



Annotation guidelines for the GerMAN corpus (German metaphor annotation)

Markus Egg^a

^aHumboldt-Universität zu Berlin

We present the guidelines for metaphor annotation in German, which we developed as part of the creation of an annotated metaphor corpus. After a short overview over the definition and compilation of the corpus, we introduce the different steps of metaphor annotation. We start with the basic principles of metaphor identification, which follow the work of Gerard Steen and his group, and the specific challenges imposed for such efforts by German morphology. Then we describe the workflow for metaphor annotation, first the distinction of relevant metaphor types and second the identification of the linguistic context on which the metaphorical interpretation of a linguistic item in the corpus is based, which we call ‘background’. We conclude with an evaluation of the guidelines and offer two appendices on dead metaphors and the challenging issue of German pronominal adverbials.

Keywords: metaphor, register, corpus annotation, issues of German morphology

1 Introduction

Reliable annotation of metaphors is an important ingredient of current approaches to metaphor processing (for an overview, see [Rai & Chakraverty 2020](#) and [Tong et al. 2021](#)). Such high-quality annotation is based on guidelines that spell out in detail the procedures of identifying and classifying the relevant items. In the field of metaphor research, the Metaphor Identification Procedure-VU¹ (MIPVU for short) of [Steen et al. \(2010\)](#) has been adopted by many annotation initiatives in different languages ([Nacey et al. 2019](#)), including German ([Herrmann et al. 2019](#)), and we also base the guidelines for our own annotation work on this procedure.

The present guidelines were developed as part of the creation of a German metaphor corpus ([Egg & Kordoni 2022, 2023, Egg 2026](#)), which comprises six

¹The ‘VU’ parts stands for *Vrije Universiteit Amsterdam* (‘Free University’), where these guidelines were created.

subcorpora of different text types (parliament speeches, sermons, debates from debating competitions, fiction, newspaper commentaries, and TEDx talks) with 30,000 words each. During the specification of the guidelines, we encountered many of the challenges that [Herrmann et al. \(2019\)](#) describe in their paper on metaphor annotation for German on the basis of MIPVU, but encountered further issues that needed specification, for instance, the challenge of the so-called pronominal adverbials like *daran* or *worauf*, which consist of a preposition and a demonstrative or interrogative adverbial.

Starting with the definition and compilation of the corpus, the guidelines introduce the different steps of metaphor annotation, from the basic principles of metaphor identification and its challenges in German discourse to the concrete way in which different types of metaphors were identified and annotated in the corpus. They conclude with an evaluation of the guidelines and offer two appendices on dead metaphors and the challenging issue of German pronominal adverbials.

The intended audience of these guidelines includes in particular, but is not restricted to, researchers who want to use the GerMAN corpus for their own research on metaphor as well as research initiatives that want to annotate their own corpora for metaphor. To address the specific interests of the second group of readers, the structure of these guidelines follows the workflow of the annotation. The first group of readers might also be interested in the description of the corpus composition in [Section 2](#); they can furthermore concentrate on specific aspects of the annotation that interest them by choosing them from the overview at the beginning of [Section 3](#).

2 The corpus

The GerMAN corpus targets the relation of metaphor and register, integrating a wide range of register variation. Its six parts (of approximately 30,000 words each; see [Table 1](#)) are parliament speeches from the German Parlamentsreden-Korpus ([Blaette 2017](#)), news commentaries (including the Potsdam Commentary Corpus; [Stede 2004](#)), sermons, light fiction (written by amateurs for their peers), debates from competitions of the organisation ‘Jugend debattiert’ ([Kemmann 2013](#)), and TEDx talks from the Multilingual TEDx Corpus ([Salesky et al. 2021](#)).

[Table 2](#) shows how register properties are distributed in the corpus, starting with those of tenor (relations between interlocutors), viz. hierarchy vs. equality and distance vs. closeness. The mode dimensions are dialogue vs. monologue and spoken vs. written. Following [Koch & Oesterreicher’s](#) (1994) distinction of conceptual literality vs. orality, speeches and sermons are classified as literal (they

Table 1: Sample and word counts for the subcorpora

subcorpus	number of samples	word count
speeches	21	30,253
commentaries	176	29,641
sermons	19	32,780
light fiction	4	30,121
debates	8	32,569
TEDx talks	17	30,178
Total	245	185,542

are prepared and fixed in advance; the ‘written to be spoken’ category of the BNC), despite their oral presentation. From the viewpoint of production, four text types are transmitted orally (speeches, sermons, debates, and TEDx talks), while commentaries and fiction are transmitted in writing. Debates and TEDx talks are conceptually oral, the debates are dialogic and the other text types are monologic.

The subcorpora also vary along two important [Biber \(2009\)](#) dimensions, the one of ‘situation-dependent vs. elaborated reference’ (how dependent is non-anaphoric reference on the situational context), and the one of overt expression of persuasion. With the annotation, we wanted to investigate the question of how the distribution of metaphors of different types depends on such register properties. Still, we believe that the corpus and the annotation guidelines can also be fruitfully used for metaphor research in general, irrespective of this specific research question.

The annotation used the INCEpTION tool ([Klie et al. 2018](#)). Metaphors were annotated independently by two annotators, any differences were subsequently adjudicated. It includes a layer of syntactic structure, derived by parsing the majority of the texts with the Stanza package ([Qi et al. 2020](#)), to allow the identification of syntactic constellations for analyses of their metaphorical potential in future work.

Fig. 1 is a snapshot of the INCEpTION tool. The menu on the right-hand side of the screen provides the labels for the features used in the annotation, which can be selected and applied to a segment (a word or a part of a word), once this segment has been marked in the running text (marked segments without labels

Table 2: SFL register properties of the subcorpora

subcorpus	hierarchical / equal	distant / close	oral / literal	dialogue / monologue	situation dependence	persuasiveness
speeches	E	D	L	M	high	medium
sermons	H	C	L	M	medium	high
commentaries	H	D	L	M	high	high
light fiction	E	C	L	M	low	low
debates	E	D	O	D	high	high
TEDx talks	H	D	O	M	medium	medium

are indicated by ‘(MRW)’). Two segments can be linked by clicking on the origin and drawing the mouse to the goal and subsequently labelling the link (the link labels appear after clicking on a link in the annotation). For the full range of options for the tool (e.g. the specific search functions, which are located on the left-hand side of the screen) we refer to the INCEPTION website.²

The annotation focuses on a subgroup of metaphors which we consider pivotal for register in that they can function as explicit register markers, consciously used by writers/speakers to indicate compliance with the requirements of specific registers and by readers/hearers to identify register properties. In particular, we investigated signalled metaphor (accompanied by additional material like quotation marks or in an explicit comparison; the ‘direct metaphors’ of Steen et al. 2010), conventionalised and non-conventionalised metaphor, extended metaphor (or metaphor chains), and potential metaphor (deliberately using tokens of the same expression first with a literal and then with a metaphorical sense in the same discourse).

In our annotation, we not only identified the metaphorical expressions themselves but also the linguistic context on which the metaphorical interpretation is based, which we call ‘background’. For instance, (1) comprises a metaphorical preposition with an NP complement, here, *einer schwierigen Situation* ‘a difficult situation’, that functions as background. The literal meaning of the preposition *in* is spatial while the background refers to an abstract state of affairs.

