



Universiteit
Leiden
The Netherlands

MIPVU: A manual for identifying metaphor-related words

Steen, G.J.; Dorst, A.G.; Herrmann, J.B.; Kaal, A.A.; Krennmayr, T.; Pasma, T.; ... ; Reijnierse, W.G.

Citation

Steen, G. J., Dorst, A. G., Herrmann, J. B., Kaal, A. A., Krennmayr, T., & Pasma, T. (2019). MIPVU: A manual for identifying metaphor-related words. In S. Nacey & W. G. Reijnierse (Eds.), *Converging Evidence in Language and Communication Research* (pp. 23-40). Amsterdam: John Benjamins. Retrieved from <https://hdl.handle.net/1887/3308917>

Version: Publisher's Version

License: [Licensed under Article 25fa Copyright Act/Law \(Amendment Taverne\)](#)

Downloaded from: <https://hdl.handle.net/1887/3308917>

Note: To cite this publication please use the final published version (if applicable).

MIPVU: A manual for identifying metaphor-related words

Gerard J. Steen¹, Aletta G. Dorst², J. Berenike Herrmann³,
Anna A. Kaal⁴, Tina Krennmayr⁴ and Tryntje Pasma⁵

¹University of Amsterdam, The Netherlands / ²Leiden University,
The Netherlands / ³Basel University, Switzerland / ⁴VU University
Amsterdam, The Netherlands / ⁵formerly affiliated with VU University
Amsterdam, The Netherlands

This chapter presents the complete procedure for finding metaphor-related words which has been utilized in our research. The style is in the form of a set of instructions. The present chapter is intended to be an independent presentation of the procedure as an autonomous tool. It may be used as a reference manual by anyone who aims to find metaphor-related words in usage. The term ‘metaphor-related words’ is used to suggest that the tool aims to identify all words in discourse that can be taken to be lexical expressions of underlying cross-domain mappings.

2.1 The basic procedure

The goal of finding metaphor in discourse can be achieved in systematic and exhaustive fashion by adhering to the following set of guidelines.

1. Find metaphor-related words (MRWs) by examining the text on a word-by-word basis.
 - For information about whether an expression counts as a word, consult Section 2.2.
2. When a word is used indirectly and that use may potentially be explained by some form of cross-domain mapping from a more basic meaning of that word, mark the word as metaphorically used (MRW).
 - For information about indirect word use that is potentially explained by cross-domain mapping, consult Section 2.3.

3. When a word is used directly and its use may potentially be explained by some form of cross-domain mapping from a more basic referent or topic in the text, mark the word as direct metaphor (MRW, direct).
 - For more information about direct word use that is potentially explained by cross-domain mapping, consult Section 2.4.
4. When words are used for the purpose of lexico-grammatical substitution, such as third person personal pronouns, or when ellipsis occurs where words may be seen as missing, as in some forms of co-ordination, and when a direct or indirect meaning is conveyed by those substitutions or ellipses that may potentially be explained by some form of cross-domain mapping from a more basic meaning, referent, or topic, insert a code for implicit metaphor (MRW, implicit).
 - For more information about implicit meaning by substitution or ellipsis that is potentially explained by cross-domain mapping, consult Section 2.5.
5. When a word functions as a signal that a cross-domain mapping may be at play, mark it as a metaphor flag (MFlag).
 - For more information about signals of cross-domain mappings, consult Section 2.6.
6. When a word is a new-formation coined by the author, examine the distinct words which are its independent parts according to steps 2 through 5.
 - For more information about new-formations coined by the author, consult Section 2.7.

The use of the phrase ‘potentially explained by a cross-domain mapping’ is intentional. It should be read with an emphasis on ‘potentially’. This links up with the tenuous connection between linguistic and conceptual metaphor identification discussed in Chapter 1 of Steen et al. (2010).

As for the relation with MIP (Pragglejaz Group 2007), points 1 and 2 are essentially the same as MIP. Points 3 and 4 deal with two additions to MIP in the area of other forms of metaphor. Point 5 is a different kind of addition to MIP and includes the identification of signals of metaphor. And point 6 takes one assumption of MIP to its linguistic conclusion by including instructions for handling new lexical units that consist of more than one free morpheme.

2.2 Deciding about words: Lexical units

The word is the unit of analysis which is examined for metaphorical use. There are other possibilities, such as the morpheme or the phrase, and these can account for additional metaphor in usage. However, we do not mark these other possibilities, because we can only do one thing at a time. Focusing on the word as the unit of analysis is already a most challenging and complex operation. It is motivated by

the functional relation between words, concepts and referents in discourse analysis, described in Chapter 1 of Steen et al. (2010).

