

Semantic Modelling of Adjective-Noun Collocations Using FrameNet

Yana Strakatova, Erhard Hinrichs

University of Tübingen, Germany

firstname.lastname@uni-tuebingen.de

Abstract

In this paper we argue that Frame Semantics (Fillmore, 1982) provides a good framework for semantic modelling of adjective-noun collocations. More specifically, the notion of a frame is rich enough to account for nouns from different semantic classes and to model semantic relations that hold between an adjective and a noun in terms of Frame Elements. We have substantiated these findings by considering a sample of adjective-noun collocations from German such as *enger Freund* ‘close friend’ and *starker Regen* ‘heavy rain’. The data sample is taken from different semantic fields identified in the German wordnet GermaNet (Hamp and Feldweg, 1997; Henrich and Hinrichs, 2010). The study is based on the electronic dictionary DWDS (Klein and Geyken, 2010) and uses the collocation extraction tool Wortprofil (Geyken et al., 2009). The FrameNet modelling is based on the online resource available at <http://framenet.icsi.berkeley.edu>. Since FrameNets are available for a range of typologically different languages, it is feasible to extend the current case study to other languages.

1 Introduction

Collocations such as *to make a mistake* and *black coffee* are multi-word expressions (MWEs) in which the choice of one constituent (*base*) is free, and the choice of the other one (*collocate*) is restricted and depends on the base (Wanner et al., 2006). Collocations are in the grey area between free phrases like *black car* and idiomatic MWEs such as *black sheep*, and in some cases it is challenging to draw the line between those concepts. As opposed to mere co-occurrences of words based on their frequencies, collocations show a certain degree of lexical rigidity which results in their partial lexicalization. This creates

difficulties for the non-native speakers when interpreting and especially producing such expressions because a substitution of the restricted component with a synonymous word is not allowed by the language (Bartsch, 2004). Therefore, combinations such as **to do a mistake* or **dark coffee* are not acceptable and sound unnatural to the native speakers, but they still can be interpreted correctly. Idiomatic MWEs such as *black sheep* are semantically opaque and belong to the domain of figurative language.

In spite of the fact that collocations have been getting more attention in the recent decades, there is a lack of systematic empirical studies on their semantic properties. Most of the previous corpus studies of collocations are concerned with their statistical properties and the ways to improve methods of automatic collocation extraction (Church et al., 1991; Smadja, 1993; Evert, 2004; Pecina, 2008; Bouma, 2009). These authors have shown that automatic and/or manual extraction of collocations is not an easy task. Our research does not attempt to contribute to this growing body of research. Rather, we focus on the classification and modelling of semantic relations that hold between a base and its collocate, e.g. the relation of degree that holds between the collocate *heavy* and its nominal base *rain*. More specifically, we will focus on the semantic relations that hold in adjective-noun collocations, since such collocations have received considerably less attention than verb-noun collocations.

In our research, we utilize existing lexical resources that reliably identify adjective-noun collocations. For purely opportunistic reasons, we have chosen German as our language of investigation since there are a number of digital resources for German, including the DWDS (short for the Digitales Wörterbuch der deutschen Sprache) (Klein and Geyken, 2010) and GermaNet (Hamp and

Feldweg, 1997; Henrich and Hinrichs, 2010), that offer a broad coverage of adjectives and nouns as the two word classes under investigation.

The remainder of this paper is structured as follows: Section 2 introduces the notion of collocation in more detail and describes the related work on the semantic classification of collocations. Section 3 presents our own proposal of how to deal with semantics of collocations; we argue that the notion of a semantic frame in the sense of FrameNet (Ruppenhofer et al., 2016) provides a suitably general semantic framework that is applicable to a wide range of semantic fields. Furthermore, we argue that collocations offer an interesting empirical domain for validating the structure of semantic frames and for further developing the FrameNet framework itself. The paper concludes with summary of our approach and with the discussion of different directions for future work.

2 Concept of collocation and related work

Following the logic of Nesselhauf (2003) and Mel'čuk (1998), we consider the following types of statistical co-occurrences true collocations:

1. the collocate has a specific sense with a limited number of words from different semantic fields, e.g. 'heavy' as intensifier: *heavy smoker, heavy rain, heavy traffic*. The adjective's sense is not prototypical, since it does not refer to the physical weight, but to intensity.
2. the collocate has a specific sense only with one or very few semantically related bases, e.g. *black coffee*. The adjective's sense here is not prototypical, since it does not refer to the colour, but to the fact, that no dairy products are added to the coffee.
3. the sense of the collocate is so specific that it can be used with only one or very few semantically closely related bases, e.g. *aquiline nose/face* (Mel'čuk, 1998). That is the adjective's only sense.

As our empirical basis we rely on the electronic dictionary DWDS. The DWDS contains a rich lexicographic treatment of collocations on the basis of the collocation extraction tool Wortprofil (Geyken et al., 2009). Figure 1 shows an excerpt of the Wortprofil for the German noun *Freund*

'friend'.¹ It illustrates the information contained in such a word profile.

