Gisela Harras & Kristel Proost

The Lemmatisation of Idioms

1. Introduction

The question of how idioms should be lemmatised is a fundamental issue in the lexicographic treatment of idioms and has been the focus of much debate ever since the first International Symposium on Lexicography. Several proposals for a systematic lexico-graphic treatment of idioms have been put forward (e.g. Cowie 1981, Burger 1983, Braasch 1988, Schemann 1991, Burger 1998 etc.). In this paper, we examine how semi- and non-literal idioms are lemmatised in some of the most widely-known dictionaries of German, English and Dutch. In what follows, we confine ourselves to the treatment of idioms in mono- and bilingual general dictionaries which are alphabetically ordered. Since the lexical status of idioms is relevant to the way in which idioms should be lemmatised, we shall first be concerned with the status of idioms as units of the lexicon.

2. Idioms and the Lexicon

2.1. Phrasemes

Idioms are phrasemes of a specific type. Phrasemes are lexical units characterised by the following two properties (cf. Baranov & Dobrovol'skij 1999: 64; Burger 1998: 14; Fleischer 1997: 36-37; Dobrovol'skij 1995: 14-17):

1. They are *multiword expressions*, i.e. they are lexicalised expressions consisting of more than one word. Because of this property, they are sometimes called "complex lexicalisations", a term which sets them apart from words, which are regarded as "simple lexicalisations" (cf. Verschueren 1985, 30).

2. They are *fixed expressions*, i.e. they are word combinations whose structure is more or less stable. The fixed nature of phrasemes is reflected by a variety of factors including restrictions on the applicability of grammatical rules. The property of structural stability distinguishes phrasemes from free combinations of words. Fixed expressions do not allow the whole range of morphological, lexical or syntactic variation which free word combinations permit. Some grammatical restrictions are relevant to the lemmatisation issue. Morphological and lexical restrictions especially bear directly on the way in which phrasemes should be lemmatised:

a) Restrictions on the applicability of morphological rules

The inflectional paradigms of verbs and nouns which are part of phrasemes show more deficiencies than those of verbs and nouns occurring in isolation. In some phrasemes of the category 'verb', the verb may occur only in a particular tense, as is the case in the following two examples:

> In jmdn. *ist* der Teufel *gefahren.* – verb form restricted to perfect tense
> In ihn *ist* wohl der Teufel *gefahren.* (present perfect)
> *In ihn *fährt* der Teufel. (simple present)
> *In ihn *fuhr* der Teufel. (simple past)
> *In ihn *wird* der Teufel *fahren.* (future)

> geen kaas van iets *gegeten hebben* (Dutch – literally: 'not having eaten cheese of something', meaning 'not to have the slightest idea of something') – verb form restricted to perfect tense
> → Van politiek *heeft* hij absoluut geen kaas *gegeten.* (present perfect) (Transl.: he has absolutely no idea of politics)
> *Van politiek *eet* hij geen kaas. (simple present)
> *Van politiek *at* hij geen kaas. (simple past)
> *Van politiek *zal* hij geen kaas *eten.* (future)

In some idioms containing a noun, the noun may often only appear either in the singular or in the plural ('#' indicates that the word combination may not be interpreted as a phraseme rather as a free combination of words; '?' indicates that the word combination is neither literally nor idiomatically well formed):

> spill the beans (noun restricted to plural)
> → #She usually did not spill *the bean* at once.
>
> etwas geht jemandem auf den Keks. – noun restricted to singular
> → ?Das geht mir gewaltig auf *die Kekse.*
>
> iets niet onder stoelen of banken steken (Dutch – literally: 'not to put something under chairs or benches', meaning: 'to talk openly about something') – both nouns restricted to plural
> → ?Hij stak zijn ergernis niet onder *een stoel* of *een bank.* (Literally: 'He did not put his annoyance under a chair or a bench'.)