²<https://inception-project.github.io>

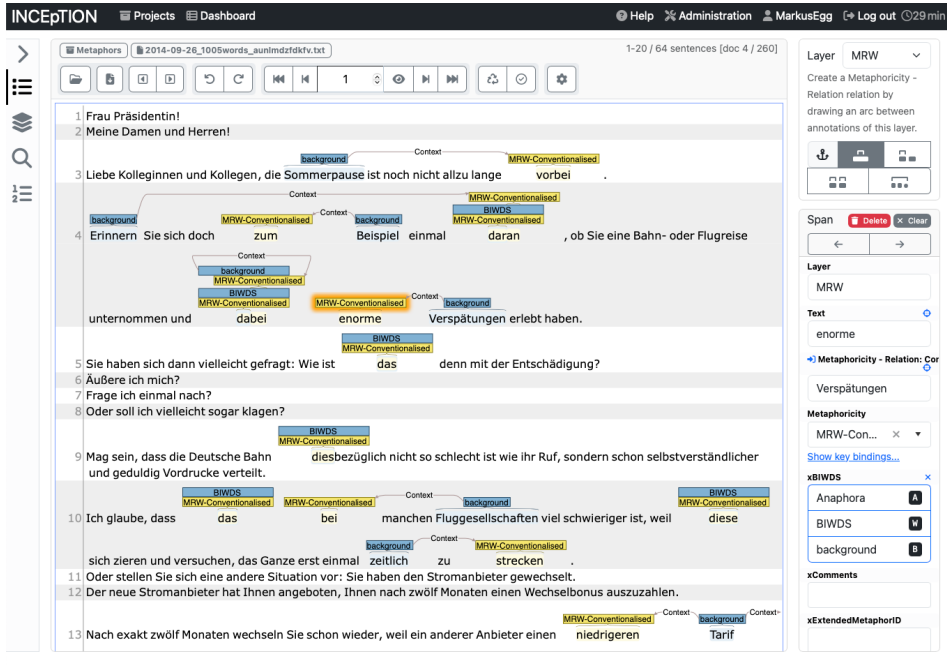


Figure 1: A snapshot of the annotation tool

- (1) *in einer schwierigen Situation*
 ‘in a difficult situation’

Metaphor types and backgrounds are described in detail in Sections 3.3-3.6 and Section 4. The set of features used in their annotation is summarised in Table 3.

The features *Non-MRW* and *WIDLII* are included for purposes of the annotation process; they do not show up in the finalised corpus. The feature *Relation-Type* shows up after creating or clicking on a link.

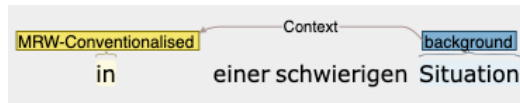


Figure 2: A first example of the annotation

As a first example of an annotation, consider the annotation of (1) in Fig. 2. The conventionally metaphorical preposition *in* bears the label *MRW-Conventionalised*, and *Situation*, the head of the corresponding background (the NP complement),

Table 3: Features employed in the annotation

feature	values	explanation
Metaphoricity	MRW-Conventionalised	conventionalised metaphor
	MRW-Non-Conventionalised	non-conventionalised metaphor
	MetaphorFlag	expressions that signal metaphor
	PotentialMetaphor	for literal expressions that later reappear as metaphors
	Non-MRW	for expressions whose status as non-metaphorical is already established
	WIDLII	for cases whose potentially metaphorical status requires further deliberation ('when in doubt, leave it in')
xBIWDS	Anaphora	anaphoric backgrounds
	BIWDS	the background is outside the clause ('background in the wider discourse')
	background	triggers a metaphoric interpretation of another expression
xComments	free	for comments during the annotation
xExtendedMetaphorID	number	for coindexing in the case of extended and potential metaphors
Relation-Type	Context	linking metaphors and backgrounds or metaphors and their flags
	Core	linking words that belong to the same metaphor or background

is labelled as *background*. Metaphor and background are connected by a *Context* link. The details of this and other annotations are expounded in Sections 3 and 4.

3 Annotating metaphor

In the following two sections, we describe the annotation of metaphors and then the annotation of the background, that part of the context that triggers the metaphorical interpretation. The description follows the workflow of the annotation:

- delimiting the range of text to be annotated (Section 3.1)
- identification of metaphors in general (Section 3.2), with reference to specific characteristics of German (Section 3.2.2)
- distinguishing conventionalised and non-conventionalised metaphors (Section 3.3)
- identifying and annotating extended metaphors (Section 3.5)
- identifying and annotating potential metaphors (Section 3.6)
- identifying and annotating backgrounds (Section 4)

3.1 Delimiting the text to be annotated

Before the annotation begins, it is necessary to define the range of the text to be annotated. In particular, the following elements of the texts have to be excluded:

- long quotes of two sentences and more (e. g. many of the sermons we annotated included the Bible passage on which the sermon was based)
- repetitions, which show up particularly in the debates

(2) *auf der Mehrheit auf der Mehrheit*
‘on the majority’

- meta-level remarks about the text (e. g. information on author or setting)

Still, it is advisable to annotate titles and quotes not longer than one sentence, however, because we found that these are integrated into the surrounding discourse.

3.2 Identifying metaphorical expressions

3.2.1 General rules

In a first step, metaphors are identified, following the rules of the MIPVU as described in chapter 2 of [Steen et al. \(2010\)](#).

An expression can be classified as a metaphor if its context-based sense differs from another, more ‘basic’ sense of the expression (e.g. one which is more concrete or related to bodily action), which is currently in use (or synchronically relevant). These senses must be related in terms of similarity but not subsumable under a common hypernym.

Basic senses are defined as synchronically relevant when they appear in a suitable lexical resource. We follow [Herrmann et al. \(2019\)](#) in using the lexical databases *Duden* (www.duden.de) and *Digitales Wörterbuch der deutschen Sprache* (www.dwds.de).

For instance, consider (3), in which *rasen* refers to time passing very quickly.

- (3) *Die Zeit rast dahin.*
‘Time is rushing by.’

The DWDS entry for *rasen* defines it as very fast spatial movement,³ which e.g. the contextual temporal sense of *rasen* ‘rush’ in (3) shares with its spatial, basic sense the element of speed.

When identifying metaphorical units, it is crucial to bear in mind what we call the subcategorisation criterion (inherent in the MIPVU rules, [Steen et al. 2010](#): page 35f.): A contextual reading of a unit can only be classified as metaphorical on the basis of a more basic reading of the same unit if both readings share the same subcategorisation frame. For instance, this applies to the (intransitive or transitive) ‘understand’ reading of *sehen*, since its basic reading ‘receive visual sensory input’ likewise appears as transitive or intransitive. In contrast, e.g. *gehen um* ‘be about’ is not metaphorical because the *um*-PP and the expletive pronoun *es* are obligatory complements in this reading, whereas for the basic spatial locomotion interpretation of *gehen* ‘walk’ a wide range of subjects and of (optional) PP-complements are available.

Another case in point is the metaphorical use of *Ziel* ‘target’ in the sense of ‘intention’, when it is used with a VP complement. These tokens of *Ziel* are not marked as metaphorical, because *Ziel* cannot have such a complement in the basic ‘target’ interpretation.

³The exact definition is *sich mit großer Geschwindigkeit bewegen, fortbewegen, sehr schnell fahren, laufen* ‘move, move away with high speed, drive, walk very quickly’.

- (4) *das Ziel, die von mir zu Kontrollierenden abzuschaffen*
 ‘the intention to abolish those who are to be controlled by me’

The issue here, pointed out by [Herrmann et al. \(2019\)](#), is that the subcategorisation frames are not listed exhaustively in our lexical resources. They note two pieces of potentially helpful grammatical information at this point, viz. differences in perfect formation for verbs and plural formation for nouns.⁴ In addition, annotators are advised to consult the examples offered in the lexical entries, e. g. in the case of *Ziel* both resources exhibit examples with a VP complement only for the ‘intention’ but not for the ‘target’ sense of the lexeme. Since this does not guarantee an exhaustive listing of all possible subcategorisation frames, there remains the possibility that the decision to apply the subcategorisation criterion is solely based on the individual annotator’s linguistic competence. In the interest of guaranteeing a high inter-rater reliability, these cases should be explicated in the comments and subsequently adjudicated collectively.

Finally, we differ from the MIPVU procedure in that we do not annotate what they have termed ‘implicit metaphors’. In these cases, pronouns and other pro-forms stand for a full expression that would be classified as a metaphor in the case of replacing the pro-form. E. g. in the (constructed) example (5), the pronoun *they* would count as an implicit metaphor, because replacing it by *the foundations* would entail a metaphorical interpretation of this NP because of *just*:

- (5) *The proposal has excellent foundations, but they are also just.*

Next, we describe the application of MIPVU to German in Section 3.2.2, and from Section 3.3 on, we characterise the different metaphor types to be annotated and the way and the order in which they are taken into account in our annotation.

3.2.2 Application to German

Since the MIPVU procedure was developed for English, its application to other languages requires specific adaptations. We describe the adaptations necessary in our corpus, which includes the ones already described in [Herrmann et al. \(2019\)](#).