A systematic and explicit approach to the relevant unit of analysis is crucial for a consistent and correct quantitative analysis of the data. Lack of clear guidelines may introduce a substantial degree of error and therefore noise into the numbers and patterns obtained. It would undermine detailed quantitative comparison between distinct studies.

For theoretical reasons, we will call the word a ‘lexical unit’. In adopting this terminology, we follow the Praggeljaz Group (2007). When you decide about the boundaries of lexical units, the following guidelines should be adopted.

2.2.1 General guideline

In our project, the data come from the British National Corpus, and we therefore follow most of BNC practice in deciding what counts as a lexical unit. In other projects with other materials, these guidelines may or may not have to be adjusted to the other source, as shown for Dutch in Chapter 7 of Steen et al. (2010).¹ In our research, the dependence on these materials means two things:

1. All words provided with an independent Part-Of-Speech (POS) tag in the corpus are taken as separate lexical units.
For instance, prepositions are coded as PRP, nouns are coded as NN, and so on. A full list of tags is available from the BNC website: www.natcorp.ox.ac.uk.
2. All so-called polywords in the corpus are taken as single lexical units.
There are a number of fixed multiword expressions that are analyzed as one lexical unit in the BNC, on the grounds that they are grammatical units which designate one specific referent in the discourse. Examples include *a good deal*, *by means of*, and *of course*. These multiword expressions are called polywords. They have special tags and are available in a finite list from the BNC website: www.natcorp.ox.ac.uk. You should follow this practice and, in particular, not examine the parts of these polywords for potential metaphorical meaning.

2.2.2 Exceptions

There are three exceptions to our overall acceptance of BNC practice: phrasal verbs, some compounds, and some proper names.

Phrasal verbs are verbal expressions consisting of more than one word, such as *look up* or *turn on*. These are not taken as single lexical units in the BNC, but as independent verbs followed by autonomous adverbial particles. We will not follow

1. Editors’ note: This is Chapter 5 of the present volume.

this practice, for phrasal verbs function as linguistic units designating one action, process, state or relation in the referential dimension of the discourse. In that respect, they are similar to polywords.

You should therefore treat all phrasal verbs as single lexical units: their individual parts do not require independent analysis for potential metaphorical meaning. The phrasal verb as a whole, however, can still be used metaphorically. For instance, setting up an organization is a metaphorical variant of setting up a roadblock. The classification of two or more words as part of one phrasal verb should be marked as such in the data.

The problem with phrasal verbs is their superficial resemblance to prepositional verbs (i.e. a frequent verb-preposition combination) and to verbs followed by free adverbs. The latter two cases should be analyzed as free combinations consisting of two independent lexical units, as opposed to phrasal verbs which should be taken as only one. Again, the motivation for this approach is the assumption of a functional and global correspondence between words, concepts, and referents.

One way to tell these three groups apart is by examining their POS tags in the BNC. Particles of phrasal verbs have received an AVP code, prepositions of prepositional verbs a PRP code, and freely occurring adverbs an AV0 code. These are classifications which have been made independently of any questions about metaphorical use; they are based on a general approach to data analysis, which is a bonus.

However, the matter is further complicated in three ways. When we go to the dictionaries used in our research for examining contextual and basic meanings, it appears that they do not distinguish between phrasal verbs and prepositional verbs. They in fact call both types phrasal verbs. An example is *look at* in a sentence like ‘it was only when you looked at their faces that you saw the difference’. According to Macmillan this is a phrasal verb, but the BNC code for *at* is PRP, indicating that it is a prepositional verb. If you follow the BNC’s decision, it means that you have to analyze *look* and *at* as two lexical units and independently examine their main senses in the dictionary to find their respective basic meanings; the contextual meaning of each of them in their combined use, even as a prepositional verb, however, will be found under the phrasal meaning of the combination.

Secondly, some of the verb+particle combinations marked as such in the BNC are in fact not conventionalized phrasal verbs. That is, they are not phrasal verbs according to the dictionary. An example is *look up* in a sentence like ‘she looked up into the sky’. Here *up* is coded as AVP in the BNC, suggesting that this is a proper phrasal verb. However, the Macmillan dictionary tells us that the contextual meaning – “to direct your eyes towards someone or something so that you can see them” – is not one of the meanings of the phrasal verb (unlike, for instance “to try to find a particular piece of information”). The contextual meaning, instead, is the

result of a free combination of a verb plus an adverb. BNC has probably made a mistake here; the words consequently have to be analyzed as two separate lexical units.