b) Restrictions on the applicability of semantic rules

Idioms are resistant in varying degrees to internal lexical substitutions. If one or more of the words that co-occur to make up an idiom are replaced by other words belonging to the same semantic field, the word combination often loses its idiomatic meaning and is interpreted as a free combination of words. The possibilities of lexical substitution which a given phraseme permits are not predictable on the basis of selection restrictions. Selection restrictions concern entire paradigms of words such as semantic classes and lexical fields. However, words which are part of phrasemes typically do not co-occur with just any item of a given paradigm. Restrictions on the co-occurrence of words in phrasemes affect only individual elements of semantic classes or lexical fields:

seize/ grasp/ #grab the nettle

jemandem Honig um den Bart schmieren
→ jemandem Honig/ Brei/ #Marmelade um den Bart/ ums Maul/ um den Mund/ #um die Lippen schmieren

iemand in het nauw drijven (Dutch – literally: 'to push somebody into a narrow space', meaning: 'to bring somebody into a difficult situation')
→ iemand in het nauw/ in een hoek/ #tegen de muur drijven (– literally: 'to push somebody into a narrow space/ into a corner/ #against the wall')

c) Restrictions on the applicability of syntactic rules

Most phrasemes allow the application of a number of syntactic operations. Phrasemes may differ considerably with respect to the range of syntactic operations they allow (cf. Dobrovol'skij 2000a; Fraser 1970; Nunberg, Sag and Wasow 1994). Yet, phrasemes are not as flexible as free combinations of words as far as the application of syntactic operations is concerned. Though restrictions on the applicability of syntactic rules are certainly interesting from a lexicographic point of view, they are not immediately relevant to the way in which phrasemes should be lemmatised or to the selection of the citation form in which a particular phraseme should be entered.

2.2. Two types of phrasemes: Collocations and idioms

All phrasemes are complex lexicalisations with a more or less stable form. Some, but not all, phrasemes are also idiomatic to a certain degree. Idiomaticity is a semantic property which phrasemes may, but need not, have. We define 'idiomaticity' in terms of semantic non-compositionality: the meaning of an idiom is not a compositional function of the *conventional*, i.e. the literal or conventional-metaphorical, meanings of its parts. Fixed expressions may vary considerably with respect to their degree of idiomaticity. Idioms show a higher degree of idiomaticity than other types of fixed expressions (cf. Dobrovol'skij 1995: 19-20). We shall use the term "collocations" to refer to such fixed expressions which are not or only marginally idiomatic. Examples are *take/make a decision* and transparent similes like *as light as a feather* and *as dry as dust*. There is no clear borderline separating idioms from collocations.
    Assuming that fixed expressions may be ordered along a scale of compositionality, the position which collocations occupy on that scale is near the borderline separating free from fixed word combinations. Idioms occupy a position between collocations and words which are simple lexicalisations and may therefore be regarded as optimal lexicalisation instances. Idioms do not constitute a homogeneous class. They are not all equally idiomatic. Some contain components which preserve their literal meaning, i.e. the meaning these components have outside the idiom. Idioms containing literal components are less idiomatic than those whose parts all have a figurative meaning. We shall call idioms containing literal components "semi-literal" and idioms which do not contain any such components "non-literal" (cf. Burger 1998: 32). The presence of literal idiom components increases the

degree of compositionality of an idiom. This means that semi-literal idioms are always compositional.

Some non-literal idioms may also be argued to be compositional to a certain degree. Non-literal idioms may be considered compositional if their components carry identifiable parts of their idiomatic meaning. This type of compositionality is known as "decomposability" or "analysability" (cf. Nunberg, Sag & Wasow 1994). The meaning of non-literal idioms which are decomposable may be explained by a paraphrase showing that parts of the lexical structure of the idiom correspond to parts of its meaning. An example is *spill the beans* (meaning 'tell somebody something which should have been kept secret'), where *spill* may be taken to denote the action of telling and *the beans* the secret which was supposed to have been kept. Idioms like *saw logs* (meaning 'sleep'), which do not distribute their meanings to their components, are non-compositional. (See Figure 1)

Phrasemes

Collocations
(not or marginally idiomatic)