3.2.2.1 Below the word level

While most relevant units for the annotation of metaphors are words, we deviate from Section 2.2 of the MIPVU chapter in recognising specific parts of words as

⁴For instance, *schmelzen* forms the perfect with *haben* in its transitive and with *sein* in its intransitive use (only the latter is used metaphorically) and *Sonnenschein* has no plural in its literal meteorological sense as opposed to its metaphorical meaning ‘darling’.

relevant units, in particular, for compounds, as illustrated by (6). In (6), *Haupt* ‘head’ in the abstract sense of ‘more important’ is metaphorical by itself, which is annotated appropriately as in Fig 3.

- (6) *Hauptamt*
 ‘more important office’ (lit. ‘head office’)

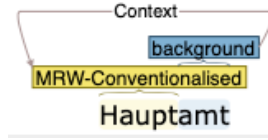


Figure 3: Annotation of a part of a word

A second group of words where parts can be relevant units are lexicalised demonstrative-preposition combinations, the so-called pronominal adverbials like *davon*, literally, ‘of it’.

- (7) *Davon müssen wir erzählen.*
 ‘We must talk about it.’

Here, both parts are metaphorical independently and marked appropriately in Fig. 4: the demonstrative *da* ‘there’ is used in a non-spatial sense, and the preposition *von* ‘of’ is used non-spatially to mark the direct object of *erzählen* ‘tell’.

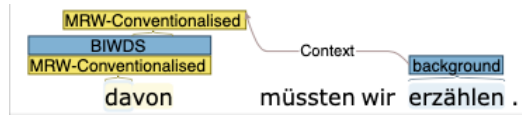


Figure 4: Annotation of pronominal adverbials

These adverbials are a challenge for metaphor analyses of German. Appendix B offers a detailed motivation for their representation in the annotation and a delimitation against similar cases that are outside the interests of metaphor annotation.

3.2.2.2 Above the word level

At the same time, and again in opposition to Section 2.2 of the MIPVU chapter, we recognise as relevant units specific expressions that consist of several words, which we call multi-word units (MWU).⁵

⁵These units are annotated only if they are relevant for metaphor.

This holds in particular for separable prefix verbs, which strand their prefixes in main clauses, like in the case of *einsetzen* ‘make use’ in (8):

- (8) *Dieses Mittel setzen wir ein.*
 ‘We make use of this measure.’

In every MWU, we distinguish a core element and link the other, peripheral elements to it by the relation *Core*. The core element is labelled for the function of the MWU, the peripheral element is not. In Fig. 5, for example, the prefix *ein* and its verb stem *setzen* constitute a conventional metaphor, whose core is the verb stem.

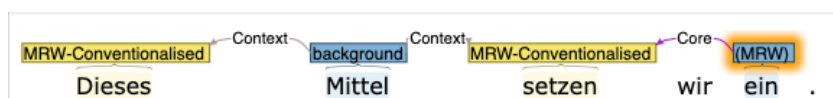


Figure 5: Linking the parts of a separable prefix verb

In the same way, some proper names consist of several parts, e. g. *Dr. Müller*, *Angela Merkel*, *Europäische Union* ‘European Union’, or *Neu Isenburg*. Their core elements are last names, or, if we can identify an internal modification structure, like in *Europäische Union*, the head.

Third, words that could or should have been written as a unit were separated occasionally, e. g. verbs like *schwermachen* ‘make difficult’, *leichtmachen* ‘facilitate’, *schwerfallen* ‘have a hard time’, or *leichtfallen* ‘find easy’.⁶ If the non-verbal part is separated from the verbal part, we identify the latter as the core of the unit.

Finally, metaphorically relevant MWUs can consist of several words that form a multi-word expression (MWE), which does not necessarily constitute a syntactic constituent, e. g. in (9) (where the complement of the transitive verb *bringen* ‘bring’ is missing).

- (9) *an einen Tisch zu bringen*
 ‘summon for negotiations’ (lit. ‘to bring to one table’)

The annotation of such multi-word expressions proceeds in two steps. First, we identify the core element of the unit. If the unit forms a syntactic constituent of its own, its head is the core element like in the case of the VP in (10), whose head *zurückspielen* ‘return’ emerges as the core element of the unit:

⁶Such cases were often triggered by orthographical vacillations in the wake of the steps toward an orthography reform for German in 1996.

- (10) *den Ball zurückspielen*
 ‘react in a way that compels a reaction from the other involved party’ (lit.
 ‘return the ball’)

In all other cases, we add the syntactically missing elements in the expression and then identify the head of the resulting constituent as the core element, e. g. for (9), adding the missing direct object of *bringen* ‘bring’ returns a VP, which allows identifying the verb as the core element:

- (11) *Schüler, Eltern und Lehrer an einen Tisch zu bringen*
 ‘to summon pupils, parents, and teachers for negotiations’

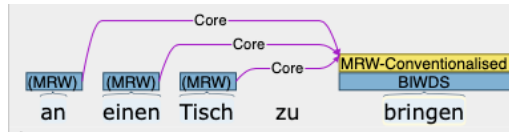


Figure 6: Linking the parts of a multi-word unit

In the second step, all non-core parts of the item are linked to the core element with a *Core* relation, but are not labelled themselves, as illustrated in Fig. 6. In this way, non-continuous MWEs can be annotated as well. (We are aware of the fact that this is a rather preliminary treatment of metaphor in MWEs.)

3.2.2.3 Conversion of verbal forms

In German, the non-finite verbal forms can undergo conversion, participles into adjectives, e. g. in *spielende Kinder* ‘playing children’ and infinitives into nouns, like in *beim Singen zuhören* ‘listen to someone sing’, respectively.

For infinitives, this form is usually not listed in the lexical resources, e. g. in the case of Herrmann et al.’s (2019) example *Brechen* ‘breaking’, which appears in our corpus as well, e. g. in the compound *Fastenbrechen* ‘break fasting’. Since this type of conversion is productive and semantically transparent,⁷ we follow their observation that the decision of whether such nominalised infinitives are metaphorical or not must fall back on the corresponding verb entry. For *Fastenbrechen*, this strategy reveals that the head *Brechen* is metaphorical in the sense

⁷It must be distinguished from non-productive cases of nominalisation, which often involve a stem vowel change, e. g. in *Bruch* ‘break’. These have lemmas of their own and should be considered independently of the underlying verb.

of ‘terminate’, while the literal interpretation is ‘divide physically by applying pressure’.

In contrast, many participles have developed into independent lexemes with lemmas of their own, e. g. *umfassend* ‘comprehensive’ or *entscheidend* ‘decisive’. Consequently, a potentially metaphorical expression that could be a participle or an adjective must be disambiguated before considering its status as a metaphor. For instance, *umfassend* and *entscheidend* could be adjectives, but *schwelend* ‘smouldering’ could not.

If the expression is identified as an adjective, the corresponding lemma is to be consulted to decide the question of whether it is metaphorical or not (irrespective of the meaning of the corresponding verb). For instance, *drückend* in the meaning ‘sweltering’ is not metaphorical, in spite of the basic meaning of *drücken* in the sense of ‘apply physical pressure’. Only if such an expression is clearly a participle (e. g. for *drückende Last* ‘heavy load’) can the decision be based on the lemma of the corresponding verb.

But if there is no specific lemma for a participle-based adjective in the lexical resources, like in the case of *schwelend* ‘smouldering’ then the question of whether its sense ‘hidden yet ongoing’ is metaphorical must be based on the lemma of the verb *schwelten* ‘smoulder’.

3.3 Conventionalised and non-conventionalised metaphor

All metaphors identified in the first step must be marked as a metaphor-related word (MRW). The next step consists in annotating them as either a conventionalised or as a non-conventionalised MRW.

To ensure a satisfactory inter-rater reliability at this point, the distinction of these two categories is based on the lexical resources: If both the basic and the metaphorical senses appear in at least one of the dictionaries, the metaphor counts as conventionalised and is labelled with the label *MRW-Conventionalised*. If only the basic sense appears in the dictionaries, the metaphor is classified as not conventionalised and labelled as *MRW-Non-Conventionalised*, like the noun in (12):

- (12) *ein sozialer Sprengsatz*
‘a social explosive device’

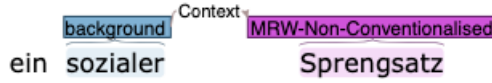


Figure 7: Annotating non-conventionalised metaphor

Once all the metaphors have been labelled as conventional or non-conventional and linked to their backgrounds, we identify signalled metaphors (Section 3.4), extended metaphors (Section 3.5), and potential metaphors (Section 3.6).