As Wanner (2006) emphasizes, collocation extraction typically only results in lists of collocations that are classified according to their morpho-syntactic structure, but that do not provide any semantic information about the combinations. Semantic modelling of collocations requires a theoretical framework with a rich inventory that can be used for describing the relations between the base and its collocate. Such an inventory is offered in the form of Lexical Functions (LFs) in Mel'čuk's *Meaning ↔ Text Theory* (Mel'čuk, 1996). A LF is a function in the mathematical sense: $f(x) = y$, where a general and abstract sense f is expressed by a certain lexical unit y depending on the lexical unit x it is associated with (Mel'čuk, 1995). The number of *standard LFs* is limited to about 60, and they have fixed names, e.g. for intensifiers the LF *Magn* is suggested: *Magn* [RAIN] = *heavy*. For other cases the *non-standard LFs* are suggested. They are very specific, and their names are formulated in a natural language: e.g. *obtained in an illegal way* [MONEY] = *dirty*. LFs have been widely used in lexicographic projects on describing French semantic derivations and collocations (Polguere, 2000), and have also been implemented in the Spanish online dictionary of collocations (DiCE) that focuses on describing emotion lexemes (Vincze et al., 2011). Mel'čuk and Wanner (1994) employ LFs to represent collocation information for German lexemes from the semantic field of emotions. Wanner (2004) conducts experiments on automatic classification of Spanish verb-noun collocations based on the typology of LFs, and continues to work on this problem using different algorithms (Wanner et al., 2006).

The works by Wanner (2004; 2006) mostly concentrate on verbal collocations, for which the Meaning-Text Theory provides at least 24 simple verbal LFs that can further be combined into complex LFs. By comparison, adjective-noun collocations have received less attention and the set of proposed adjectival LFs is relatively small: there are six simple adjectival LFs (Mel'čuk, 2015). Thus, our main objective is to find a suitable framework for describing adjectival collocations. Jousse (2007) proposes a way of formalizing non-standard adjectival LFs through assign-

¹DWDS-Wortprofil for "Freund", generated from Digitales Wörterbuch der deutschen Sprache, <https://www.dwds.de/wp/Freund>, accessed on 04.29.2019.

hat Adjektivattribut	logDice ↓ ^q	Freq. ↓ ^q
1. eng	9.4	5005
2. lieb	8.5	1473
3. gut	8.5	15174
4. treu	8.5	1153
5. langjährig	8.4	1396
6. väterlich	8.3	926
7. alt	7.9	7896
8. falsch	7.1	967
9. gleichaltrig	7.0	362
10. verstorben	7.0	429

Figure 1: The top-10 adjective collocates as listed in the Wortprofil for the noun *Freund* ‘friend’.

ing attributes to the base word, e.g. shape, size, colour, function. These attributes can be compared to Frame Elements in Frame Semantics (Fillmore, 1982) and to the Qualia Roles in the theory of Generative Lexicon by J. Pustejovsky (1991). Qualia roles have been implemented as the underlying framework in the construction of SIMPLE lexicon (Bel et al., 2000). While they are easily applicable for the treatment of concrete nouns, they fail to suitably generalize the semantics of abstract nouns.

By contrast, the concept of semantic roles in Frame Semantics is not restricted to concrete nouns, but applies equally well to other semantic fields as well (for details see section 3 below). The main idea of Frame Semantics is that word meanings are defined relative to a set of semantic frames, which represent non-linguistic entities such as events, states of affairs, beliefs, and emotions, and which are evoked by the use of corresponding words in a particular language. Semantic Frames for English are described in the lexical database *FrameNet* (FrameNet-Database) in terms of *Frame Elements* (FEs) (Ruppenhofer et al., 2016). The database provides a rich coverage of nouns and adjectives from different semantic fields, currently there are 5558 nouns and 2396 adjectives, and the resource is under further development. The further advantage of FrameNet is that it can be adapted for other languages. As demonstrated by Boas (2005) and Padó (2007), a transfer of existing frame annotations from English to other languages is possible: there is a high degree of cross-lingual parallelism both for frames (70%) and for Frame Elements (90%) (Padó, 2007). For the reasons outlined above, we will use Frame El-

ements in the sense of FrameNet for the semantic modelling of adjective-noun collocations.

3 Semantic Modelling of Collocations

As motivated in the previous section, the main objective of this study is to develop a framework for semantic modelling of German adjective-noun collocations. To assess the applicability of FrameNet for modelling of collocations, we have investigated eleven frames for nouns from various semantic fields (see Table 1). The corresponding semantic fields were assigned according to the information from the German wordnet GermaNet, and the estimates about the degree of concreteness of the chosen nouns are provided by the MRC Psycholinguistic Database (Wilson, 1988). The nominal bases have been chosen on the basis of frequency and richness of collocates. The stage of choosing the candidates for modelling showed that there are significant differences in the behaviour of concrete and abstract nouns: the latter ones have a greater number and a richer variety of collocates (see Table 2). As explained in the previous section, we employ English FrameNet for German collocations. Semantic Frames in FrameNet describe non-linguistic concepts and deal with meanings rather than with particular lexical units in a language. Thus, a correct translation of the target German word into English makes it possible to apply the information contained in the English FrameNet to German data. In collocations, it is only the collocate (the adjective) that is language specific, and thus is problematic to translate. However, we consider the semantically transparent base (noun) to be the frame-evoking word, and such words do not cause any difficulties for translation.