Thirdly, there is the matter of complex phrasal verbs, such as *make up for* or *do away with*. These may be easily confused with combinations of simple phrasal verbs with a preposition (*make up + for* or *do away + with*). However, they are typically listed as complete, complex phrasal verbs in the Macmillan dictionary, as run-ons after the main verb, and they can be replaced by a synonym (*compensate* and *get rid of*). Because of this referential unity, we follow the dictionary for complex phrasal verbs and take the dictionary classification of these complex verbs as single units as our guideline.

Taking all of this into consideration, we have established the following rules for simple phrasal verbs (complex phrasal verbs being recognizable by the criteria above):

- a. If the POS tag is PRP then we are dealing with a prepositional verb → analyze the verb and the preposition separately (i.e. two lexical units).
- b. If the POS tag is AVP then check in the dictionary whether the combination of verb+particle has been listed as a phrasal verb meaning in the relevant contextual meaning
 - if this is the case, then we accept it is a phrasal verb and analyze the combination as one lexical unit;
 - if this is not the case, then we do not take the combination to be a conventionalized phrasal verb and therefore we analyze the verb and the particle separately (i.e. two lexical units).
- c. If the POS tag is AV0 then we are dealing with a verb followed by a free adverb → analyze as two lexical units.
- d. If the POS tag is PRP/AVP then apply the tests below to determine whether we are dealing with a phrasal or a prepositional verb; as it happens, these are all **transitive** constructions.
- e. If the BNC code is clearly wrong (supported by the above criteria or the tests below) then apply the proper analysis and add a comment in the materials stating “incorrect POS tag: PRP not AVP”.

Tests for deciding between phrasal/prepositional verbs

In **prepositional** verbs:

- The preposition and following noun can be moved to the front of the sentence, which is not possible with phrasal verb particles (e.g. *Up into the sky she looked* but not **Up the information she looked*).

- An adverb can be inserted before the preposition (e.g. *She ran quickly down the hill but not *She ran viciously down her best friends*).
- The preposition can be moved to the front of a *wh*-word (e.g. *Up which hill did he run? but not *Up which hill did he run?*).

In phrasal verbs:

- The adverbial particle can be placed before or after the noun phrase acting as object of the verb, which is not possible for the prepositional verbs (e.g. *She looked the information up but not *She looked his face at*).
- If the noun phrase is replaced by a pronoun, the pronoun has to be placed in front of the particle (e.g. *The dentist took all my teeth out > The dentist took them out but not She went through the gate > *She went it through*).

Compounds are single lexical units consisting of two distinct parts, which may cause orthographical problems. They can be spelt in three ways: as one word, as two hyphenated words, and as two separate words.

- a. When a compound noun is spelt as one word, such as *underpass*, and can be found as such in the dictionary we treat it as one lexical unit designating one referent in the discourse.
- b. When a compound noun is spelt as two hyphenated words and can be found as such in the dictionary, such as *pitter-patter*, we similarly treat it as one lexical unit. However, if we are dealing with a novel formation unknown to the dictionary, the compound noun is analyzed as two separate units, even though it may have one POS tag in the corpus. Our reason for this practice is that the language user is forced to parse the compound into its two component parts in order to establish the relation between the two related concepts and referents. This also applies to hyphenated compound nouns created through a productive morphological rule but that are not listed as a conventionalized compound in the dictionary (such as *under-five*),
- c. In the BNC, compound nouns that have been spelt as two separate words are not taken as single lexical units, but analyzed as combinations of two independent words which each receive their own POS tags. When such compounds are conventionalized and, again, function as lexical units designating one referent in the discourse, we will not follow the BNC solution. For then they are like polywords, and should be treated as single lexical units, whose parts do not require analysis for potential metaphorical meaning.

The Macmillan dictionary has a tell-tale signal for identifying conventionalized compounds that are spelled as two distinct words: when a fixed expression is taken

to be a compound noun, there is primary stress on the first word and secondary stress on the second word (e.g. ‘**power** **plant**'). In cases where the Macmillan dictionary treats a multiword combination as having one meaning, but displays a reversed stress pattern (such as ‘*nuclear*’ *power*), we do not treat the multiword expression as a compound noun, and analyze it as consisting of two separate lexical units.

- Rules a and b also apply to compound adverbs and adjectives, such as *honey-hunting*. This example is a novel formation unknown to Macmillan. Therefore, following rule b, the adjective is analyzed as comprising two separate lexical units, even though BNC has given it one POS tag.
- Words may be spelt in more than one way, which may cause problems about the independent status of their components in some cases. An example is when the preposition *onto* is spelt as two words instead of one. When this happens, we will adhere to the spelling of the dictionary instead of the spelling of the document under analysis, because the dictionary is the more general reference work and related to accepted norms for language users. You should therefore analyze words according to their spelling in the dictionary, not according to their spelling and POS tagging in the corpus.