Idioms

Semi-literal

Non-literal

decomposable   not decomposable

make/take a decision
as light as a feather

promise sb.
the moon
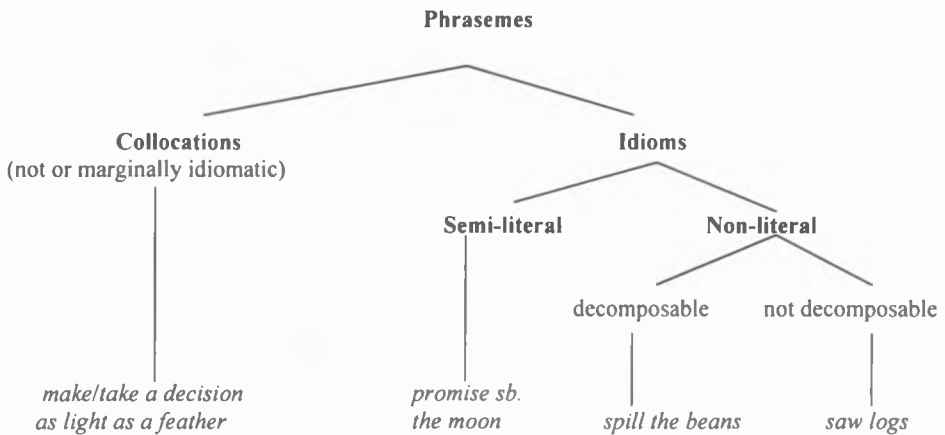
spill the beans

saw logs

Figure 1: A Classification of Fixed Expressions

The claim that many non-literal idioms are decomposable should not be taken to entail that these idioms are compositional in the same way as free combinations of words. A speaker who knows the literal, i.e. conventional, meanings of each of the components of a non-literal idiom but has never encountered the idiom would not be able to define the meaning of the idiom on the basis of the conventional meanings of its components. Speakers are capable in principle of decomposing non-literal idioms in their native language(s). They may recognise the compositionality of a non-literal idiom, for example, when they are asked what a particular idiom means. The ability to decompose non-literal idioms is part of their competence as speakers of their native language. This does not mean, though, that they also actually decompose idioms when they produce or understand sentences. A native speaker's ability to decompose an idiom when needed or to recognise the compositionality of an idiom is a matter of competence rather than performance (cf. Nunberg, Sag & Wasow 1994: 496). For this reason, the decomposability of idioms has sometimes been argued to involve a kind of "compositionality after the fact" (cf. Dobrovol'skij 2000b: 115).

Structural stability, idiomaticity and decomposability all contribute to determine the lexical status of a given phraseme. They are also directly relevant to the way in which these expressions should be lemmatised. In what follows, we shall focus on the lemmatisation of semi- and non-literal idioms, especially of the category 'verb'.

3. The Lemmatisation of Idioms: Three Crucial Issues

The following three issues are crucial to the question of how idiomatic expressions should be lemmatised (cf. Braasch 1988; Burger 1983: 57-62; Burger 1998: 170-178; Schemann 1991: 2790-2792):

1. Under which headword should an idiom be entered into a general dictionary?

2. Which position should a given idiom occupy within the entry of the headword under which it is listed? Does it make sense for the idiom to be associated with one of the senses of the headword? If so, under which of the senses of the headword should the idiom be entered?

3. What is the appropriate citation form in which the idiom should be entered?

We argue that these questions should be answered in accordance with the following four general principles:

Principle 1: The way in which idioms are lemmatised should reflect the fact that idioms are lexical units. As such they have to be treated in exactly the same way as all other lexical units entered into the dictionary, i.e. they have to be entered as headwords of their own and must be represented graphically like all other headwords (e.g. in bold type).

Principle 2: The way in which idioms are lemmatised should be compatible with theoretical insights into the lexical status of idioms.

Principle 3: The way in which idioms are lemmatised must not create additional semantic problems such as an unnecessary multiplication of the senses of the headword.

Principle 4: The way in which idioms are lemmatised should be compatible with current lexicographic practice.

We are able to follow the fourth principle only if we are sufficiently acquainted with current lexicographic practice. We shall therefore turn to the lexicographic tradition next and examine how the three questions mentioned above have been dealt with in some of the most widely-known general dictionaries of German, English and Dutch.

4. The Lexicographic Tradition

4.1. The Headword Issue

There are basically two options for entering an idiom into a general mono- or bilingual dictionary:

1) Idioms are entered and explained (or translated) under each of their components. This procedure is applied consistently in the German WDG, (Klappenbach & Steinitz (eds.) 1980). Some dictionaries such as Collins, (Hanks et al.1979), and Van Dale DN-ND (Duits Nederlands - Nederlands Duits) (Cox (ed.) 1983/1995 and 1986), use it in alternation with the second procedure, which will be illustrated under point 2. The following cases exemplify the procedure of multiple lemmatisation:

> eine Stange angeben (WDG) – entered and explained under *Stange* and *angeben*

> de loftrompet over iemand steken (Van Dale DN-ND) (Dutch – literally: 'to put the trumpet of praise over somebody'; meaning: 'to praise somebody') – entered and translated under *loftrompet* and *steken*