3.1 Modelling concrete nouns

The number of true collocates for concrete nouns is relatively small due to several reasons. First of all, when combined with concrete nouns, most adjectives retain their prototypical meaning: *enge Straße* ‘narrow street’, *großes Haus* ‘big house’, *hoher Turm* ‘tall tower’, such expressions are considered free phrases. In addition, there are a lot of cases where a concrete noun is part of an idiomatic expression.²

²Depending on the context, a combination of two words with concrete meaning can either be a free phrase or an idiom: *roter Faden* lit. ‘red thread’, fig. ‘common theme’; *raues Pflaster* lit. ‘rough pavement’, fig. ‘harsh environment’.

Lexical Unit	MRC Rating	Semantic field in GermaNet	Frame in FrameNet
<i>Schokolade</i> ‘chocolate’	576	<i>Nahrung</i> ‘food’	Food
<i>Droge</i> ‘drug’	555	<i>Substanz</i> ‘substance’	Intoxicants
<i>Schuh</i> ‘shoe’	600	<i>Artifakt</i> ‘artefact’	Clothing
<i>Wald</i> ‘forest’	609	<i>Ort</i> ‘location’	Biological_area
<i>Regen</i> ‘rain’	600	<i>natPhaenomen</i> ‘phenomenon’	Precipitation
<i>Freund</i> ‘friend’	450	<i>Mensch</i> ‘person’	Personal_relationship
<i>Interesse</i> ‘interest’	305	<i>Kognition</i> ‘cognition’	Emotion_directed
<i>Angst</i> ‘fear’	326	<i>Gefühl</i> ‘feeling’	Fear
<i>Thema</i> ‘issue/topic’	338	<i>Kommunikation</i> ‘communication’	Point_of_dispute
<i>Strafe</i> ‘punishment’	358	<i>Geschehen</i> ‘event’	Rewards_and_punishments
<i>Preis</i> ‘price’	-	<i>Besitz</i> ‘possession’	Commerce_scenario

Table 1: The chosen nominal bases from different semantic fields. The ratings from the MRC Psycholinguistic Database (Wilson, 1988) indicate the level of concreteness of the nouns (in the range 100 to 700).

When concrete nouns do form true collocations, the sense of their collocates is not prototypical, yet it is highly conventionalized. Consider the following collocates of the word *Schokolade* ‘chocolate’: *schwarz* lit. ‘black’, *dunkel* ‘dark’, *weiß* ‘white’. In FrameNet the lexical unit (LU) ‘chocolate’ evokes the frame “Food” with Frame Elements (FEs) FOOD, CONSTITUENT_PARTS, DESCRIPTOR, and TYPE. Although it is true that dark chocolate has a darker colour than milk chocolate, when we use the expression *dunkle Schokolade*, we do not refer to the colour of the product, but to the fact that it contains a high percentage of cocoa and little or no milk. The same is true for *weiße Schokolade* ‘white chocolate’: it indeed has a very light colour, but it is due to the fact that such type of chocolate is made of cocoa butter and does not contain cocoa powder. FrameNet offers a suitable FE TYPE for describing the relation that holds between these adjectives and the noun. It is defined in FrameNet as follows: “This FE identifies a particular Type of the food item” (FrameNet-Database). A similar logic is applied to the collocates of the noun *Droge* ‘drug’: the collocates *hart* ‘hard’, *weich* ‘soft’, and *leicht* ‘light’ are accommodated by the FE TYPE within the frame “Intoxicants”.³ In the case of the artefact *Schuh* ‘shoe’, there are only two collocates (*hochhackig* ‘high-heeled’, *flach* ‘flat’) and the corresponding frame “Clothing” offers a suit-

goldene Nase (verdienen) lit. ‘(to earn) golden nose’, fig. ‘to earn a lot of money’; etc. Some cases include metonymy: *offenes Ohr* lit. ‘open ear’, fig. ‘person ready to listen’; *heller Kopf* lit. ‘bright head’, fig. ‘smart person’.

³Other similar examples are the collocations *grüner Tee* ‘green tea’ and *schwarzer Tee* ‘black tea’.

able semantic role STYLE.

When a noun is less concrete, e.g. *Regen* ‘rain’ that is a natural phenomenon and thus is a process, the list of its collocates is longer. The noun evokes the frame “Precipitation” and all the collocates are accommodated by the suitable frame elements. For example, under QUANTITY the following attributes are found: *sintflutartig* ‘torrential’, *stark* ‘heavy’, *kräftig* ‘heavy’, *leicht* ‘light’. All those adjectives describe rain in terms of the amount of water that falls in the process. The same is true for the modifier *strömend* ‘pouring’, however, it carries an extra meaning of the manner in which it can rain and is therefore assigned to the FE MANNER.