Proper names appear to form a special group in our analyses. There are several subclasses which we have encountered, which may not all technically qualify as genuine proper names. They will be discussed one by one. In general, however, proper names do not require any specific additional coding.

Our general strategy is to reduce the number of exceptions to POS tagging as provided by the BNC corpus. The solution to annotation problems proposed below is maximally simple: every separate word will be treated as a separate lexical unit, except for the underlined cases.

- a. Proper names: all parts of genuine proper names are to be treated in the way of regular POS tagging. That is, *Roy Wood* and *Madame Mattli* are coded as two separate words and taken as two lexical units. This can be extended to addresses, with house numbers as well as road names all being cut up into separate lexical units. As a result, *New York* (in *New York Herald Tribune*) is also two units.
- b. Some proper names have been bestowed on public entities and may appear in the dictionary. If they do, they are to be treated as all other expressions in the dictionary: thus, *Labour Party* becomes one lexical unit because it has the stress pattern of a compound.

The same holds for some titles that appear in the dictionary, such as *Pulitzer Prize*, which is also treated as one lexical unit on the basis of the stress pattern.

In our annotations, these expressions should be treated similarly to phrasal verbs, compounds, and polywords and should therefore receive a code to indicate that the words form single lexical units.

Green Paper and *White Paper*, by contrast, are to be treated as containing two lexical units, because they have rising stress (*Green* and *White* would always be marked as related to metaphor, and *Paper* would always be treated as metonymically used).

The elements of names of countries (e.g. *United Kingdom*) and organizations (e.g. *United Nations*) that have rising stress in the dictionary should also be treated as separate units.

- c. Other names and titles do not appear in the dictionary. They are also treated as composites of their independent words, both by the BNC and by us. This accounts for two lexical units in *Labour Law*, *Executive Committee*, *European Plan*, *Scarman Report*, and even more lexical units in *the Student Winter Games*, *the Henley Royal Regatta*, *the Criminal Law Revision Committee*, *House of Oliver*, and so on.
- d. A separate problem is constituted by genuine titles, that is, titles of texts:
 - If titles are used as titles, that is, as headings of newspaper articles or chapters and sections of novels and academic writing, they need to be taken on a word-by-word basis. This is because they summarize or indicate content by means of words, concepts, and referents. They are regular cases, if linguistically sometimes odd.
 - If titles are mentioned, however, to refer to for example a text or a TV programme, they function as names, like proper names. If they are in the dictionary, check their stress pattern; if they are not, use BNC-Baby as a guide.

2.3 Indirect use potentially explained by cross-domain mapping

Indirect use of lexical units which may be explained by a cross-domain mapping is basically identified by means of MIP, with some adjustments. This means that the following guidelines should be adopted.

1. Identify the contextual meaning of the lexical unit.
 - For more information, see Section 2.3.1.
2. Check if there is a more basic meaning of the lexical unit. If there is, establish its identity.
 - For more information, consult Section 2.3.2.

3. Determine whether the more basic meaning of the lexical unit is sufficiently distinct from the contextual meaning.
 - For more information, see Section 2.3.3.
4. Examine whether the contextual meaning of the lexical unit can be related to the more basic meaning by some form of similarity.
 - For more information, consult Section 2.3.4.

If the results of instructions 2, 3, and 4 are positive, then a lexical unit should be marked as a metaphor-related word ('MRW'), which may be made more precise by adding the information that it is 'indirect' (as opposed to 'direct' or 'implicit', see below).

2.3.1 Identifying contextual meanings

The contextual meaning of a lexical unit is the meaning it has in the situation in which it is used. It may be conventionalized and attested, and will then be found in a general users' dictionary; but it may also be novel, specialized, or highly specific, in which case it cannot be found in a general users' dictionary.

When you identify the contextual meaning of a lexical unit, several problems may arise.

1. When utterances are not finished, there is not enough contextual knowledge to determine the precise intended meaning of a lexical unit in context. In such cases, it may be that the lexical unit has been used indirectly on the basis of a metaphorical mapping, but this is impossible to decide. In such cases, we will discard for metaphor analysis all relevant lexical units in aborted utterances.

An example is 'Yeah I had somebody [come round] and *stuck* their *bloody* ...' The lexical units in the incomplete utterance in question (beginning with *stuck*) that could or could not have been related to metaphor should be marked as Discarded For Metaphor Analysis (add code 'DFMA' to each of them).

