Semantics/Pragmatics Interface 13). Amsterdam: Elsevier.

Erbach, Gregor. 1992. Head-driven lexical representation of idioms in HPSG. In Martin Everaert, Erik-Jan van der Linden, André Schenk & Rob Schreuder (eds.), *Proceedings of the International Conference on Idioms*. Tilburg, The Netherlands: ITK.

Ernst, Thomas. 1981. Grist for the linguistic mill: Idioms and 'extra' adjectives. *Journal of Linguistic Research* 1(3). 51–68.

Everaert, Martin. 2010. The lexical encoding of idioms. In Malka Rappaport Hovav, Edit Doron & Ivy Sichel (eds.), *Lexical semantics, syntax, and event structure* (Oxford Studies in Theoretical Linguistics 27), 76–97. Oxford: Oxford University Press.

Fellbaum, Christiane. 2019. How flexible are idioms? A corpus-based study. *Journal of Linguistics* 57(4). 735–767. DOI: https://doi.org/10.1515/ling-2019-0015.

Fillmore, Charles J., Paul Kay & Mary Catherine O'Connor. 1988. Regularity and idiomaticity in grammatical constructions: The case of *let alone*. *Language* 64(3). 501–538.

Findlay, Jamie Y. 2017. Multiword expressions and lexicalism. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG'17 Conference*, 209–229. Stanford, CA: CSLI Publications. http://web.stanford.edu/group/cslipublications/cslipublications/LFG/LFG-2017/lfg2017-findlay.pdf (22 June, 2018).

Fleischer, Wolfgang. 1997. *Phraseologie der deutschen Gegenwartssprache*. 2nd edn. Tübingen: Max Niemeyer Verlag.

Flickinger, Dan. 2015. *Multi-word expressions in HPSG*. Slides to a course at the 1st Parseme Training School, Prague, January 2015. https://ufal.mff.cuni.cz/events/parseme-1st-training-school (14 August, 2018).

Fraser, Bruce. 1970. Idioms within a transformational grammar. *Foundations of Language* 6(1). 22–42.

Gazdar, Gerald, Ewan Klein, Geoffrey K. Pullum & Ivan A. Sag. 1985. *Generalized Phrase Structure Grammar*. Cambridge, MA: Harvard University Press.

Gibbs, Raymond W., Nandini P. Nayak, John L. Bolton & Melissa E. Keppel. 1989. Speakers' assumptions about the lexical flexibility of idioms. *Memory and Cognition* 17(1). 58–68.

Gibbs, Raymond W. Jr. & Herbert L. Colston. 2007. Psycholinguistic aspects of phraseology: The American tradition. In Harald Burger, Dmitrij Dobrovol'skij, Peter Kühn & Neal R. Norrick (eds.), *Phraseologie/phraseology*, vol. 2 (Ein internationales Handbuch der zeitgenössischen Forschung/An International Handbook of Contemporary Research), 819–836. Berlin: De Gruyter.

Glucksberg, Sam. 2001. *Understanding figurative language: From metaphors to idioms* (Oxford Psychology Series 36). Oxford: Oxford University Press.

Gross, Maurice. 1982. Une classification des phrases "figées" du français. *Revue Québecoise de Linguistique* 11(2). 151–185. DOI: 10.7202/602492ar.

Haspelmath, Martin. 1999. External possession in a European areal perspective. In Doris L. Payne & Immanuel Barshi (eds.), *External possession*, 109–135. Amsterdam: Benjamins.

Higgins, Francis Roger. 1974. On the use of idioms as evidence for movement. A cautionary note. Unpublished manuscript of a talk given at LSA 1974, New York. Amherst, MA.

Ho, Jia Qian. 2015. *Losing one's mind over meaning: analysing the behaviour of English possessive idioms*. Tech. rep. Linguistics & Multilingual Studies, Nanyang Technological University, Singapore. http://compling.hss.ntu.edu.sg/pdf/2015-fyp-ho-jia-qian.pdf (3 September, 2019).