3.2 Modelling abstract nouns

Abstract concepts have a complex meaning which is reflected in the amount of semantic roles describing the corresponding frame and in the amount of attributes through which the semantic roles are realised in the language. For instance, according to the FrameNet Database (FrameNet-Database), the frame “Personal_relationship” evoked by the noun *Freund* ‘friend’ has the following non-core FEs:

- **Depictive:** Depictive phrase describing the Partners.
- **Degree:** Degree to which event occurs
- **Duration:** The length of the relationship.
- **Manner:** Manner of performing an action.
- **Means:** An act whereby a focal participant achieves an action indicated by the target.

LU	FE name	Collocates
<i>Schokolade</i>	Food (Core)	<i>heiß</i> ‘hot’
	Descriptor	<i>edel</i> ‘premium’
	Type	<i>schwarz</i> ‘dark’, <i>dunkel</i> ‘dark’, <i>weiß</i> ‘white’
<i>Droge</i>	Type	<i>hart</i> ‘hard’, <i>weich</i> ‘soft’, <i>leicht</i> ‘soft’
<i>Schuh</i>	Style	<i>hochhackig</i> ‘high-heeled’, <i>flach</i> ‘flat’
<i>Wald</i>	Descriptor	<i>tief</i> ‘deep’
	Contituent_parts	<i>licht</i> ‘open’
<i>Regen</i>	Precipitation (Core)	<i>sauer</i> ‘acid’
	Manner	<i>strömend</i> ‘pouring’
	Quantity	<i>sintflutartig</i> ‘torrential’, <i>stark</i> ‘heavy’, <i>kräftig</i> ‘heavy’, <i>leicht</i> ‘light’
<i>Freund</i>	Degree	<i>eng</i> ‘close’, <i>dick</i> ‘close’ (Pl)
	Duration	<i>alt</i> ‘old’
	Manner	<i>wahr</i> ‘true’, <i>echt</i> ‘real’, <i>falsch</i> ‘fake’
	Relationship	<i>fest</i> ‘boyfriend’
<i>Interesse</i>	Experiencer(Core)	<i>ureigen</i> ‘vested’, <i>widerstreitend</i> ‘conflicting’
	Topic(Core)	<i>materiell</i> ‘material’
	Degree	<i>groß</i> ‘strong/big’, <i>stark</i> ‘strong’, <i>hoch</i> ‘strong’, <i>massiv</i> ‘massive’
	Manner	<i>rege</i> ‘active’, <i>lebhaft</i> ‘lively’, <i>vital</i> ‘lively’, <i>echt</i> ‘genuine’, <i>wahr</i> ‘genuine’
	Parameter	<i>breit</i> ‘broad’, <i>handfest</i> ‘concrete’, <i>elementar</i> ‘fundamental’, <i>vital</i> ‘vital’
	Circumstances	<i>unmittelbar</i> ‘direct’
<i>Angst</i>	Degree	<i>groß</i> ‘strong/big’, <i>tief</i> ‘deep’, <i>höllisch</i> ‘hellish’
	Circumstances	<i>panisch</i> ‘panic’, <i>unterschwellig</i> ‘subconscious’, <i>krankhaft</i> ‘pathological’
	Manner	<i>blank</i> ‘sheer’, <i>pur</i> ‘pure’, <i>nackt</i> ‘pure’, <i>diffus</i> ‘vague’, <i>dumpf</i> ‘vague’
	Topic	<i>existenziell</i> ‘existential’
<i>Thema</i>	Domain	<i>sperrig</i> ‘unwieldy’, <i>weich</i> ‘vague’
	Time	<i>brennend</i> ‘urgent’, <i>drängend</i> ‘pressing’
	Group	<i>unbequem</i> ‘uncomfortable’, <i>heikel</i> ‘delicate’, <i>sensibel</i> ‘sensitive’, <i>brisant</i> ‘controversial’, <i>leidig</i> ‘vexed’
	Status	<i>groß</i> ‘big/major’, <i>heiß</i> ‘hot’, <i>beherrschend</i> ‘dominant’
<i>Strafe</i>	Instrument	<i>symbolisch</i> ‘symbolic’, <i>unmenschlich</i> ‘inhumane’
	Degree	<i>drakonisch</i> ‘draconian’, <i>hart</i> ‘harsh’, <i>empfindlich</i> ‘severe’ <i>saftig</i> ‘stiff’, <i>streng</i> ‘strict’, <i>scharf</i> ‘harsh’, <i>schwer</i> ‘heavy’, <i>deftig</i> ‘severe’, <i>hoch</i> ‘high’, <i>mild</i> ‘mild’, <i>niedrig</i> ‘weak’
<i>Preis</i>	Rate	<i>horrend</i> ‘horrendous’, <i>stolz</i> ‘stiff’, <i>hoch</i> ‘high’, <i>erschwinglich</i> ‘affordable’, <i>vernünftig</i> ‘reasonable’, <i>niedrig</i> ‘low’, <i>fest</i> ‘fixed’, <i>stabil</i> ‘stable’

Table 2: Semantic modelling of German adjective-noun collocations using Frame Elements from FrameNet.