2) Idioms are entered and explained under only one of their components. Cross-references to the headword under which the idiom is explained appear at other idiom components. Cross-references appear systematically in some dictionaries such as the German HWDG or the English OALD (Hornby 2000) but only sporadically in others. In some dictionaries, cross-references are lacking altogether. This is true of the German Duden dictionary. When idioms are entered and explained only once, they appear under the first content word which is part of the idiom. In some German dictionaries such as Duden and DUW, idioms are entered under the first noun which is part of the idiom, regardless of whether there is another content word preceding it. In the absence of nouns, the idiom is entered under the first word which is taken to be "semantically significant". What counts as "semantically significant" words are mostly content words. The following examples illustrate the procedure according to which idioms are entered and explained only once:

> jmdm. einen Bären aufbinden (Duden & Duden/Oxford) – entered and explained (or translated) under *Bär*; no cross-reference at *aufbinden*

> frieren wie ein Schneider (Duden & Duden/Oxford) – entered and explained (or translated) under *Schneider*; no cross-reference at *frieren*

> break the ice (Duden/Oxford) – entered and translated under *ice*, no cross-reference at *break*

> spill the beans (OALD) – entered under *spill*; cross-reference at *bean*

> iemand een pluim op zijn hoed steken (Van Dale ND) (Dutch – literally: ‚to put a feather onto someone's hat', meaning: ‚praise someone') – entered and explained under *pluim*; no cross-references at *hoed* or *steken*

jmdn. über den grünen Klee loben (Van Dale DN) – entered and translated under *Klee*; cross-references at *grün* and *loben*

Some dictionaries alternate between the two procedures. Alternation is especially frequent in Collins and Van Dale DN-ND and also sporadically occurs in Duden/Oxford:

cool as a cucumber (Collins) – entered and explained under *cucumber*; no cross-references at *cool*

mad as a hatter (Collins) – entered and explained under *mad* and *hatter*

jmdn. über den grünen Klee loben (Van Dale DN) – entered and translated under *Klee*, cross-references at *grün* and *loben*

angeben wie eine Tüte Mücken (Duden/Oxford) – entered under *angeben*; no cross-references at *Tüte* or *Mücken*

These examples also illustrate some of the main problems which arise when a decision must be made with respect to the headword under which a particular idiom has to be entered:

– All of the English and German idioms listed above have the same lexical status: they are all semi-literal idioms, i.e. idioms containing a component which preserves its literal meaning. In spite of their identical lexical status, they are treated differently with respect to the headword under which they are entered. Particularly, they are entered under their literal component in one case and under their non-literal component in the other.

– Another problem is that lexical variants of one and the same semi-literal idiom are automatically separated and occur at different places in the dictionary when idioms are entered only under the first of their components which happens to be a content word:

Duden & DUW:
lügen wie gedruckt (entered and explained under *lügen* – first content word)
lügen, dass sich die Balken biegen (entered and explained under *Balken* – first noun)
jmdn. die Hucke voll lügen (entered and explained under *Hucke* – first noun)
jmdn. die Jacke voll lügen (entered and explained under *Jacke* – first noun)

– The criterion requiring that idioms be entered under the first of their components which is a content word is not always transparent to dictionary users. A dictionary user is more likely to look up an idiom like *angeben wie eine Tüte Mücken* under *angeben* than under *Tüte* or *Mücke*. Occasionally, such considerations also seem to be taken into account by lexicographers:

Duden/Oxford:
angeben wie eine Tüte Mücken
entered and explained under *angeben*
(but: frieren wie ein Schneider – entered under *Schneider*)

The representation of idioms as headwords of their own, which is required by the first principle, is treated inconsistently. When idioms appear in bold type, they are represented

as headwords of their own. Idioms occurring in italics may erroneously be taken to be examples of typical patterns of usage.

4.2. The Position of the Idiom within a Dictionary Entry

General dictionaries differ with respect to the position which a given idiom occupies within the entry of the headword under which it is listed. On the whole, two tendencies may be observed:

1) The idiom is entered under one of the senses of the headword.

> Duden & DUW:
> jmdm. einen Bären aufbinden
> entered and explained under **Bär**: 1. großes Raubtier
> (2. Rammklotz, großer Hammer)

As Burger has shown, the association of an idiom with one of the senses of the headword is often arbitrary and may lead to a remotivation of idioms which are in fact completely opaque.