2. When there is not enough contextual knowledge to determine the precise intended meaning of a lexical unit in context, it may be that it has been used indirectly on the basis of a metaphorical mapping, but this may be impossible to decide.
 - a. An example is the use of *up* to indicate movement towards, where it is possible that the target is either higher (not metaphorical) or not higher (metaphorical) than the speaker.
 - b. Another example is the use of idioms such as *gasp for breath* or *turn your shoulder*, approached as three lexical units, where it is possible that the designated action in fact takes place and thereby stands for the emotion

(metonymy), or the designated action in fact does not take place so that the phrase is used metaphorically to indicate the concomitant emotion.

- c. A third example involves anaphora which may be interpreted in more than one way, as in ‘all that’ in the following example, where a possible metaphorical interpretation is applicable: ‘he said I come to sup be supervisor he said, I don’t know, I don’t wish to learn all that!’

In such cases of lack of situational knowledge but with a potential for metaphorical meaning, you have to treat the word as if it was used indirectly and metaphorically, on the basis of the general rule ‘When In Doubt, Leave It In’ and add the special code ‘WIDLII’.

3. Specialist terminology may constitute a specific case of insufficient contextual knowledge to determine the precise intended meaning of a lexical unit in context. When there is not enough contextual knowledge to determine the specific technical and/or scientific meaning of a word in context, regular dictionaries cannot help. In such cases, it would of course be possible to use other, preferably specialized dictionaries to find out the specific contextual meaning of a term. However, in our project we assume that metaphor is ‘metaphor to the general language user’: if we as general language users cannot establish the meaning of the lexical unit with the contemporary dictionaries alone but the lexical unit could be metaphorical on the basis of some contextual meaning projected from the basic–nontechnical–meaning, we also mark the word as metaphor-related based on ‘WIDLII’.
4. Sometimes the contextual meaning of a lexical unit may be taken as either metaphorical or as not metaphorical. This seems to be the case for many personifications, such as furious debate or this essay thinks. These examples may be analyzed as involving a metaphorical use of furious and thinks, respectively, but they may also be resolved by a metonymic interpretation of the other terms, i.e. debate and essay, in which case furious and thinks automatically turn not metaphorical. In such cases, the possibility of the metaphorical interpretation should not be lost, and you should mark the relevant ambiguous words furious and thinks as metaphor-related words, and add a comment that this is due to a possible personification.

2.3.2 Deciding about more basic meanings

A more basic meaning of a lexical unit is defined as a more concrete, specific, and human-oriented sense in contemporary language use. Since these meanings are basic, they are always to be found in a general users’ dictionary. A meaning cannot be more basic if it is not included in a contemporary users’ dictionary.

From a linguistic point of view, a more basic meaning of a word is its historically older meaning. However, from a behavioural point of view, this definition may not be optimal. Most language users are not aware of the relative ages of the various meanings of most words in the contemporary language. This means that the linguistic notion of basic sense as the historically prior sense has little relevance to the behavioural, in particular cognitive notion of basic sense.

However, it is one of the fundamental claims of contemporary metaphor theory that most of the historically older meanings of words are also more concrete, specific, and human-oriented. This is explained by the cognitive-linguistic assumption of experientialism. As a result, concrete meanings are typically also basic meanings from a historical perspective.

The still largely programmatic assumption of a connection between historically prior meanings and concrete, specific, and human-oriented meanings makes it possible for us to adopt one practical and consistent general starting point about basic meanings: they can be operationalized in terms of concrete, specific, and human-oriented meanings. This is our general definition for basic meanings.

As a result, we will not check the history of each lexical unit as an integral part of our procedure. This is a huge practical advantage, which is based in general cognitive linguistic practice. Diachronic considerations of basic meanings may only come in when specific problems arise.

When attempting to find basic meanings in the dictionary, the following guidelines should be adopted.

1. A more basic sense has to be present for the *relevant grammatical category of the word form* as it is used in context. This is because a grammatical category in a text specifies a particular class of concept and referent, which may not be altered when looking for basic meanings, for otherwise the basis of comparison is shifted. When the dictionary shows that a word may be used in more than one grammatical category, you hence have to examine the various meanings of the word within its grammatical category.

Contextual and basic meanings are therefore contrasted as two alternative uses for the same word form *in the particular grammatical role that it has in the text*. As a result,

- a. the contextual meaning of nouns, verbs, adjectives, adverbs, prepositions, and interjections cannot be compared with the meaning of other word classes for the same lemma (conversions); for instance, the meaning of *shift* as a noun should be analyzed irrespective of the meaning of *shift* as a verb.
- b. the contextual meaning of verbs used as linking verbs, primary verbs, modal verbs, verbs initiating complex verb constructions such as *start*, *stop*, *continue*, *quit*, *keep*, and so on, causative verbs (*have*, *get*, and so on),

and full verbs cannot be compared with the meaning of the same verbs used in other roles.