- **Relationship:** The Relationship between Partners.
- **Source_of_relationship:** The source of the relationship.

The semantic roles as well as the name of the frame suggest that, in many contexts, the word ‘friend’ does not refer to a person as a human being of certain age, appearance, ethnicity, etc., but to the relationship people are engaged in. In German, the adjectives *eng* lit. ‘narrow’ or *dick* lit. ‘thick’ are both used with *Freund* in the sense ‘close’, thus describing the DEGREE of friendship. The collocate *alt* ‘old’ implies that the friendship has lasted for some time to the moment of speaking and can therefore be accommodate by the FE DURATION. When using *wahr* ‘true’, *echt* ‘real’, *falsch* ‘fake’ in connection with friendship, we refer to its quality, the most suitable FE of that kind in this case is MANNER. There are also borderline cases, when the suitable FE is not obvious, as in the case of the word *fest* ‘steady’ (lit. ‘solid’). At first glance, the modifier characterizes MANNER; however, in German, the expression *fester Freund* means ‘boyfriend’ that actually refers to the nature of the relationship between the partners. Therefore, the most suitable FE for that adjective is RELATIONSHIP. All the adjectival modifiers find corresponding semantic roles, however, not all the FEs are realised through adjectives and some of the slots such as MEANS or DEPICTIVE are left empty. Such unrealised FEs are not listed in Table 2.

An accurate mapping of collocates to corresponding FEs is possible for other semantic fields as well. Consider an example from the field of cognition: *Interesse* ‘interest’. In FrameNet it evokes the frame “Emotion_directed”. It has an EXPERIENCER referred to by the adjectives *ureigen* ‘vested’ and *widerstreitend* ‘conflicting’; MANNER (*rege* ‘active’, *lebhaft* ‘lively’, *vital* ‘lively’, *echt* ‘genuine’, and *wahr* ‘genuine’); TOPIC (*materiell* ‘material’); PARAMETER (*breit* ‘wide’, *handfest* ‘concrete’, *elementar* ‘fundamental’, and *vital* ‘vital’); and CIRCUMSTANCES (*unmittelbar* ‘direct’). It also has a property of intensity described in the frame as DEGREE. This FE accommodates the collocates *groß* ‘strong’, *stark* ‘strong’, *hoch* ‘strong’, and *massiv* ‘massive’.

A similar pattern is found for the emotion noun *Angst* ‘fear’. Consider its collocates:

groß ‘strong’, *nackt* ‘pure’, *höllisch* ‘hellish’, *panisch* ‘panic’, *pur* ‘pure’, *unterschwellig* ‘subconscious’, *blank* ‘sheer’, *diffus* ‘vague’, *tief* ‘deep’, *dumpf* ‘vague’, *existenziell* ‘existential’, *krankhaft* ‘pathological’

The identified relevant FEs are as follows (FrameNet-Database):

- **Degree:** The extent to which the Experiencer’s emotion deviates from the norm for the emotion.
- **Circumstances:** The Circumstances is the condition(s) under which the Stimulus evokes its response. In some cases it may appear without an explicit Stimulus. Quite often in such cases, the Stimulus can be inferred from the Circumstances.
- **Manner:** Any description of the way in which the Experiencer experiences the Stimulus which is not covered by more specific FEs, including secondary effects (quietly, loudly), and general descriptions comparing events (the same way). Manner may also describe a state of the Experiencer that affects the details of the emotional experience.
- **Topic:** The Topic is the general area in which the emotion occurs. It indicates a range of possible Stimulus.

The interpretation of some collocates is straightforward: the adjective *existenziell* ‘existential’ indicates the area of the stimulus and is modelled as TOPIC. The collocates *groß* ‘strong’ and *tief* ‘deep’ are used as intensifiers and are, therefore, assigned to the FE DEGREE. The word *höllisch* ‘hellish’ is frequently used as an intensifier with *Schmerz* ‘pain’ and carries the same meaning with ‘fear’, thus it is also assigned to DEGREE. The other adjectives do not reveal any information about the intensity of the experienced emotion: *blank* ‘sheer’, *pur* ‘pure’, and *nackt* ‘pure’ rather imply that, at a particular moment, fear is the only emotion guiding the behaviour of a person. This interpretation fits the definition of MANNER, and so do the collocates *diffus* ‘vague’ and *dumpf* ‘vague’. The remaining three adjectives (*panisch*, *unterschwellig*, *krankhaft*) could also be assigned to MANNER, however, there is more information in their meaning than it may seem. These collocations are very close to psychological terms, as well as ‘existential’, but they refer to certain conditions under which fear might be

experienced rather than to the area of the stimulus. In such cases context is helpful; consider the following examples from the DWDS-Wortprofil for the noun *Angst*⁴:

1. *Deshalb habe die Frau panische Angst vor ihrem sehr dominanten Mann gehabt.*
eng. 'That is why the woman had a panic fear of her dominant husband'.
2. *Dann spricht man von Erythrophobie, der krankhaften Angst zu erröten.*
eng. 'This is referred to as erythrophobia, a pathological fear of blushing'.
3. *Es ist eine unterschwellige, alltägliche Angst, mit der die Bürger leben.*
eng. 'It is a subconscious everyday fear the citizens live with'.