> Duden & DUW:
> jmdm. einen Korb geben
> entered and explained under *Korb*: fourth sense: eine ablehnende Antwort auf ein Angebot, einen [Heirats]antrag (cf. Burger 1998: 175)

In this example, the word *Korb* has been assigned the sense 'ablehnende Antwort' (i.e. 'refusal'), because the other three senses do not allow an association of the idiom with the meaning of the headword. The problem with this case is that *Korb* does not have the meaning 'refusal' when it occurs in isolation.

2) The idiom is entered at the end of the entry, i.e. underneath all of the senses of the headword. The idiom is preceded by a lexicographic device indicating its lexical status as a specific type of fixed expression.

> OALD:
> spill the beans
> entered under *spill*
> *spill*
> 1. (usually of liquids) to accidentally flow over the edge of a container
> 2. (of people) to come out of a place in large numbers
> IDM *spill the beans*
>
> HWDG:
> jmdm. einen Korb geben
> Entered under *Korb*
> 1. geflochtener fester Behälter
> 2. Korbgeflecht

3. (Sport) Ring, an dem ein offenes Netz befestigt ist, in das der Ball beim Korbball, Basketball geworfen werden muss
4. Handschutz am Griff von Degen, Säbeln
+ jmdm. einen Korb geben, sich einen Korb holen, einen Korb bekommen

4.3. The Selection of the Citation Form

Idioms are not always entered by means of a genuine citation form. Instead, they often appear as examples where verbs occur in a specific inflected verb form and argument positions are occupied by specific nominal expressions. Listing idioms by means of examples containing specific forms is likely to create confusion, because it suggests that restrictions are stronger than they actually are:

Duden/Oxford:
under *trouble* (sense 1a):
Are you looking for trouble?
You are looking for trouble.
That's asking for trouble.

Like verbs, verbal idioms are entered in a citation form containing the infinitive form of the verb. Obligatory restrictions on the range of items which may occur in the subject position are therefore often ignored:

jemandem aufs Gemüt schlagen (Duden/Oxford, Van Dale DN & Duden 11)

This idiom is treated as if it were possible to say:

*jemand schlägt jemandem aufs Gemüt.

On the other hand, the citation form often contains too many restrictions:

hij liegt dat hij zwart ziet (de Groot /ed.) 1999: Van Dale Idioomwoordenboek)

Yet, the corresponding idiom is not restricted to this particular form. In fact, the literal component *liegen* may occur in practically all inflectional forms. Constituents which are free are often listed along with idiom components:

einen Streit vom Zaun brechen (Duden 11)

This citation form also suggests too many restrictions, because the slot occupied by *einen Streit* may in fact also be filled by a whole series of other nominal expressions:

einen Streit/ein Streitgespräch/eine Debatte/eine Diskussion/ein Gespräch.
But: ˀeine Rede/ ˀeine Gardinenpredigt vom Zaun brechen.

In this case, the range of items which may occur in the argument position seems to be predictable on the basis of selection restrictions. The idiom may be more adequately represented in a form which allows for more flexibility:

X vom Zaun brechen (where 'X' stands for some kind of communicative event involving two or more participants.)

If what appears in the argument position is really predictable by selection restrictions, the constituent represented by "X" is an idiom-external rather than an idiom-internal component.

5. Maxims for the Lemmatisation of Idioms

To avoid lexicographic inconsistencies, we suggest the following set of maxims for the lemmatisation of idioms. The lemmatisation maxims we propose are based on theoretical insights into the lexical status of idioms.

5.1. The Headword Maxim

The way in which idioms are lemmatised should reflect their lexical status. This means that semi- and non- literal idioms have to be treated differently. Semi-literal idioms should be entered under the idiom component which preserves its literal meaning. Cross-references may appear at the other idiom components to lead the user to the literal component where the idiom is entered and explained. Examples:

promise somebody the earth/the moon (entered under *promise* – cross-references at *earth/moon*)

jemandem den Himmel auf Erden versprechen (entered under *versprechen* – cross-reference at *Himmel* and *Erde*)

jemanden über den grünen Klee loben (entered under *loben* – cross-references at *grün* and *Klee*)

angeben wie ein Wald voll Affen (entered under *angeben* – cross-references at *Wald* and *Affen*)

frieren wie ein Schneider (entered under *frieren* – cross-reference at *Schneider*)
jemands Lob singen (entered under *Lob* – cross-reference at *singen*)

sing somebody's praises (entered under *praise* – cross-reference at *sing*)