Horn, George M. 2003. Idioms, metaphors and syntactic mobility. *Journal of Linguistics* 39(2). 245–273.

Kay, Paul & Charles J. Fillmore. 1999. Grammatical constructions and linguistic generalizations: The What's X Doing Y? Construction. *Language* 75(1). 1–33.

Kay, Paul, Ivan A. Sag & Daniel P. Flickinger. 2015. *A lexical theory of phrasal idioms*. Ms. CSLI Stanford.

King, Paul. 1989. *A logical formalism for Head-Driven Phrase Structure Grammar*. University of Manchester. (Doctoral dissertation).

Koenig, Jean-Pierre & Frank Richter. 2020. Semantics. In Stefan Müller, Anne Abeillé, Robert D. Borsley & Jean-Pierre Koenig (eds.), *Head-Driven Phrase Structure Grammar: The handbook*, 909–948. Berlin: Language Science Press. DOI: ??.

Krenn, Brigitte & Gregor Erbach. 1994. Idioms and support verb constructions. In John Nerbonne, Klaus Netter & Carl Pollard (eds.), *German in Head-Driven Phrase Structure Grammar* (CSLI Lecture Notes 46), 365–396. Stanford, CA: CSLI Publications.

Kuiper, Koenraad. 2018. Multiword expressions and the Law of Exception. In Manfred Sailer & Stella Markantonatou (eds.), *Multiword expressions: Insights from a multi-lingual perspective* (Phraseology and Multiword Expressions 1), 121–141. Berlin: Language Science Press. DOI: 10.5281/zenodo.1182595. (15 October, 2018).

Lambrecht, Knud. 1990. ''What, me worry?" — 'Mad Magazine sentences' revisited. In *Proceedings of the 16th Annual Meeting of the Berkeley Linguistics Society*, 215–228. Berkeley, CA: Berkley Linguistics Society.

Laporte, Éric. 2018. Choosing features for classifying multiword expressions. In Manfred Sailer & Stella Markantonatou (eds.), *Multiword expressions: Insights from a multi-lingual perspective* (Phraseology and Multiword Expressions 1), 143–186. Berlin: Language Science Press. DOI: DOI:10.5281/zenodo.1182597. (26 July, 2018).

Lødrup, Helge. 2009. Looking possessor raising in the mouth: Norwegian possessor raising with unergatives. In Miriam Butt & Tracy Holloway King (eds.), *Proceedings of the LFG09 conference*, 420–440. Stanford, CA: CSLI Publications. http://web.stanford.edu/group/cslipublications/cslipublications/LFG/14/papers/lfg09lodrup.pdf (15 October, 2018).

Markantonatou, Stella & Manfred Sailer. 2016. *Affectees in MWEs: German and Modern Greek*. Poster presented at PARSEME's 6th general meeting in Struga. https://typo.uni-konstanz.de/parseme/images/Meeting/2016-04-07-Struga-meeting/WG1-MARKANTONATOU-SAILER-poster-1.pdf (14 October, 2018).

Mateu, Jaume & Teresa Espinal. 2007. Argument structure and compositionality in idiomatic constructions. *The Linguistic Review* 24(1). 33–59. DOI: https://doi-org.proxy.ub.uni-frankfurt.de/10.1515/TLR.2007.002.

McCawley, James D. 1981. The syntax and semantics of English relative clauses. *Lingua* 53(2). 99–149. DOI: 10.1016/0024-3841(81)90014-0.

Müller, Stefan. 2010. Persian complex predicates and the limits of inheritance-based analyses. *Journal of Linguistics* 46(3). 601–655. DOI: 10.1017/S0022226709990284.

Müller, Stefan. 2014. Kernigkeit: Anmerkungen zur Kern-Peripherie-Unterscheidung. In Antonio Machicao y Priemer, Andreas Nolda & Athina Sioupi (eds.), *Zwischen Kern und Peripherie* (studia grammatica 76), 25–39. Berlin: de Gruyter. DOI: 10.1524/9783050065335.