The examples illustrate that these three collocates describe a certain kind of fear triggered by a particular stimulus, but the stimulus itself can only be derived from the context. Thus, the most suitable semantic role for accommodating the collocates is CIRCUMSTANCES.

All the above described cases demonstrate that semantic roles present in abstract collocations are quite diverse, and the relations can well be generalized using FrameNet's inventory of frame elements. There are, however, nouns, that seem to be less diverse when it comes to the number of attributes realized through adjectives. This is the case when a noun has a certain kind of scale at the core of its meaning. For instance, the noun *Strafe* 'punishment/penalty' is mostly modified in terms of how strict the inflicted punishment is:

drakonisch 'draconian', *mild* 'mild', *hart* 'harsh', *empfindlich* 'severe', *hoch* 'high', *niedrig* 'weak', *saftig* 'stiff', *streng* 'strict', *scharf* 'harsh', *unmenschlich* 'inhumane', *schwer* 'heavy', *symbolisch* 'symbolic', *deftig* 'severe'

They can all be accommodated by the FE DEGREE. However, two adjectives from this list stand out in their meaning: *symbolisch* 'symbolic' and *unmenschlich* 'inhumane', they carry an extra meaning describing a kind of penalty, which is reflected in the FE INSTRUMENT ("The Instrument with which the reward or punishment is carried out" (FrameNet-Database)).

⁴DWDS-Wortprofil for "Angst", generated from Digitales Wörterbuch der deutschen Sprache, <https://www.dwds.de/wp/Angst>, accessed on 04.29.2019.

A similar situation holds for nouns from other semantic fields. Consider the noun 'price': it is defined in FrameNet as "the amount of money expected, required, or given in payment for something" (FrameNet-Database). The list of its collocates contains the following adjectives:

horrend 'horrendous', *vernünftig* 'reasonable', *erschwinglich* 'affordable', *stolz* 'stiff', *hoch* 'high', *niedrig* 'low', *fest* 'fixed', *stabil* 'stable'

They all refer to the scale "the amount of money", the latter two emphasize that there are no changes on the scale, whereas the others show the degree of how high the certain amount is from the point of view of the customer. The noun 'price' evokes the frame 'Commerce_scenario' with the following FEs: BUYER, SELLER, GOODS, MONEY, MEANS, PURPOSE, RATE, UNIT. The most suitable FE in this case is RATE that according to FrameNet describes price or payment per unit of Goods and is therefore the closest to the concept of a scale in this frame.

The examples illustrate that frame semantics offers a varied inventory for modelling semantic relations between the constituents of collocations independently of the semantic field of the noun, either concrete or abstract. FrameNet provides frame semantic information about many lexical units; however, it is still under development and there are cases, when the frame evoked by a noun does not reflect all the aspects of its meaning. This issue is discussed in more detail in the next subsection.

3.3 Challenges

More than one thousand frames are described in FrameNet, thus providing a rich coverage of the lexicon. However, there is always the fundamental issue of granularity that affects the groupings of LUs into frames. There are cases when adjectival collocates provide additional information about a word's semantics, but where there are no suitable FEs to accommodate this additional aspect of a word's meaning. The following examples illustrate the issue. Consider the collocates of the noun *Zukunft* 'future':

nah 'near', *unmittelbar* 'immediate', *fern* 'distant', *weit* 'distant', *entfernt* 'distant', *rosig* 'rosy', *glänzend* 'bright', *licht* 'bright', *golden* 'golden', *strahlend* 'bright', *hell* 'bright', *blühend* 'prosper-

ous’, *leuchtend* ‘bright’, *groß* ‘great’, *glanzvoll* ‘bright’, *dunkel* ‘dark’, *düster* ‘dark’, *stabil* ‘stable’

Some of them refer to the temporal proximity of future, the others are evaluative descriptors (mostly positive ones). The frame evoked by ‘future’ in FrameNet is “Alternatives” with the following FEs ([FrameNet-Database](#)):

- **Agent:** An individual involved in the Event.
- **Salient_entity:** An entity intimately involved in the Event.
- **Situation:** Something that may happen in the future, or at least whose factual status is unresolved. -
- **Number_of_possibilities:** The number of different future Events under consideration.
- **Purpose:** The state-of-affairs that the Agent hopes to bring about which is associated with some of the possible Events but not others.