Non-literal idioms are entered under the first of their components which happens to be a content word. Cross-references may appear under other idiom components to guide the user to the component under which the idiom is entered and explained.

jmdm. einen Bären aufbinden (entered under *Bär*; cross-reference at *aufbinden*)

kill two birds with one stone (entered under *kill*; cross-references at *bird* and *stone*)

de koe bij de hoorns vatten (Dutch – literally: 'fetch the cow by the horns'; meaning: 'deal with a difficult situation directly') (entered and explained under *cow*; cross-references at *hoorn* and *vatten*)

The lemmatisation of German and Dutch non-literal idioms may require a slightly different procedure. To conform as much as possible to current lexicographic practice, German and Dutch non-literal idioms may have to be entered and explained under the first noun which is part of the idiom. For most German and Dutch idioms of the category VP, lemmatisation under the first content word coincides with lemmatisation under the first noun (as in *jemandem einen **Bären** aufbinden*). This is due to the fact that the word order pattern of the citation form is identical with that of German and Dutch subordinate clauses. Word order is SOV in German and Dutch subordinate clauses, i.e. the complement precedes the head verb of the verb phrase. Since SOV is also the word order pattern of the citation form in German as well as in Dutch, the first content word of the citation form is usually a noun, i.e. the head noun of the NP complement preceding the verb. Exceptions are idioms containing a noun preceded by an adjective (e.g. *kleine Brötchen backen*). When such idioms are lemmatised, the word preceding the noun is usually ignored.

Lemmatisation under the first noun would ignore the word order properties of English. Word order in English is SVO in main as well as in subordinate clauses. Thus the complement of a verb follows the head verb within the VP. Lemmatisation of English non-literal idioms under the first noun would require that the preceding head verb of the verb phrase be ignored in idioms like *kill two birds with one stone* and *count one's chickens before they are hatched*. This is likely to be the reason why **systematic** lemmatisation only under the first noun of a VP-idiom did not occur in any of the English dictionaries we examined. These observations suggest that the selection of a citation form should also be guided by the word order properties of a particular language.

According to the headword maxim, idioms are entered under a given headword depending on whether that idiom is a semi- or non-literal idiom. The distinction between semi- and non-literal idioms has the following advantages:

– In the large majority of cases, the criterion of literal meaning is transparent both to lexicographers and for dictionary users. Lexicographers are likely to find it easy to apply it consistently. The dictionary user has to be conceived of as someone looking up **an idiom**, and not, for example, an isolated word. This means that the dictionary user must have recognised the formal structure of an idiom in order to be able to consult a dictionary. As Wittgenstein said: Somebody asking about the meaning of a word in a particular language must already know a lot about that language.

– The distinction between semi- and non-literal idioms enables us to deal with a variety of idiomatic expressions in one and the same way. It may be applied to all similes of the form [V *like* XP] or [(as) A as NP] as well as to phrasemes of the type [A + N]:

> sich schwarz / grün und blau / grün und gelb ärgern (all entered and explained under the literal component *ärgern*)

> dumm wie Bohnenstroh (entered and explained under the literal component *dumm*)

> blinder Passagier (entered and explained under *Passagier,* because the meaning of *blinder Passagier* is an extension of the meaning of *Passagier*)

> Iron Curtain (entered and explained under *iron,* because the expression is a non-literal idiom)

blind date (entered and explained under *date*, because the expression is a semi-literal idiom)

– Treating semi- and non-literal idioms differently means that there are two different principles for lemmatising idioms where we might have done with only one. However, the total number of lemmatisation principles for phrasemes does not increase, because semi-literal idioms are lemmatised in exactly the same way as collocations. Semi-literal idioms and collocations are both entered under their literal component.

5.2. The Position in Entry Maxim

This maxim is irrelevant to the lemmatisation of semi-literal idioms: semi-literal idioms appear under the component which preserves its literal meaning.

Non-literal idioms should appear at the end of the entry, i.e. underneath all of the senses of the headword. This procedure avoids the association of one of the idiom components with one of the senses of the headword under which it is listed. It thereby also avoids an unnecessary multiplication of the senses of the headword (such as the assignment of the reading 'ablehnende Antwort' to *Korb* in *jemandem einen Korb geben*).