Müller, Stefan. 2020. HPSG and Construction Grammar. In Stefan Müller, Anne Abeillé, Robert D. Borsley & Jean-Pierre Koenig (eds.), *Head-Driven Phrase Structure Grammar: The handbook*, 1383–1435. Berlin: Language Science Press. DOI: ??.

Newmeyer, Frederick J. 1974. The regularity of idiom behavior. *Lingua* 34(4). 327–342.

Nunberg, Geoffrey, Ivan A. Sag & Thomas Wasow. 1994. Idioms. *Language* 70(3). 491–538.

Pollard, Carl & Ivan A. Sag. 1987. *Information-based syntax and semantics* (CSLI Lecture Notes 13). Stanford, CA: CSLI Publications.

Pollard, Carl & Ivan A. Sag. 1994. *Head-Driven Phrase Structure Grammar* (Studies in Contemporary Linguistics). Chicago: The University of Chicago Press.

Pulman, Stephen G. 1993. The recognition and interpretation of idioms. In Cristina Cacciari & Patrizia Tabossi (eds.), *Idioms: Processing, structure, and interpretation*, 249–270. Hillsdale, NJ: Lawrence Erlbaum Associates.

Richards, Norvin. 2001. An idiomatic argument for lexical decomposition. *Linguistic Inquiry* 32(1). 183–192.

Richter, Frank. 2020. Formal background. In Stefan Müller, Anne Abeillé, Robert D. Borsley & Jean-Pierre Koenig (eds.), *Head-Driven Phrase Structure Grammar: The handbook*, 81–116. Berlin: Language Science Press. DOI: ??.

Richter, Frank & Manfred Sailer. 2009. Phraseological clauses in constructional HPSG. In Stefan Müller (ed.), *Proceedings of the 16th International Conference on Head-Driven Phrase Structure Grammar, Göttingen 2009*, 297–317. Stanford, CA: CSLI Publications. cslipublications.stanford.edu/HPSG/2009/richter-sailer.pdf.

Richter, Frank & Manfred Sailer. 2014. Idiome mit phraseologisierten Teilsätzen: Eine Fallstudie zur Formalisierung von Konstruktionen im Rahmen der HPSG. In Alexander Lasch & Alexander Ziem (eds.), *Grammatik als Netzwerk von Konstruktionen*, 291–312. Berlin: De Gruyter.

Richter, Frank & Jan-Philipp Soehn. 2006. *Braucht niemanden zu scheren*: a survey of NPI licensing in German. In Stefan Müller (ed.), *Proceedings of the 13th International Conference on Head-Driven Phrase Structure Grammar, Varna*, 421–440. http://cslipublications.stanford.edu/HPSG/2006/richter-soehn.pdf.

Riehemann, Susanne. 1997. Idiomatic constructions in HPSG. Paper presented at the 1997 HPSG conference, Ithaca. http://doors.stanford.edu/~sr/idioms.ps (28 June, 2018).

Riehemann, Susanne Z. 2001. *A Constructional approach to idioms and word formation*. Stanford University. (Dissertation). http://mally.stanford.edu/~sr/sr-diss.pdf (20 March, 2018).

Riehemann, Susanne Z. & Emily Bender. 1999. Absolute constructions: On the distribution of predicative idioms. In Sonya Bird, Andrew Carnie, Jason D. Haugen & Peter Norquest (eds.), *Proceedings of the WCCFL 18 conference*, 476–489.

Somerville, MA: Cascadilla Press. http://doors.stanford.edu/~sr/wccfl.ps (18 August, 2020).

Ruwet, Nicolas. 1991. On the use and abuse of idioms in syntactic argumentation. In John Goldsmith (ed.), *Syntax and human experience* (Studies in Contemporary Linguistics), 171–251. Edited and translated by John Goldsmith. Chicago: University of Chicago Press.

Sag, Ivan A. 1997. English relative clause constructions. *Journal of Linguistics* 33(2). 431–484. DOI: 10.1017/S002222679700652X.