- c. the contextual meaning of verbs used transitively can as a rule not be compared with the meaning of the same verbs used intransitively.
- d. the contextual meaning of nouns used to designate countable entities can as a rule not be compared with the meaning of the same nouns used to designate uncountable entities.

However, there are a number of complications:

2. When a word may be used in more than one grammatical category, but its description in the dictionary is limited to one of those categories only, you inevitably have to compare the various meanings of the word in the other grammatical categories with reference to that one grammatical category. Example: the contextual and basic meanings of *suppression* have to be examined with reference to the description of *suppress*.
3. When verbs are described under a single sense description in the dictionary as both Transitive and Intransitive, then you may compare these Transitive and Intransitive meanings with each other in order to determine whether the contextual meaning may be differentiated from a more basic meaning in the same sense description.
4. Sometimes lexical units have an abstract contextual meaning that is general which has to be contrasted with a concrete meaning that is specialized, for instance because it is limited to a style (e.g. very [in]formal), a subject (business, computing, journalism, law, linguistics, medicine, science, and so on), or period (literary, old-fashioned). In that case, we abide by our general rule for finding basic senses and take the most concrete sense as basic, even if it is specialized. Example: the concrete medical sense of *palliate* is basic and the general abstract sense of *palliate* is therefore metaphorical.
5. The reverse of [4] also applies: when a lexical unit with an abstract but specialized contextual meaning has to be contrasted with a concrete but general meaning, we also take the concrete sense as basic. Example: the abstract religious sense of *father*, *mother*, and so on is not basic, whereas the concrete general sense is. Therefore the religious senses are metaphorical.
6. When the contextual meaning of a lexical unit is just as abstract/concrete as some of its alternative meanings, we have to check whether there is any indication of the (original) domain from which the word derives. For instance, there are verbs such as *trot* and *roar* which may be applied with equal ease to a range of concrete entities, but the nonhuman, animal origin (basic sense) of the lexical units decides which applications are metaphorical and which are not.

7. However, other lexical units may have a less clear domain of origin, such as the verb *ride*. It is presented in the Macmillan dictionary as monosemous between animal and artefact. If we suspect that there is a problem with the dictionary description because of its function as an advanced learners' dictionary, we check the evidence in a second advanced learners' dictionary, Longman. For instance, the verb *to groom* does not have distinct senses for people and animals in Macmillan, but it does in Longman; as a result, we rely on Longman to conclude that the two senses are sufficiently distinct. By contrast, *transform* has one general sense in Macmillan, which is corroborated by the Longman dictionary.

2.3.3 Deciding about sufficient distinctness

Metaphorical meanings depend on a contrast between a contextual meaning and a more basic meaning. This suggests that the more basic meaning has to be sufficiently distinct from the contextual meaning for the latter to be seen as potentially participating in another semantic or conceptual domain. The following practical guideline should be followed: When a lexical unit has more than one separate, numbered sense description within its grammatical category, these senses are regarded as sufficiently distinct. When a lexical unit has only one numbered sense description within its grammatical category, this counts as the basic sense and any difference with the contextual sense of the item under investigation will count as sufficient distinctness.

2.3.4 Deciding about the role of similarity

When you have two sufficiently distinct meanings of a lexical unit and one seems more basic than the other, these senses are potentially metaphorically related to each other when they display some form of similarity. This typically happens because they capitalize on external or functional resemblances (attributes and relations) between the concepts they designate. It is immaterial whether these resemblances are highly schematic or fairly rich.

In deciding about a relation of similarity between the contextual and the basic sense of a lexical unit, the following practical guidelines should be followed:

1. When a lexical unit has a general and vague contextual sense which looks like a bleached, abstracted relation of a rather specific and concrete sense, you should mark the word as metaphorically used when the two senses are distinct enough and can be related via similarity. This is typically the case for senses that may be distinguished as concrete versus abstract.

It should be noted that similarity is not the same as class-inclusion, as in the case of synecdoche. Thus, for *appeal* we have an abstract general sense and a more concrete but also specialized legal sense. If we decide that the latter is basic because it is more concrete, then the general sense of *appeal* is a case of generalization instead of similarity, and it can therefore be treated as a case of synecdoche instead of metaphor. This should be contrasted with a case like *palliate*, where we see both generalization and similarity based on metaphorical mapping from concrete (relieve physical pain) to abstract (relieve generally bad situations of their most serious aspects).