3.4 Signalled metaphors

Signalled metaphors (conventionalised or non-conventionalised) are accompanied by additional material, so-called ‘metaphor flags’, among them hedges such as *sozusagen* ‘in a way’ or expressions that introduce explicit comparisons like *wie* and *als* ‘like’ (which gives rise to similes). These flags highlight the metaphorical status of the expressions they accompany, which is why we talk about ‘signalled metaphor’ in these cases. Usually, there is only one flag per metaphor.

- (13) *die Urgemeinde als eine Familie*
‘the early community as a family’
- (14) *Jean Claude Juncker, der Oberfinanzminister, Chef sozusagen der EU*
‘JCJ, the senior minister of finances, boss of the EU, in a way’

Goatly (2011) offers an extensive list and a classification of these flags. We follow Steen et al. (2010) in annotating a subset of these expressions, viz. expressions that introduce similes such as *als* ‘like’, hedges, expressions referring to a perceptual process like *wirken* ‘come across’ in (15), intensifiers like *regelrecht* ‘downright’ in (16), and quotation marks like in (17).

- (15) *Das Ganze wirkt wie eine letzte große kolonialistische Geste.*
‘The whole [procedure] comes across as a last grand colonial gesture.’
- (16) *regelrechte Nervendaten-Autobahnen*
‘downright motorways for neural data’
- (17) *alle „niederen“ Bedürfnisse*
‘all “basic” (lit. low) needs’

We annotate signalled metaphor by labelling the flag *MetaphorFlag* and linking it to the metaphorical expression by a *Context* relation, like illustrated in Fig. 8 for (14). For quotation marks, only the first one is annotated.

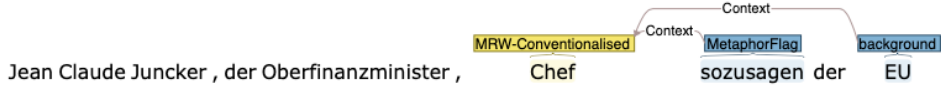


Figure 8: Annotating signalled metaphor

3.5 Extended metaphors

Extended metaphors consist of several individual metaphors which all share the same metaphorical mapping. We distinguish it from mere repetitions of the same metaphor and only annotate cases in which the mapping is instantiated by at least two different metaphors. For instance, an extended metaphor in one of the debates describes social rank in terms of height, starting with *oben* ‘above’, then the mapping reappears in *herabsehen* ‘look down upon’ and three repetitions of *auf Augenhöhe* ‘at eye level’:

- (18) *bevor wir wieder von oben auf sie herabsehen*
 ‘before we look down upon them from above’
- (19) *endlich auf Augenhöhe miteinander reden*
 ‘talk on a level playing field’ (lit. ‘at eye level’)

Extended metaphors may span several sentences, like in (18) and (19), but can also appear sentence-internally, as in (20), where *bauen* ‘build’, *fest* ‘firm’, and *Fundament* ‘foundation’ all exhibit the same metaphorical mapping of physical onto organisational structure (Grady 1997):

- (20) *ein festes Fundament für eine glückliche Zukunft zu bauen*
 ‘to build a firm foundation for a happy future’

We indicate the individual metaphors that make up an extended metaphor (regardless of whether they are conventional or not) by assigning each one of them the same index (in terms of a specific number in the field ‘xExtended-MetaphorID’). Such extended metaphors are characteristic for sermons and debates, like in the example quoted above, but can be found outside these text types, too.

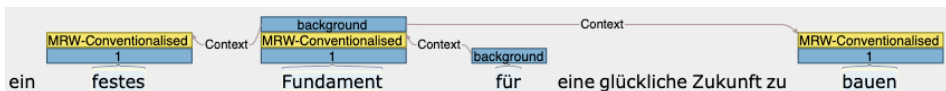


Figure 9: Annotating extended metaphor

3.6 Potential metaphors

Potential metaphors emerge when expressions are first used in a literal sense, and then occur in a metaphorical interpretation in the same discourse at some point. In such a constellation, all the tokens of the expression with a literal sense are called potential metaphor.

Consider for instance the *Lumpen* ‘rags’ metaphor in one of our sermons. The sermon first refers to people who literally wear rags occasionally in (21), and then reuses the expression with a metaphorical interpretation in (22):

(21) *in alten Lumpen*
‘in old rags’

(22) *in den Lumpen deiner zerschissenen Gerechtigkeit*
‘in the rags of your worn out justice’

We first annotate the potential metaphors (the literal tokens) with the label *PotentialMetaphor* and indicate the relation between the literal and the metaphorical tokens of the expression in terms of a shared index in the ‘xExtendedMetaphorID’ field. Potential metaphors are typical for sermons.⁸

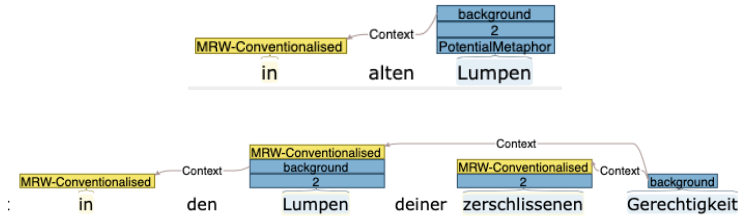


Figure 10: Annotating potential metaphor

4 Metaphors and their background

4.1 Kinds of backgrounds and how to find them

After annotating the metaphors, the next step is linking them to their background, the immediate linguistic context on which the metaphorical interpretation

⁸Fig.10 illustrates the observation that potential metaphors can partake in extended metaphors. In (22), the metaphors *Lumpen* ‘rags’ and *zerschissenen* ‘worn out’ share a metaphorical mapping from clothes to character traits.

tation is based. Consider example (1) and its annotation in Fig. 2, repeated here for convenience:

- (23) *in einer schwierigen Situation*
 ‘in a difficult situation’

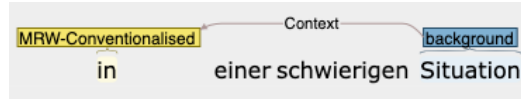


Figure 11: Annotating background

Backgrounds introduce the target domain of the respective metaphor, in (23), for the mapping STATES ARE LOCATIONS. The basic strategy for linking metaphors to their backgrounds in the annotation uses *Context* links from the background (or, in the case of complex backgrounds, its core element, see below) to the metaphor. For (23), the link in Fig. 11 connects the head of the background (the complement NP) with the metaphorical preposition *in*.

Two principles guide the identification of backgrounds during the annotation. First, metaphors and backgrounds are syntactically or morphologically linked. The relations between metaphors and their backgrounds that are relevant for our corpus are defined exhaustively in Section 4.2, e.g. as a head-complement constellation like in (23). We expect them to be relevant for other corpora as well, which does not preclude the possibility that these corpora might exhibit additional metaphorically relevant syntactic or morphological relations.

Second, we typically find a semantic incongruity between the literal interpretation of the metaphor and the background. For instance, the spatial meaning of the preposition *in* in (23) is incompatible with the abstract sense of its complement *einer schwierigen Situation* ‘a difficult situation’, and the concrete sense of *Sprengsatz* ‘explosive device’ is not compatible with the abstract meaning of *sozialer* ‘social’ in (12).

Both metaphorical expressions and backgrounds can consist of several words. If the background is composed of a group of several words, the *Context* link starts at the core element of this group. Such a group can be a phrase, e.g. for the NP background in (23), the link goes from its head *Situation* to *in*. The group can also be one of the multi-word units described in Section 3.2.2, e.g. for the separable prefix verb *sprechen ... aus* ‘opt’ in (24), which is the background for the metaphorical preposition *gegen* ‘against’. Here the *Context* link starts at the stem *sprechen*, the core element of the background:

- (24) *70% der Deutschen sprechen sich gegen gentechnisch veränderte Nahrungsmittel aus.*

‘70% of the Germans opt against genetically altered foods.’