None of the FEs reflects the evaluative or the temporal aspect of the meaning of the noun ‘future’ expressed by the collocates above. This means that additional FEs need to be inserted into the frame “Alternatives”. The most appropriate FEs appear to be **DESCRIPTOR** which in FrameNet refers to descriptive characteristics and properties, and **TIME**.

Consider another example: the frame “Calendric_unit” is evoked by LUs denoting seasons, days of the week, months, times of the day, etc. The FEs describing this frame refer to different aspects of time. However, some, but not all of the LUs that evoke this frame have collocates referring to the weather or the state of nature: *winter* can be ‘mild’ or ‘harsh’ (in the sense of temperature/weather), *autumn*, and *September* or *October* are ‘golden’. Such LUs should be accommodated by a subframe that inherits from the frame “Calendric_unit” and contains additional FEs referring to weather and/or state of nature.

4 Conclusion and future work

In this paper we have argued that Frame Semantics provides a good framework for semantic modelling of adjective-noun collocations. More specifically, the notion of a frame is rich enough to account for nouns from different semantic classes and to model semantic relations that hold between an adjective and a noun in terms of Frame Elements. We have substantiated these findings by

considering a sample of adjective-noun collocations from German that are taken from different semantic fields identified in the German wordnet GermaNet. We are grateful to the anonymous reviewer for raising an interesting question concerning the applicability of FrameNet’s semantic relations to adjective-noun free phrases as well.

In future research, we plan to perform the modelling on a larger scale. For this purpose, we are currently preparing a large dataset containing more than 2000 German adjective-noun collocations. We will continue to use the dictionary DWDS and its collocation extraction tool Wortprofil as the empirical basis for obtaining the data. The resulting data sample will cover nouns and adjectives from all the semantic classes identified in GermaNet. We will use this dataset to examine FrameNet’s coverage of lexical units from different semantic fields. But even if a lexical frame exists for a given noun, the Frame Elements included in the lexical frame may not suffice. As described in the previous subsection, the structure of some semantic frames lacks important FEs, which therefore need to be added. Therefore, the overall objective in the future work is to examine various semantic frames and their Frame Elements in terms of their comprehensiveness and applicability for modelling diverse relations that hold between collocation constituents.

A second important objective of our future research will be to address the question of reliability of annotations for the semantics of collocations on the basis of FrameNet. To this end, we plan to conduct an inter-annotator agreement study. This study will be informed by detailed instructions to the annotators in the form of written guidelines on how to identify the correct Frame Elements for a given collocation.

As mentioned in Section 2, one of the advantages of FrameNet is that it can be adapted for other languages. Therefore, it is worthwhile to conduct a comparative study on semantic annotation of collocations based on FrameNet for languages other than German. We plan to conduct such a study for Russian and English, since relevant resources and points of comparison are available for each of those two languages. For Russian, the Explanatory Combinatorial Dictionary of Russian ([Mel’cuk and Zholkovsky, 1984](#)) describes collocations in terms of Lexical Functions à la Mel’čuk. The Macmillan Collocations Dictionary

for Learners of English (Macmillan, 2010) provides a rich coverage of English lexicon with semantic grouping of collocates for each base word and uses short definitions to describe such semantic sets. We plan to evaluate the relative merits of different annotation schemes and expect that it will be of further benefit for our research on collocations as MWEs.

Extending the present study to Russian will also provide an opportunity to compare the present approach that classifies collocations in terms of Frame Elements with Mel'čuk's classification according to Lexical Functions. One noteworthy difference that is apparent already at this point is that FrameNet's semantic relations can also be applied to describe free phrases, whereas the application of LFs is limited to lexically restricted combinations (Mel'čuk, 1995; Mel'čuk, 2015).⁵

Acknowledgements

The authors gratefully acknowledge the financial support of the research reported here by the grant *Modellierung lexikalisch-semantischer Beziehungen von Kollokationen* awarded by the Deutsche Forschungsgemeinschaft (DFG). We would also like to thank three anonymous reviewers for their constructive remarks on an earlier version of this paper.