2. When a lexical unit has an abstract contextual sense and a sufficiently distinct, concrete more basic sense, but there does not seem to be a relation of similarity between the two even though there does seem to be *some* sort of relation, check the Oxford English Dictionary to deepen your understanding of the word. In such a case, the two senses may be historically related via a common source which may have disappeared from the language. Checking the OED may explain the strange relation between the current abstract and concrete senses and support the decision *not* to take the concrete sense as basic for the abstract sense, but instead to take both senses as equally basic because there is no transparent relation of similarity for the contemporary language user. We have seen this for a word like *order* ('arrangement' and 'bringing about of order by speech act').
3. When two senses appear to be metonymically related, this does not mean that you should not also consider the possibility that they are metaphorically related at the same time. Sense relations may have more than one motivation.

2.4 Direct use potentially explained by cross-domain mapping

Directly used lexical units that are related to metaphor are identified as follows:

1. Find local referent and topic shifts.
 - Good clues are provided by lexis which is “incongruous” (Cameron 2003; Charteris-Black 2004) with the rest of the text.
2. Test whether the incongruous lexical units are to be integrated within the overall referential and/or topical framework by means of some form of comparison.
 - Good clues are provided by lexis which flags the need for some form of similarity or projection (Goatly 1997).
3. Test whether the comparison is nonliteral or cross-domain.
 - Cameron (2003: 74) suggests that we should include any comparison that is not obviously non-metaphorical, such as *the campsite was like a holiday village*. Consequently, whenever two concepts are compared and they can be

constructed, in context, as somehow belonging to two distinct and contrasted domains, the comparison should be seen as expressing a cross-domain mapping. Cameron refers to these as two incongruous domains.

4. Test whether the comparison can be seen as some form of indirect discourse about the local or main referent or topic of the text.
 - A provisional sketch of a mapping between the incongruous material functioning as source domain on the one hand and elements from the co-text functioning as target domain on the other should be possible.

If the findings of tests 2, 3, and 4 are positive, then a word should be marked for direct metaphor (MRW, direct).

2.5 Implicit meaning potentially explained by cross-domain mapping

The previous forms of metaphor were explicit in that there is at least one word in the discourse which comes from another semantic or conceptual domain. Implicit metaphor is different and does not have words that clearly stand out as coming from an alien domain. It comes in two forms, implicit metaphor by substitution and implicit metaphor by ellipsis. Following Halliday and Hasan (1976), metaphor by substitution works through pro-forms such as pronouns, and metaphor by ellipsis works through non-existent words which may be inserted into grammatical gaps. Both types therefore do not exhibit ostensibly incongruous words, but still need to be analyzed as the linguistic expression of metaphor in natural discourse.

When a discourse uses lexical units for the purpose of substitution and thereby still conveys a direct or indirect meaning that may be explained by some form of cross-domain mapping from a more basic meaning, referent, or topic, insert a code for implicit metaphor ('Implicit'). An example is: 'Naturally, to embark on such a step is not necessarily to succeed immediately in realizing *it*'. Here *step* is related to metaphor, and *it* is a substitution for the notion of 'step' and hence receives a code for implicit metaphor (MRW, impl).

When a text displays ellipsis and still conveys a direct or indirect meaning that may be explained by some form of cross-domain mapping from a more basic meaning or referent than the contextual meaning recoverable from the presumably understood lexical units, insert a code for implicit metaphor ('Implicit'). An example is *but he is*, which may be read as *but he is [an ignorant pig]*, when that expression is taken as a description of a male colleague discussed before. The verb *is* may be coded as a place filler by the code <MRW, impl>.

In general, for implicit metaphor, we need one linguistic element of cohesion (which means substitution or ellipsis, including what Halliday and Hasan call 'reference') that is not necessarily metaphorical by itself but refers back to a previous

word and concept that was metaphorically used. Potential elements of cohesion include third person pronouns, primary and modal verbs, and so on.

- The first step in finding implicit metaphor will therefore be to decide whether a particular linguistic form from a list of potentially cohesive devices has in fact been used for cohesion as opposed to another function.
- The second step is to decide whether the cohesion device is related to another word that was related to metaphor.

In principle it is possible for both demonstratives as well as general words such as *thing* and *stuff* to refer back to a metaphorically used expression. In that case, they are both indirectly metaphorical (because of their linguistic status) as well as implicitly metaphorical (because of their connection to a metaphorical concept in the text base). For this type of case we should add a code which combines ‘met’ with ‘impl’: metimpl.