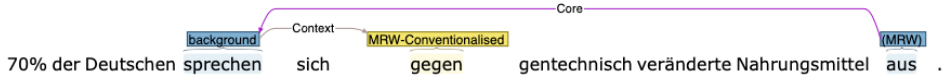


Figure 12: Annotating multi-word units as backgrounds

The metaphorical expression can consist of several words, too, for instance, for the separable prefix verb *setzen ... ein* ‘make use’ in (8). Here the *Context* link from the corresponding background *dieses Mittel* ‘this measure’ terminates at the core element *setzen*, as shown in Fig. 5.

Finally, both background and/or metaphor can consist of items below the word level. In the compound *Hauptamt* ‘more important office’ in Fig. 3, the head *Amt* ‘office’ functions as the background for the metaphorical non-head *Haupt* ‘head’.

Apart from backgrounds that consist of full (non-anaphoric) constituents, we also annotate backgrounds that are introduced by anaphors, which includes relative pronouns. These backgrounds are labelled as *Anaphora*. For instance, the pronoun *sie* (literally, ‘she’, since German has grammatical gender), refers back to the preceding NP *die Energiepolitik* in (25), wherefore it serves as background for the metaphorical *angesprochen* ‘mentioned’ (literally, ‘addressed’):

- (25) *die Energiepolitik – sie ist schon angesprochen worden –*
‘the energy policy – it has already been addressed –’

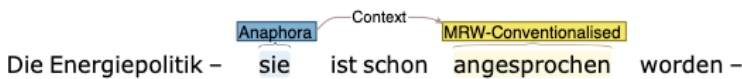


Figure 13: Annotating an anaphoric background

An anaphor can serve as background if replacing it by its antecedent would amount to introducing a full, non-anaphoric background. For example, in (25), the personal pronoun *sie* serves as background for *angesprochen* ‘mentioned’, because replacing it by its antecedent *die Energiepolitik* would return a non-anaphoric background. We do not identify the antecedents of the involved anaphors, however.⁹

⁹In the case of relative pronouns, the modified element counts as the antecedent.

- (26) *die Energiepolitik ist schon angesprochen worden*
‘the energy policy has already been addressed’

Backgrounds are only annotated when metaphors and their backgrounds are in the same clause. If a metaphor has no background in its clause, the metaphorical expression bears the additional label *BIWDS* (‘background in the wider discourse’), for instance, in Fig. 4 or 6. Labelling a metaphorical multi-word unit by *BIWDS* affects the core element only, no such label appears on the non-core elements, as illustrated, e. g. for *an einen Tisch bringen* ‘summon’ in (9) in Fig. 6.

4.2 Syntactic environments for metaphor-background pairs

4.2.1 Copulative constructions

In copulative constructions, the *Context* relation holds between the two arguments (subject and predicative complement) of the respective verbs; we distinguish three kinds of copulative constructions:

- purely copulative relations with verbs like *sein* ‘to be’, *repräsentieren* ‘represent’ or *bilden* ‘form’

- (27) *die Entscheidung (...) ist noch offen*
‘the decision is still pending’

- (28) *die europäischen Grundwerte bilden unsere Basis*
‘the European fundamental values form our foundation’

- verbs like *bleiben* and *werden*, which express the continuation or the beginning of a copulative relation, respectively

- (29) *an vielen Stellen bleibt er zu weich*
‘in many places, it [the annual report of the EU on asylum politics] remains too soft’

- (30) *den Heiland, der schwach wurde*
‘the Saviour, who became weak’

- verbs like *machen* ‘make’, when they express the causation of such a copulative relation

- (31) *die Auferstehungswirklichkeit sichtbar machen*
 ‘to make visible the reality of the Resurrection’

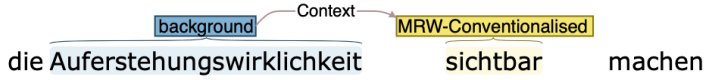


Figure 14: Sentential background of juxtaposed elements

4.2.2 Head-Complement

Head-complement constructions exhibit a considerable amount of metaphoricity, especially for prepositions and verbs (for verbs we subsume subject-verb pairs here). The metaphoricity can go in both directions, with the complement serving as background for a metaphorical head like in (20) or vice versa (*Liebe* ‘love’ as the background for *zu* ‘to’) in (32):

- (32) *aus Liebe zu dir*
 ‘out of love for you’

Several complements can simultaneously be backgrounds of their metaphorical head, like in (33), where the verb *stehen* ‘stand’ gets a non-spatial interpretation through its arguments *Europa* and *an vorderster Stelle* ‘at the forefront’:

- (33) *Europa muss an vorderster Stelle stehen*
 ‘Europe must stand at the forefront’

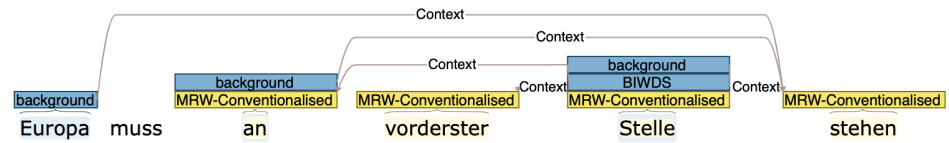


Figure 15: A metaphorical head with multiple background complements

Heads (mostly, verbs, but also nouns) can subcategorise for a PP complement and determine the complement’s head, like in *ringen um* ‘fight (lit. wrestle) for’. In such cases, the preposition is often metaphorical, with the head providing its background. This holds in particular for the prepositions *durch* ‘through’ and *von* ‘from’ in their role as indicating the subject for passive verbs and (in the case of *von*) the object of nominalised verbs and other event-denoting nouns, e. g. in (34):

- (34) *Export von Rüstungsgütern*
 ‘export of military equipment’

Some metaphorically used prepositions seem to have two potential backgrounds: the element that subcategorises for them (mostly, a verb) or their argument. Consider e. g. (35) and (36), where the metaphorical status of the prepositions *an* ‘at’ and *aus* ‘out of’ could be due to the fact that they are subcategorised for by the respective verbs *denken* ‘think’ and *herauskommen* ‘exit’. Alternatively, their metaphorical status might be triggered by the non-spatial complements of the PPs (*den Lobbyismus* ‘lobbyism’ and *extremer Armut* ‘extreme poverty’, respectively). For these constellations, we distinguish two cases.

- (35) *an den Lobbyismus denken*
 ‘think of lobbyism’
- (36) *mehr als 100 Millionen Menschen aus extremer Armut herausgekommen sind*
 ‘more than 100 million people left behind extreme poverty’

- First case: The status of the subcategorising verb as metaphorical or not does not depend on the complement of the PP, like in the case of *denken* in (35), which refers to a mental activity, irrespective of the nature of the topic of this activity (as specified in the complement of *an*). In this case, the verb forms the background for the metaphorical preposition, which is annotated like in Fig. 16 for (35).

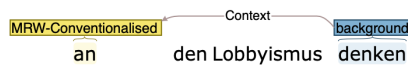


Figure 16: PP-subcategorising verb as background

- Second case: The status of the verb as metaphorical or not depends on the complement of the PP, which can be shown by changing it like in (37), where the verb *herauskommen* ‘exit’ is not metaphorical, as opposed to (36):

- (37) *aus dem Haus herauskommen*
 ‘leave the house’

Hence, if the verb is metaphorical due to the choice of complement of the PP like in (36), we analyse this complement as the background of the metaphorical preposition. The preposition is the head of the complement PP that forms the background of its metaphorical verbal head. In (36), the PP complement is *extremer Armut* ‘extreme poverty’, the complement PP is *aus extremer Armut* ‘out of extreme poverty’, and the verb is *herausgekommen* ‘exited’:

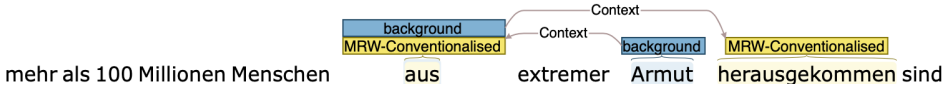


Figure 17: Subcategorised PP as background

Sentential complements as a specific case of a Head-Complement constellation are detailed in Section 4.2.7.

4.2.3 Modifier-Head and Specifier-Head

Mostly, modifiers form backgrounds of modified metaphors, e. g. in example (12) and its annotation in Fig. 7. The opposite direction is illustrated in (38), where the head *Jahren* ‘years’ functions as the background of its modifier *nächsten* ‘next’.