5.3. The Citation Form Maxim

Idioms should basically be entered in their basic or canonical form. This means that the citation form should contain only general pronouns like *someone* or *somebody* and *something*. VP-idioms should basically be entered in the infinitive form of the head verb. Where deviations from the canonical citation form are required, these should be in accordance with the following submaxims:

(1) The citation form must indicate as many restrictions a possible:

The range of arguments which a given VP-idiom permits cannot be predicted on the basis of the meaning of its head verb. For example, the arguments of the German idiom *Etwas schlägt jemandem auf den Magen* are not predictable from the lexical meaning of the verb *schlagen*. The verb *schlagen* subcategorises the external argument *jemand* and the internal argument *jemandem*, but the idiom does not allow *jemand* in the subject position. Obligatory restrictions on the range of items which may occur in the subject position should be indicated by the citation form:

etwas hält sich die Waage. (not: sich die Waage halten)

etwas schlägt jemandem auf den Magen. (not: jemandem auf den Magen schlagen)

iets komt uit de lucht gevallen. (Dutch – literally: ‚something is falling down from the air'; meaning: 'something happens quite unexpectedly') (not: *uit de lucht gevallen komen* or *uit de lucht vallen*)

(2) Morphological restrictions should also be indicated by the citation form:

> jemandem ist eine Laus über die Leber gelaufen.

> geen kaas van iets gegeten hebben.

(The verb form is restricted to occurring in the perfect tense.)

(3) The citation form should not be too restrictive:

> etwas vom Zaun brechen – not: einen Streit vom Zaun brechen (*Einen Streit* is only one possibility among many)

> iemand door de mangel halen (Dutch – literally: 'to draw somebody through the mangle'; meaning: 'to question somebody sharply' or 'to reprimand somebody severely') – not: door de mangel gehaald worden (as it is entered in Van Dale Idioomwoordenboek) (This idiom is not restricted to occurring in the passive voice.)

Text corpora often contain examples of uses of idioms where syntactic and morphological restrictions are not observed. An example is *Er frisst immer wieder einen Narren an ihr*, which is a variant of the idiom *einen Narren an jemandem gefressen haben*. In this German idiom, the verb is restricted to occurring in the perfect tense. Such cases typically occur in contexts which are highly marked, and it is up to the lexicographer to decide whether these phenomena should also be taken into account. In our view, marked forms should not appear as citation forms.

6. Conclusion

Theoretical issues such as the lexical status of idioms are relevant to the way in which idioms should be lemmatised. This is especially true of the distinction between semi- and non-literal idioms, which is transparent to lexicographers as well as dictionary users.

Whether idioms are decomposable or not turned out to be irrelevant to the lemmatisation issue. Taking the decomposability of idioms into account may even cause additional semantic problems, such as an unnecessary multiplication of the senses of the headword. Difficulties arising from the association of idioms with one of the senses of a particular headword may be avoided when the idiom is entered at the end of the entry, i.e. underneath all of the senses of the headword.

Relevant to the selection of the citation form are obligatory restrictions on the range of expressions which may occur in a given argument position as well as the distinction between idiom-internal and idiom-external components.

References

Baranov, Anatolij & Dmitrij Dobrovol'skij (1999): "Idioms from a Cognitive Perspective", in: *Moscow State University Bulletin* 19, 1/1999, 64-75. Moscow: Moscow State University.

Braasch, Anna (1988): "Zur lexikographischen Kodifizierung von Phrasemen in einsprachigen deutschen Wörterbüchern aus der Sicht eines ausländischen Wörterbuchbenutzers", in: Karl Hyldgaard-Jensen and Arne Zettersten (eds.): *Proceedings of the Fourth International Symposium on Lexicography*, April 20-22, 1988, University of Copenhagen, Lexicographica Series Maior; 26, 83-100. Tübingen: Niemeyer.

Burger, Harald (1983): „Phraseologie in den Wörterbüchern des heutigen Deutsch", in: Herbert Ernst Wiegand (ed.): *Studien zur Neuhochdeutschen Lexikographie III*, 13-66. Hildesheim: Olms.

– (1998): *Phraseologie: Eine Einführung am Beispiel des Deutschen,* Grundlagen der Germanistik; 36. Berlin: Schmidt.