Finally, tag questions within the same utterance are not included in our view of cohesion. They are grammatical forms enabling a particular form of asking a question. There is no alternative where the pro-forms in the tag could be replaced by full NPs or VPs. This is why these are not part of cohesion. (However, when parts of utterances are repeated by subsequent speakers in order to ask or confirm or deny what the preceding speaker said, these are core cases of cohesion.)

2.6 Signals of potential cross-domain mappings

Lexical signals of cross-domain mappings are those words which alert the language user to the fact that some form of contrast or comparison is at play (cf. Goatly 1997).

1. We focus on potential markers of simile and analogy and so on, such as *like*, *as*, *more*, *less*, *more/less ... than*, comparative inflection plus *than*, and so on. But we also include more substantial lexical markers such as *compare*, *comparison*, *comparative*; *same*, *similar*; *analogy*, *analogue*; and so on. Complex mental conception markers are also annotated as metaphor signals; they include *regard as*, *conceive of*, *see as*; *imagine*, *think*, *talk*, *behave as if* and so on; or simply *as if*. All of these lexical units are coded with ‘MFlag’.
2. We exclude more general signals of all indirectness, such as *sort of*, *kind of*, and so on, since it is not always clear that they signal metaphoricity or other aspects of discourse. We have also excluded what Goatly (1997) calls topic domain signalling, such as *intellectual stagnation*, since its nature and demarcation were not clear from the beginning of the project.

2.7 New formations and parts that may be potentially explained by cross-domain mapping

We assume that new-formations, such as *honey-hunting* discussed above, have to be analyzed as if they were phrases consisting of more than one lexical unit: each part of such new lexical units activates a concept and relates to a distinct referent in the discourse, which both have to be checked for metaphor. As a result, we sometimes have to mark parts of lexical units (free morphemes) as indicating metaphorical meaning.

The guidelines for finding metaphor-related words in new-formations are a variant on the basic procedure for finding all metaphor-related lexical units described in Section 2.1.

1. Find metaphor-related words in new-formations by going through the text on a word-by-word basis and identifying all new-formations.
 - A new-formation is a complex lexical unit consisting of at least one independent lexical unit which, as a whole, is not defined in the dictionary.
 - A special group is formed by specialized technical and scientific terms which may be missing from the regular dictionary but may therefore be seen as new-formations for the general language user.
2. When a lexical unit in a new-formation is used indirectly and its meaning in the discourse may be explained by some form of cross-domain mapping, mark the word as related to metaphor. (MRW, indirect)
 - If you're not sure about indirect word use that is explained by cross-domain mapping, go to Section 2.3.
3. When a lexical unit in a new-formation is used directly and its meaning may be explained by some form of cross-domain mapping, mark the word as direct metaphor. (MRW, direct)
 - If you're not sure about direct use of lexical units that is explained by cross-domain mapping, go to Section 2.4.
4. When a lexical unit in a new-formation implicitly conveys a direct or indirect meaning that may be explained by some form of cross-domain mapping, insert a code for implicit metaphor. ('Implicit')
 - If you're not sure about implicit indirect meaning that is explained by cross-domain mapping, go to Section 2.5.
5. When a lexical unit in a new-formation functions as a signal that a cross-domain mapping may be at play, mark it as a metaphor flag. (MFlag)
 - If you're not sure about signals of cross-domain mappings, go to Section 2.6.

Acknowledgements

NOTE: This chapter was originally published as Chapter 2 of Steen, G. J., Dorst, A. G., Herrmann, J. B., Kaal, A. A., Krennmayr, T., & Pasma, T. (2010). *A method for linguistic metaphor identification: From MIP to MIPVU*. Amsterdam: John Benjamins. It has been reprinted here with permission from John Benjamins and has only been revised in terms of cross-referencing, numbering of sections, and formatting in order to fit into the current volume.

References

- Cameron, L. (2003). *Metaphor in educational discourse*. London: Continuum.
- Charteris-Black, J. (2004). *Corpus approaches to critical metaphor analysis*. Basingstoke: Palgrave Macmillan. <https://doi.org/10.1057/9780230000612>
- Goatly, A. (1997). *The language of metaphors*. London: Routledge. <https://doi.org/10.4324/9780203210000>
- Halliday, M., & Hasan, R. (1976). *Cohesion in English*. London: Longman.
- Pragglejaz Group (2007). MIP: A method for identifying metaphorically used words in discourse. *Metaphor and Symbol* 22(1), 1–39. <https://doi.org/10.1080/10926480709336752>
- Steen, G. J., Dorst, A. G., Herrmann, J. B., Kaal, A. A., Krennmayr, T., & Pasma, T. (2010). *A method for linguistic metaphor identification: From MIP to MIPVU*. Amsterdam: John Benjamins. <https://doi.org/10.1075/celcr.14>