- (38) *in den nächsten Jahren*
 ‘in the next years’

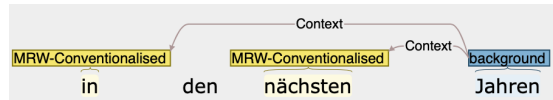


Figure 18: Modifier-Head and Specifier-Head relations

The relation between specifier and head is predominantly relevant for non-spatial uses of the demonstrative determiners, e. g. for discourse deixis like in (8) and its annotation in Fig. 5, where the head *Mittel* ‘measure’ functions as the background of the metaphorical determiner *dieses* ‘this’.

4.2.4 Apposition

Appositive constructions can also serve as syntactic contexts. In (39), the metaphorical *Unteren* ‘low classes’ has the second part of the apposition as its background:

- (39) *die Unteren, sozusagen die Wenigverdiener*
 ‘the low classes, those that are not well paid’

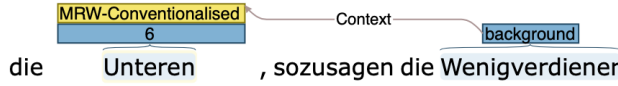


Figure 19: Apposition as background

4.2.5 Subject-and object-control adjectives

We found only one example in which *gebrochen* ‘crushed’ is metaphorical and its background is *ich*, because the adjective is under the control of the subject:

- (40) *ich sehe sie gebrochen an*
 ‘crushed, I look at her’



Figure 20: Controlling subject as background

4.2.6 Compound head and non-head

One part of a compound can be the background of the other, like in *Wertefundament* ‘foundation of values’, where *Wert* ‘value’ is the background and *Fundament* ‘foundation’, the metaphor. The head as well as the non-head can be background or metaphor, e. g. in *Fachpolitiker* ‘specialised politician’¹⁰ or *migrationsfluesse* ‘migration currents’. Sometimes the two parts are connected by a linking element (*Fugenelement*), like the *s* between *Migration* and *Flüsse*, which should not be included in the annotation.

These cases presuppose semantic transparency of the compounds. Semantically opaque compounds like *Stichpunkt* ‘bullet point’ or *Leidwesen* ‘chagrin’ can no longer be interpreted in terms of the meanings of their parts, hence, their internal structure is never annotated. For the question of where to draw the line

¹⁰The metaphorical expression *Fach* is used to refer to an abstract area, its basic interpretation is ‘compartment’ or ‘shelf’.

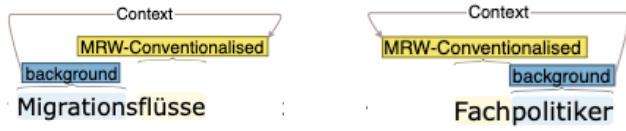


Figure 21: Compound head and non-head as background for the other part

with respect to semantic opacity, we fall back upon our lexical resources. For instance, they regard *Höhepunkt* ‘top, apex’ as opaque, even though its meaning still is related to the meaning of its two parts *Höhe* ‘height’ and *Punkt* ‘point’. In contrast, *Tiefpunkt* ‘bottom, nadir’ is regarded as transparent and has a literal, spatial interpretation.

Compounds can be metaphorical in their entirety, irrespective of their internal structure, first, semantically opaque compounds like *Schwerpunkt* ‘centre of gravity’ in the sense of ‘focus’, or *Schieflage* ‘tilt’ in the sense of ‘tricky situation’. Compounds with a transparent internal background-metaphor structure can be used metaphorically in their entirety, too, e. g. in (41). In this example, *Grundeinstellung* ‘standard setting’ breaks down into the metaphorical *Grund* ‘bottom’ (in the sense of ‘basic’) and its background *Einstellung* ‘setting’. Accompanied by the background *politische* ‘political’, the entire compound is used metaphorically to refer to a state of mind. Another example is *Angriffsfläche* in the sense of ‘weak point’ (literally, ‘effective area’, composed of the metaphorical *Angriff* ‘attack’ and its background *Fläche* ‘surface’). The annotation for (41) is shown in Fig. 22.

- (41) *die politische Grundeinstellung*
‘the fundamental political position’

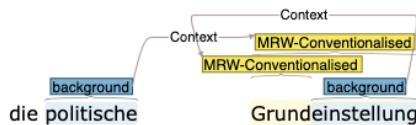


Figure 22: Metaphorical interpretation of a metaphorically complex compound

In (fortunately) extremely rare cases, this recursive metaphor-background structure can apply within one single word, e. g. in *Arbeitsplatzsuche* ‘job search’, where *Arbeit* ‘work’ is the background of *Platz* ‘place’, and *Arbeitsplatz* ‘work-place’ is the background for *Suche* ‘search’; the same kind of word-internal re-

cursion is found in *darüberhinaus* ‘what is more’. The resulting annotation is depicted in Fig. 23.

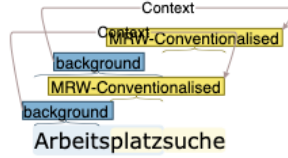


Figure 23: Recursive metaphor-background structure inside a complex compound

4.2.7 Sentential complements

The basic rule that for complex syntactic constituents as backgrounds, the *Context* relation starts from the head of the constituent applies to sentential constituents as well. For these constituents, we assume the following heads for this part of the annotation:

- for sentential backgrounds with complementisers (conjunctions or w-elements), these are the heads, e. g. *dass* ‘that’ and *was* ‘what’ carry the label *background* and are linked by a *Context* relation to *sehen* in (42) and (43):

(42) *Wir sehen, dass es neben Veränderungen im wirtschaftlichen Bereich auch Unsicherheiten gibt.*

‘We see that there is - apart from changes in the economic sector - also insecurity.’

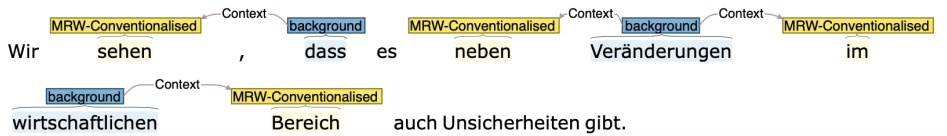


Figure 24: Sentential background with a complementiser

(43) *In diesen Tagen sehen wir ja, was in Afrika los ist.*

‘Nowadays we can see what is going on in Africa.’

- quotes of all syntactic categories can occur as syntactic constituents, e. g. in (44)-(46); if the quote is a finite sentence like in (44), the finite verb is the head:

(44) Vom „Wir schaffen das“ müssen wir jetzt übergehen zum „Wir machen das“.

‘We now have to proceed from “We can do it” to “We will do it”.’

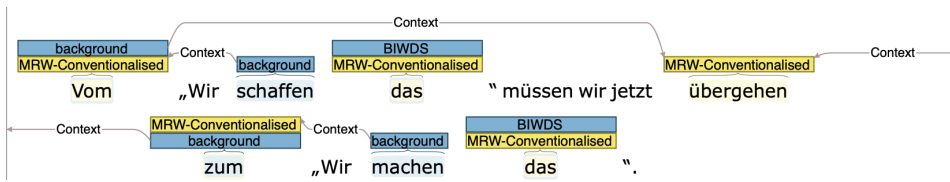


Figure 25: Sentential quote as background

In constituents that consist of two juxtaposed elements, the determination of the head is not obvious; we opt for the second element (or its head, if it is complex itself). Concrete examples for this phenomenon are (45) and (46), in which „Ja, aber“ ‘yes, but’ and „Je besser, desto mehr“ ‘The better, the more’ are backgrounds. The *Context* link from these backgrounds to their respective metaphors *kommt* ‘comes’ and *Formel* ‘formula’ starts at *aber* and *mehr*:

(45) da kommt nie ein „Ja, aber“
 ‘the response is never a “yes, but”’

(46) Die Formel dafür, an der wir uns orientieren können, ist: „Je besser, desto mehr“.
 ‘The formula for that, which might provide orientation for us, is: “The better, the more”.’