Sag, Ivan A., Timothy Baldwin, Francis Bond, Ann Copestake & Dan Flickinger. 2002. Multiword expressions: A pain in the neck for NLP. In *Proceedings of the Third International Conference on Intelligent Text Processing and Computational Linguistics (CICLING 2002)*, 1–15. Mexico City, Mexico. http://lingo.stanford.edu/pubs/WP-2001-03.pdf (20 March, 2018).

Sailer, Manfred. 2000. *Combinatorial semantics and idiomatic expressions in Head-Driven Phrase Structure Grammar*. Eberhard-Karls-Universität Tübingen. (Dissertation). https://publikationen.uni-tuebingen.de/xmlui/handle/10900/46191 (18 August, 2020).

Sailer, Manfred. 2012. Phraseologie und Syntax (Head-Driven Phrase Structure Grammar). In Michael Prinz & Ulrike Richter-Vapaatalo (eds.), *Idiome, Konstruktionen, „verblümte rede". Beiträge zur Geschichte der germanistischen Phraseologieforschung*, vol. 3 (Beiträge zur Geschichte der Germanistik), 241–262. Stuttgart: S. Hirzel Verlag.

Samvelian, Pollet & Pegah Faghiri. 2016. Re-thinking compositionality in Persian complex predicates. In *Proceedings of Berkeley Linguistic Society 39th annual meeting, February 16-17 2013*, 212–227. Berkeley: Berkeley Linguistics Society.

Schenk, André. 1995. The syntactic behavior of idioms. In Martin Everaert, Erik-Jan van der Linden, André Schenk & Rob Schreuder (eds.), *Idioms: structural and psychological perspectives*, 253–271. Hillsdale NJ: Erlbaum Assoc.

Sheinfux, Livnat Herzig, Tali Arad Greshler, Nurit Melnik & Shuly Wintner. 2019. Verbal MWEs: Idiomaticity and flexibility. In Yannick Parmentier & Jakub Waszczuk (eds.), *Representation and parsing of multiword expressions: Current trends* (Phraseology and Multiword Expressions 2), 35–68. Berlin: Language Science Press.

Soehn, Jan-Philipp. 2006. *Über Bärendienste und erstaunte Bauklötze – Idiome ohne freie Lesart in der HPSG* (Deutsche Sprache und Literatur 1930). Frankfurt/Main: Peter Lang.

Soehn, Jan-Philipp. 2009. *Lexical licensing in formal grammar*. Online publication. Universität Tübingen. Universität Tübingen. http://nbn-resolving.de/urn:nbn:de:bsz:21-opus-42035 (15 August, 2018).

Swinney, David A. & Anne Cutler. 1979. The access and processing of idiomatic expressions. *Journal of Verbal Learning and Verbal Behavior* 18(3). 523–534.

Trawiński, Beata, Jan-Philipp Soehn, Manfred Sailer, Frank Richter & Lothar Lemnitzer. 2008. Cranberry expressions in English and in German. In Nicole Grégoire, Stefan Evert & Brigitte Krenn (eds.), *Proceedings of the LREC workshop towards a shared task for multiword expressions (MWE 2008)*, 35–38. Marrakech, Morokko.

Villavicencio, Aline & Ann Copestake. 2002. *On the nature of idioms*. LinGO Working Paper 2002-04. Stanford, CA: CSLI Stanford. http://lingo.stanford.edu/working-papers.html (2 September, 2019).

Wasow, Thomas, Geoffrey Nunberg & Ivan A. Sag. 1984. Idioms: An interim report. In Hattori Shiro & Inoue Kazuko (eds.), *Proceedings of the XIIIth International Congress of Linguistics*. The Hague: CIPL.

Weinreich, Uriel. 1969. Problems in the analysis of idioms. In Jaan Puhvel (ed.), *Substance and structure of language*, 23–81. Reprinted as Weinreich (1980). University of California Press.

Weinreich, Uriel. 1980. Problems in the analysis of idioms. In William Labov & Beatrice S. Weinreich (eds.), *On semantics*, 208–264. Philadelphia: University of Pennsylvania Press. DOI: 10.9783/9781512819267-007.