Cox, H. L. (ed.) (1986): *Van Dale Groot woordenboek Nederlands-Duits.* Utrecht: Van Dale Lexicografie.

– (ed.) (1983/1995): *Van Dale Groot woordenboek Duits-Nederlands,* tweede druk. Utrecht: Van Dale Lexicografie.

Cowie, Anthony P. (1981): „The Treatment of Collocations and Idioms in Learner's Dictionaries", in: *Applied Linguistics* Vol. II, No. 3, 223-235.

Dobrovol'skij, Dmitrij (1995): *Kognitive Aspekte der Idiom-Semantik: Studien zum Thesaurus deutscher Idiome,* Eurogermanistik; 8. Tübingen: Narr.

Dobrolvol'skij, Dmitrij (2000a): „Syntaktische Modifizierbarkeit von Idiomen aus lexikographischer Perspektive", in: Ulrich Heid et al. (eds.): *The Ninth EURALEX International Congress, August 8-12, 2000, Universität Stuttgart, Proceedings,* Vol. II, 557-568. Stuttgart: Universität Stuttgart.

Dobrolvol'skij, Dmitrij (2000b): „Ist die Semantik von Idiomen nichtkompositionell?" in: Susanne Beckmann, Peter-Paul König und Georg Wolf (eds.): *Sprachspiel und Bedeutung: Festschrift für Franz Hundsnurscher zum 65. Geburtstag,* 113-124. Tübingen: Niemeyer.

DUDEN Das große Wörterbuch der deutschen Sprache in zehn Bänden (1999), 3, völlig neu bearbeitete und erweiterte Auflage, herausgegeben vom Wissenschaftlichen Rat der Dudenredaktion. Mannheim: Dudenverlag.

DUDEN Deutsches Universalwörterbuch A-Z. (1996), 3. neu bearbeitete und erweiterte Auflage, auf der Grundlage der neuen amtlichen Rechtschreibregeln, bearbeitet von Günther Drosdowski und der Dudenredaktion. Mannheim: Dudenverlag.

DUDEN Redewendungen und sprichwörtliche Redensarten: Wörterbuch der deutschen Idiomatik (1998), bearbeitet von Günther Drosdowski und Werner Scholze-Stubenrecht, nach den Regeln der neuen deutschen Rechtschreibung überarbeitet, Nachdruck der 1. Auflage. Mannheim: Dudenverlag.

DUDEN/OXFORD Großwörterbuch Englisch (1990), herausgegeben von der Dudenredaktion und Oxford University Press. Mannheim: Dudenverlag.

Fleischer, Wolfgang (1997): *Phraseologie der deutschen Gegenwartssprache,* 2. durchgesehene und ergänzte Auflage. Tübingen: Niemeyer.

Fraser, Bruce (1970): "Idioms within a Transformational Grammar", in: *Foundations of Language* 6. 22-42.

de Groot, Hans (ed.) (1999): *Van Dale Idioomwoordenboek: Verklaring en herkomst van uitdrukkingen en gezegden.* Utrecht: Van Dale Lexicografie.

Handwörterbuch der deutschen Gegenwartssprache in zwei Bänden (1984). Von einem Autorenkollektiv unter der Leitung von Günter Kempcke. Berlin: Akademie-Verlag.

Hanks, Peter et al. (1979): *Collins English Dictionary.* London: Collins.

Hornby, A. S. (2000): *Oxford Advanced Learner's Dictionary of Current English,* ed. Sally Wehmeier. Oxford: Oxford University Press.

Klappenbach, Ruth and Wolfgang Steinitz (eds.) (1980): *Wörterbuch der deutschen Gegenwartssprache* (1980), 10., bearbeitete Auflage. Berlin: Akademie-Verlag.

Nunberg, Geoffrey; Ivan A. Sag and Thomas Wasow (1994): „Idioms", in: *Language* 70, 491-538.

Schemann, Hans (1991): „Die Phraseologie im zweisprachigen Wörterbuch“, in: Franz Josef Hausmann et al. (eds): *Wörterbücher*: *Ein Internationales Handbuch zur Lexikographie*, 3. Teilband. 2789-94. Berlin: de Gruyter.

Verschueren, Jef (1985): *What People Say They Do with Words*: *Prolegomena to an Empirical-Conceptual Approach to Linguistic Action*, Advances in Discourse Processes; 14. Norwood, NJ: Ablex.