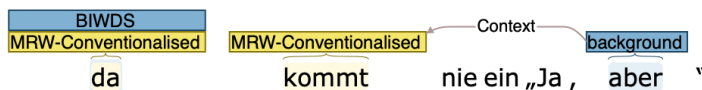


Figure 26: Sentential background of juxtaposed elements

4.2.8 Conjoined phrases as backgrounds

Conjoined phrases are of the type ‘A conj B’, where A and B are of the same syntactic type, and ‘conj’ is taken from a very small set of expressions, in particular, *und* ‘and’ and *oder* ‘or’, but also including *beziehungsweise/bzw.* ‘or rather’:

- (47) *angesichts des zunehmenden Missbrauchs und leerer Kassen*
 ‘in view of the increasing abuse and empty coffers’

Backgrounds can consist of conjoined phrases. Assuming that the conjunction is the head of a conjoined phrase, we follow the strategy of indicating the background status of a syntactically complex phrase by annotating its head with the label *background* and choosing it as the starting point of the *Context* relation to its respective metaphor. For instance, in (47), the metaphorical *angesichts* ‘in view of’ has as its background its whole complement, which consists of two NPs conjoined by *und* ‘and’. Consequently, the *Context* link goes from *und* to *angesichts*:

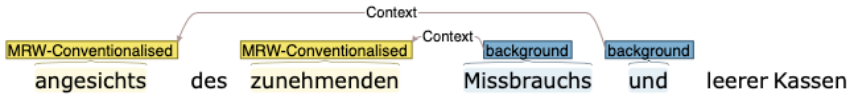


Figure 27: A conjoined phrase as background

When compounds are conjoined, repeated parts can be omitted, e. g. in (48). Such cases are treated analogously: To indicate the background status of *Gedanken- bzw. Informationsaustausch* ‘exchange of thoughts, or rather, information’ for the preposition *bei* ‘at, with’, the conjunction *bzw.* ‘or, rather’ is the origin of the *Context* link to the metaphorical *bei*:

- (48) *bei dem Gedanken- bzw. Informationsaustausch*
 ‘during the exchange of thoughts, or rather, information’

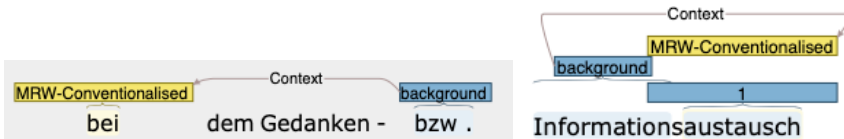


Figure 28: A conjoined phrase as background

4.2.9 Additional observations on backgrounds

The relation between backgrounds and metaphors is many-many: First, metaphors can have more than one background, which happens typically when a metaphorical verb has several of its arguments as backgrounds, as illustrated by example (33) and its annotation in Fig. 15.

Second, a single item can serve as the background of several metaphors, e. g. in PPs like (38), where *Jahren* ‘years’ is labelled as a *background* and linked by *Context* relations to its metaphorical modifier *nächsten* ‘next’ and (as the head of the complement of *in*) to its metaphorical head *in*. Both metaphors convey a non-spatial, temporal sense. For the annotation, see Fig. 18 above.

Furthermore, there are chains of metaphor-background pairs, in which the metaphor in one pair is simultaneously an (im-)proper part of the background in the next pair. For instance, in (49), *der Bitterkeit* ‘of (the) bitterness’ is the background for the metaphorical *Strudel* ‘vortex’, then *dem Strudel der Bitterkeit* ‘the vortex of bitterness’ is the background for *aus* ‘from’. The PP *aus dem Strudel der Bitterkeit* is then the background for *herausziehen* ‘pull out’:

- (49) *Mit seiner Intervention zieht Jesus den Mann heraus aus dem Strudel der Bitterkeit.*
 ‘With his intervention, Jesus pulls the man from the vortex of bitterness.’

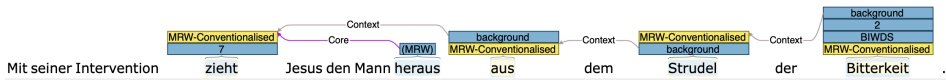


Figure 29: A metaphor-background chain

5 Evaluating the guidelines

To test the quality of the guidelines, we conducted a study in which two annotators independently annotated an additional text (not included in the corpus). The annotators had previously annotated parts of the corpus and adjudicated their results subsequently. That is, the study tried to replicate a realistic setting, in which annotators undergo training on the basis of the guidelines before their annotations are actually included into the respective corpus.

For this study, we chose a sermon because sermons emerged as a challenging text type in our corpus due to their high degree of metaphoricity; consequently, the study presents a representative test case for our annotation enterprise. For

the task of identifying the metaphorical units in the text of 1455 words, the annotators agreed on 95.9 % of the cases, jointly identifying 258 metaphors. The corresponding value of Cohen's kappa was .87 (almost perfect agreement between the annotators according to Landis & Koch 1977).

Furthermore, for the jointly identified 258 metaphors, the annotators agreed on the background in 224 cases (86.8 %). There was less agreement on extended metaphors (the first annotator found six and the second one identified four, with an overlap of three); also, the first annotator annotated one non-conventionalised metaphor, which was not confirmed by the second annotator. These results are in line with our experiences from the adjudication process, in which most of the time was dedicated to these metaphor phenomena.

Funding information

This work was funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) – SFB 1412, 416591334.

Acknowledgements

These guidelines were compiled as part of our work in the projects 'Metaphor and metonymy as register phenomena' and 'GeRMaN: German register marking by non-literal expressions' in the Collaborative Research Centre 1412 *Register: language users' knowledge of situational-functional variation*. Thanks to the whole team, past and present, Jenny Lee, Leon Rake, Patrick Saliger, and especially to Valia Kordoni and Lottix Feder.

References

- Biber, Douglas. 2009. Multi-dimensional approaches. In Anke Lüdeling & Merja Kytö (eds.), *Corpus linguistics. An international handbook*, 822–855. Berlin: Mouton de Gruyter.
- Blaette, Andreas. 2017. *GermaParl. Corpus of plenary protocols of the German Bundestag*. TEI files. <https://github.com/PolMine/GermaParlTEI>.
- Egg, Markus. 2026. Metaphors as register markers. *Corpus Linguistics and Linguistic Theory*. DOI: [doi:10.1515/cllt-2025-0096](https://doi.org/10.1515/cllt-2025-0096).
- Egg, Markus & Valia Kordoni. 2022. Metaphor annotation for German. In *Proceedings of LREC 2022*, 2556–2562.

- Egg, Markus & Valia Kordoni. 2023. A corpus of metaphors as register markers. In *Findings of EACL 2023*, 220–226.
- Goatly, Andrew. 2011. *The language of metaphors*. 2nd. Abingdon: Routledge.
- Grady, Joseph. 1997. THEORIES ARE BUILDINGS revisited. *Cognitive Linguistics* 8. 267–290.
- Herrmann, Berenike, Karola Woll & Aletta Dorst. 2019. Linguistic metaphor identification in German. In Susan Nacey, Aletta Dorst, Tina Krennmayr & Gudrun Reijnierse (eds.), *Metaphor identification in multiple languages. MIPVU around the world*, 113–135. Amsterdam: Benjamins.
- Kemmann, Ansgar. 2013. Debatte. In Björn Rothstein & Claudia Müller (eds.), *Kernbegriffe der Sprachdidaktik Deutsch. Ein Handbuch*, 41–43. Hohengehren: Schneider.
- Klie, Jan-Christoph, Michael Bugert, Beto Boullosa, Richard Eckart de Castilho & Iryna Gurevych. 2018. The Inception platform: machine-assisted and knowledge-oriented interactive annotation. In *Proceedings of COLING 2018: system demonstrations*, 5–9.
- Koch, Peter & Wulf Oesterreicher. 1994. Schriftlichkeit und Sprache. In Hartmut Günther & Otto Ludwig (eds.), *Schrift und Schriftlichkeit. Writing and its use. Ein interdisziplinäres Handbuch internationaler Forschung. An interdisciplinary handbook of international research*, vol. 1, 587–604. Berlin: de Gruyter.
- Landis, Richard & Gary Koch. 1977. The measurement of observer agreement for categorical data. *Biometrics* 33(1). 159–174.
- Nacey, Susan, Aletta Dorst, Tina Krennmayr & Gudrun Reijnierse (eds.). 2019. *Metaphor identification in multiple languages. mipvu around the world*. Amsterdam: Benjamins.
- Qi, Peng, Yuhao Zhang, Yuhui Zhang, Jason Bolton & Christopher Manning. 2020. Stanza: a Python Natural Language Processing toolkit for many human languages. In *Proceedings of ACL 2020: system demonstrations*, 101–108.
- Rai, Sunny & Shampa Chakraverty. 2020. A survey on computational metaphor processing. *ACM Computing Surveys (CSUR)* 53. 1–37.
- Salesky, Elizabeth, Matthew Wiesner, Jacob Bremerman, Roldano Cattoni, Matteo Negri, Marco Turchi, Douglas Oard & Matt Post. 2021. The multilingual TEDx corpus for speech recognition and translation. In *Interspeech 2021*, 3655–3659.
- Stede, Manfred. 2004. The Potsdam Commentary Corpus. In Bonnie Webber & Donna Byron (eds.), *Acl 2004 workshop on discourse annotation*, 96–102. Barcelona, Spain.