References

- Sabine Bartsch. 2004. *Structural and functional properties of collocations in English: A corpus study of lexical and pragmatic constraints on lexical co-occurrence*. Gunter Narr Verlag.
- Nuria Bel, Federica Busa, Nicoletta Calzolari, Elisabetta Gola, Alessandro Lenci, Monica Monachini, Antoine Ogonowski, Ivonne Peters, Wim Peters, Nilda Ruimy, Marta Villegas, and Antonio Zampolli. 2000. *SIMPLE: A general framework for the development of multilingual lexicons*. In *Proceedings of the Second International Conference on Language Resources and Evaluation (LREC'00)*, Athens, Greece. European Language Resources Association (ELRA).
- Hans C. Boas. 2005. *Semantic Frames as Interlingual Representations for Multilingual Lexical Databases*. *International Journal of Lexicography*, 18(4):445–478.
- Gerlof Bouma. 2009. Normalized (pointwise) mutual information in collocation extraction. *Proceedings of GSCL*, pages 31–40.
- Kenneth Church, William Gale, Patrick Hanks, and Donald Hindle. 1991. Using statistics in lexical analysis. In *Lexical Acquisition: Exploiting On-Line Resources to Build a Lexicon*, pages 115–164. Erlbaum.
- Stefan Evert. 2004. *The Statistics of Word Cooccurrences: Word Pairs and Collocations*. Ph.D. thesis, Institut für maschinelle Sprachverarbeitung, University of Stuttgart.
- Charles J. Fillmore. 1982. Frame Semantics. In *Linguistics in the Morning Calm*, pages 111–137. Hanshin Publishing Co., Seoul, South Korea.
- FrameNet-Database. [Berkeley FrameNet Project](#). FrameNet Project.
- Alexander Geyken, Jorg Didakowski, and Alexander Siebert. 2009. Generation of Word Profiles for Large German Corpora. *Corpus Analysis and Variation in Linguistics*, 1:141–157.
- Birgit Hamp and Helmut Feldweg. 1997. *GermaNet - a Lexical-Semantic Net for German*. In *Proceedings of the ACL workshop Automatic Information Extraction and Building of Lexical Semantic Resources for NLP Applications*.
- Verena Henrich and Erhard Hinrichs. 2010. *GernEiT - the GermaNet Editing Tool*. In *Proceedings of the Seventh conference on International Language Resources and Evaluation (LREC'10)*, Valletta, Malta. European Language Resources Association (ELRA).
- Anne-Laure Jousse. 2007. Formalizing non-standard lexical functions. *Wiener Slawistischer Almanach*, pages 219–228.
- Wolfgang Klein and Alexander Geyken. 2010. Das digitale Wörterbuch der deutschen Sprache (DWDS). In *Lexicographica: International annual for lexicography*, pages 79–96. De Gruyter.
- Macmillan. 2010. *Macmillan Collocations Dictionary for Learners of English*. Macmillan Education.
- Igor Mel'čuk. 1998. Collocations and lexical functions. In A P Cowie, editor, *Phraseology: Theory, analysis, and applications*, pages 23–53. Clarendon Press, Oxford.
- Igor Mel'čuk. 2015. *Semantics: From meaning to text*, volume 3. John Benjamins Publishing Company.
- Igor A Mel'čuk and Leo Wanner. 1994. Lexical co-occurrence and lexical inheritance. Emotion lexemes in German: A lexicographic case study. *Lexikos*, 4:86–161.
- Igor Mel'čuk. 1996. Lexical functions: a tool for the description of lexical relations in a lexicon. *Lexical functions in lexicography and natural language processing*, 31:37–102.

⁵We would like to thank an anonymous reviewer for raising this issue.

- Igor A Mel'čuk. 1995. *Russkij jazyk v modeli "Smysl-Tekst"*. Moskva, Vena.
- Igor A Mel'čuk and Alexander K Zholkovsky. 1984. Explanatory Combinatorial Dictionary of Modern Russian. *Wiener Slawistischer Almanach, Vienna*.
- Nadja Nesselhauf. 2003. The use of collocations by advanced learners of English and some implications for teaching. *Applied linguistics*, 24(2):223–242.
- Sebastian Padó. 2007. *Cross-lingual annotation projection models for role-semantic information*. Ph.D. thesis, Saarland University.
- Pavel Pecina. 2008. *Lexical Association Measures: Collocation Extraction*. Ph.D. thesis, Faculty of Mathematics and Physics, Charles University in Prague, Prague, Czech Republic.
- Alain Polguere. 2000. Towards a theoretically-motivated general public dictionary of semantic derivations and collocations for French. In *Proceedings of the 9th EURALEX International Congress*, pages 517–527, Stuttgart, Germany. Institut für Maschinelle Sprachverarbeitung.
- James Pustejovsky. 1991. [The Generative Lexicon](#). *Computational Linguistics*, 17(4):409–441.
- Josef Ruppenhofer, Michael Ellsworth, Myriam Schwarzer-Petruck, Christopher R Johnson, and Jan Scheffczyk. 2016. FrameNet II: Extended theory and practice.
- Frank Smadja. 1993. Retrieving collocations from text: Xtract. *Computational linguistics*, 19(1):143–177.
- Orsolya Vincze, Estela Mosqueira, and Margarita Alonso Ramos. 2011. An online collocation dictionary of Spanish. In *Proceedings of the 5th International Conference on Meaning-Text Theory. Barcelona*, pages 275–286.
- Leo Wanner. 2004. Towards automatic fine-grained semantic classification of verb-noun collocations. *Natural Language Engineering*, 10(2):95–143.
- Leo Wanner, Bernd Bohnet, and Mark Giereth. 2006. Making sense of collocations. *Computer Speech & Language*, 20(4):609–624.
- Michael Wilson. 1988. [MRC Psycholinguistic Database: Machine-usable dictionary, version 2.00](#). *Behavior Research Methods, Instruments, & Computers*, 20(1):6–10.