- Steen, Gerard, Aletta Dorst, Berenike Herrmann, Anna Kaal, Tina Krennmayr & Trijntje Pasma. 2010. *A method for linguistic metaphor identification: From MIP to MIPVU*. Amsterdam: Benjamins.
- Tong, Xiaoyu, Ekaterina Shutova & Martha Lewis. 2021. Recent advances in neural metaphor processing: a linguistic, cognitive and social perspective. In *Proceedings of NAACL 2021*, 4673–4686.

Appendix 1: Dead metaphors

In the annotation, we usually ignore dead metaphors, that is, expressions that do not exhibit synchronically both a more basic and a context-based more abstract sense which are related in terms of similarity.

E. g. cases like *begreifen* ‘understand’ do not count as metaphorical, despite their clear metaphorical origin in earlier stages of the language. In the concrete case of *begreifen*, the present-day meaning emerged through a metaphor like the one in the English *grasp*. This also applies to originally metaphorical idiomatic expressions that have no literal interpretation today (e. g. in *in Bausch und Bogen* ‘completely’).

In very rare cases, however, such dead metaphors can be revived, typically, by being part of an extended metaphor. Such cases of revived metaphors are annotated (as a conventionalised metaphor). The only corpus example is (50), where *Herzstück* ‘centrepiece’ (formerly, ‘piece of the heart’) is revived (and consequently annotated as conventionalised metaphor), as it forms an extended metaphor with *Skalpell* ‘scalpel’ and *OP* ‘operation’:

- (50) *das Skalpell in die Hand nehmen und an das Herzstück des deutschen Zivilrechts herangehen (...) die OP an der Zivilprozessordnung*
‘grasp the scalpel and tackle the centerpiece of the German civil law (...) the operation of the code of civil procedure’

Appendix 2: Pronominal adverbials and similar formations

Pronominal adverbials are decomposed because they show systematic variability between *da(r)*- ‘there’, *hier*- ‘here’, and *wo(r)*- ‘where’ as first elements and because they can stand in for full PPs in cases in which a verb marks a complement with a specific preposition (so-called ‘prepositional objects’), e. g. in the following examples:

- (51) *Ich beziehe mich auf das Gesetz/darauf/hierauf.*
‘I refer to the law/to that/to this’
- (52) *Worauf beziehst du dich?*
‘What are you referring to?’

In prepositional objects, the preposition can be metaphorical, with the verb serving as its background, e. g. in *sich beziehen auf* ‘refer to’, where the preposition is used in an abstract, non-spatial sense. But then the regular variation between

pronominal adverbial and full prepositional objects entails that a verb can be the background for the (metaphorical) prepositional part of a pronominal adverbial in the very same way, as in *sich darauf beziehen* ‘refer to that’, which hence must be annotated. At the same time, the demonstrative/interrogative part is often metaphorical (when used in a non-spatial sense) and must be annotated independently. See (7) and Fig. 4 for a concrete example. This affects the following pronominal adverbials:¹¹

<i>daran, hieran, woran</i>	<i>festklammern an</i> ‘cling to’
<i>darauf, hierauf, worauf</i>	<i>bauen auf</i> ‘rely on’
<i>daraus, hieraus, woraus</i>	<i>folgen aus</i> ‘result from’
<i>dabei, hierbei, wobei</i>	<i>sich entschuldigen bei</i> ‘roam’
<i>dadurch, hierdurch, wodurch</i>	<i>streifen durch</i> ‘apologise to’
<i>dafür, hierfür, wofür</i>	<i>stehen für</i> ‘support’
<i>dagegen, hiergegen, wogegen</i>	<i>sich wenden gegen</i> ‘oppose’
<i>dahinter, hierhinter, wohinter</i>	<i>stehen hinter</i> ‘support’
<i>darin, hierin, worin</i>	<i>bestehen in</i> ‘consist of’
<i>damit, hiermit, womit</i>	<i>sich begnügen mit</i> ‘make do with’
<i>danach, hiernach, wonach</i>	<i>streben nach</i> ‘aspire’
<i>daneben, hierneben, woneben</i>	<i>bestehen neben</i> ‘be on a par with’
<i>darüber, hierüber, worüber</i>	<i>sprechen über</i> ‘discuss’
<i>darum, hierum, worum</i>	<i>kämpfen um</i> ‘fight for’
<i>darunter, hierunter, worunter</i>	<i>fallen unter</i> ‘be subsumed under’
<i>davon, hiervon, wovon</i>	<i>sprechen von</i> ‘talk about’
<i>davor, hiervoor, wovor</i>	<i>sich fürchten vor</i> ‘fear’
<i>dazu, hierzu, wozu</i>	<i>dienen zu</i> ‘serve’
<i>dazwischen, hierzwischen, wozwischen</i>	<i>unterscheiden zwischen</i> ‘distinguish’

Similar formations, which are no pronominal adverbials though, are formed with an inflected demonstrative part as first or second element, but are invariable and never used as verbal complements. They are ignored in the annotation,

¹¹*Damit* can also be a subordinating conjunction with the sense ‘in order to’, it can be distinguished from the pronominal adverbial by the verb-final word order of the clauses it occurs in, the different stress pattern, and by its obligatory clause-initial position.

because they no longer have a spatial interpretation. (Some expressions that are formed in this way are subordinating conjunctions, in particular, *indem* ‘by’ and *nachdem* ‘after’, with *seitdem* being ambiguous between an adverbial with the meaning ‘ever since’ and a conjunction ‘since’.) This group comprises the following expressions:

<i>demgegenüber</i>	<i>demgemäß</i>	<i>demnach</i>	<i>demnächst</i>
<i>demzufolge</i>	<i>deswegen</i>	<i>außerdem</i>	<i>indessen</i>
<i>infolgedessen</i>	<i>stattdessen</i>	<i>seitdem</i>	<i>trotzdem</i>
<i>währenddessen</i>	<i>zudem</i>		

Next, there are formations with the (originally spatial) *her* ‘towards the speaker (or the deictic origo)’ and *hin* ‘away from the speaker’ whose second element is either a demonstrative or a preposition. We do not decompose those but annotate them as metaphors if they are used in a non-spatial way, e. g. in (53), because they have a basic spatial interpretation, as illustrated by (54):

- (53) *Er hat dann hinterher einen Bericht darüber geschrieben*
‘Afterwards, he wrote a report about it’
- (54) *bevor er flink seinem Vater hinterher rennt*
‘before he quickly runs after his father’

Such formations include *daher*, *dahin*, *hierher*, *hierhin*, *hinterher*, and *nebenher* as well as *heraus*, *herbei*, or *herzu*. Inherently temporal – and hence never metaphorical – are *nachher* ‘later’, *vorher* ‘earlier’, *vorhin* ‘before’, *seither* ‘ever since’, and *weiterhin* ‘still’, however.

Finally, the reflexive *einander* ‘one another’ combines with prepositions as well. In analogy to pronominal adverbials, we decompose them in case the preposition is used metaphorically, e. g. *in aneinander denken* ‘think of each other’. This can affect the following combinations:

<i>aneinander</i>	<i>aufeinander</i>	<i>auseinander</i>	<i>beieinander</i>
<i>durcheinander</i>	<i>gegeneinander</i>	<i>hintereinander</i>	<i>ineinander</i>
<i>nebeneinander</i>	<i>übereinander</i>	<i>umeinander</i>	<i>untereinander</i>
<i>voneinander</i>	<i>voreinander</i>	<i>zueinander</i>	<i>zwischeneinander</i>