

Maria Ermakova/Alexander Geyken/  
Lothar Lemnitzer/Bernhard Roll

# INTEGRATION OF MULTI-WORD EXPRESSIONS INTO THE DIGITAL DICTIONARY OF GERMAN LANGUAGE (DWDS)

## Towards a lexicographic representation of phraseological variation

**Abstract** One central goal of the project ‘Zentrum für digitale Lexikographie der deutschen Sprache’ (Center for digital lexicography for the German Language, [www.zdl.org](http://www.zdl.org)) is to provide a corpus-based lexicographic component of common German multi-word expressions (MWE), including idioms, for DWDS ([www.dwds.de](http://www.dwds.de)), a general language dictionary of contemporary German. As a central challenge of this task, we have identified an adequate lexicographic representation of such common properties of MWE as variation and modification. To document the variation, we have developed a special entry-clustering model, which we call *hub-node entry*. This model comprises a core hub entry headed by a short nuclear form of the MWE and several node entries, which represent the most common variants in their full lexical forms.

**Keywords** Multi-word expressions, phraseological variation, dictionary entry structure

### 1. Introduction

The Center for Digital Lexicography for the German Language (ZDL), a project that is currently funded by the German Federal Ministry of Education and Research for the period from 2019 through 2023, pursues the goal to describe the German language comprehensively, while remaining true to a scientifically pure approach – on a digital platform that is accessible free of charge for everyone. This platform links several resources, dedicated to synchronic and diachronic descriptions of words. Users can query all resources simultaneously through a single search engine. The ZDL brings together digital dictionaries, both legacy dictionaries and dictionaries that are currently under development, extensive corpus resources covering several centuries of German language history, and digital linguistic tools that visualise syntactic and historical data for specific words.

The ZDL encompasses the “Digitales Wörterbuch der deutschen Sprache” (DWDS – Digital Dictionary of the German Language), a long-term project that is funded by the Berlin-Brandenburg academy of Sciences and the Humanities from 2007 through 2024. The DWDS provides a part of the contemporary language component of the project and the ZDL is currently building upon it. The DWDS provides an information system for the historical and present vocabulary of the German language ([www.dwds.de](http://www.dwds.de)). The various information sources are continuously updated. Both the ZDL and DWDS can be consulted free of charge.

ZDL and DWDS pursue two goals. Firstly, they aim at pooling and updating the lexical information from the large dictionaries published until now. Secondly, they strive to provide an information system that connects a reliable and scientifically sound lexicographical description of words with the possibility of researching the various uses of a word in well-documented text corpora.

One of the main goals of the project is to fully describe the majority of the frequently used multi-word expressions (MWE), focusing on idiomatic expressions (IE).<sup>1,2</sup>

The major source for the acquisition and description of these lexical items are large corpora of contemporary German as well as several existing lexical resources. Special attention is given to the formal variability of the MWE, a variability that can be observed in our corpora. Many MWE turn out to be semi-fixed, i. e. they can be modified in certain ways but are less flexible than freely composed phrases. A challenge of this project is to appropriately describe the range of regular variation for (groups of) MWE.

In this article, we will firstly review the traditional placement of the MWE in general dictionaries of German and shortly outline, how we integrate MWE in DWDS as full-fledged entries (section 2). The next section (section 3) addresses the types of phraseological variation that we deal with in our project and shows, how we treat the typical cases of moderate phraseological variation in DWDS. Further to this, we will direct the readers' attention on our modelling approach for idiomatic expressions with a wider range of internal variation. For the comprehensive record of the full range of (observed) variations, we developed an architecture of hub entries and node entries. In section 4 we will present and discuss this model. In section 5 we will present our conclusion and suggestions for further work.

## 2. Idiomatic expressions as part of traditional (print) dictionaries and the DWDS

MWE have since long been the object of lexicographical description, particularly in German lexicography. Up until now, there is a range of specialized dictionaries describing (semi-) fixed expressions such as collocations and idiomatic expressions, e. g. “Deutsche Idiomatik” (ed. H. Schemann) (2011) and Duden 11 (2020). Multi-word expressions of all kinds are also part and parcel of large monolingual general dictionaries of contemporary German. The most prominent examples of this species are the six-volume “Wörterbuch der deutschen Gegenwartssprache” (WDG 1962–1977), and, in the wake of this pioneering work, the one-volume Duden Universalwörterbuch (DUW 2018) and the one-volume Wahrig Deutsches Wörterbuch (WDW 2018). A retro-digitized version of the WDG is also part of the lexical stock of the DWDS.

In general language dictionaries, MWE are usually listed within the microstructure of single-word articles and highlighted with a specific typographic marker like an asterisk or bold type for them to be found easily by the users (Burger 2015, p. 185). WDG/DUW on the one hand and WDW on the other hand treat these items differently. While the former integrate these units in the sense part of the entry and subsume these units under the closest or most appropriate sense of the component word, the WDW does with lists of MWE at the end of

<sup>1</sup> We consider idiomatic expressions to be a subclass of multi-word expressions. The latter also contains lexical items such as complex nouns, complex adverbs, light verb constructions, etc.

<sup>2</sup> For an overview of the essential properties of multi-word expressions, see Hüning/Schlücker (2015). However, idiomatic expressions pose the largest challenges with regard to a description of their variability. They are therefore dealt with in the rest of the article and MWE and IE are used interchangeably.

an article, i. e. outside the proper sense descriptions of the headword. Even specialized dictionaries for idiomatic expression order the entries in accordance with this principle: MWE are grouped under the headwords, more precisely: under the first sense bearing headword of the MWE.

In contrast to this approach, our project is going to create full-fledged entries for multi-word expressions and therefore make them a part of the macro-structure. Furthermore, we intertwine both the entries for single word expressions and the entries for multi-word expressions through a system of cross-references and embedding (see Fig. 1 and Fig. 2 below for an example).

MWE are lexical units of their own right<sup>3</sup> and are as complex as single words: a) they sometimes unfold to a broad range of formal variants; b) in many cases they are used in more than one sense; c) each of these senses deserves a full lexicographical description, in particular i) a definition, ii) usage notes about style as well as temporal and local restrictions, iii) typical combinations with other words (i. e. typical collocations and internal as well as external modification), iv) a sufficiently rich set of usage examples drawn from corpora. Such a rich and detailed description is not possible if these MWE were part of the micro-structure(s) of their component words<sup>4</sup>.

In other words, articles for multi-word expressions follow (with few exceptions) the micro-structure of articles for single words. Technically speaking, one and the same XML-schema is used for both single word entries and multi-word entries. Of course, the same holds for the presentation of articles on the website.

On the other hand, MWE are embedded into the articles of their meaning bearing component words. Each single word entry has as one of its parts a list of MWE entries of which it is part (see Fig. 2). This list is automatically generated and draws on the “Bestandteile” (= components) section of the MWE. Figure 1 contains a list of the four meaning bearing components of the idiomatic expression *weit/ weitab vom Schuss* (lit. far/far away from shot; engl.: ‘far away; in the far distance’).

## weit vom Schuss

Grammatik	Mehrwortausdruck
Nebenform	<b>weitab vom Schuss</b>
Aussprache	■() ■()
Bestandteile	↗ <i>weit</i> ↗ <i>Schuss</i> , ↗ <i>weitab</i> ↗ <i>Schuss</i>
Rechtschreibregeln	§ 2, § 25 (E1)

**Fig. 1:** A screenshot of the article *weit vom Schuss/ weitab vom Schuss* in the DWDS, cf. <https://www.dwds.de/wb/weitab%20vom%20Schuss>

<sup>3</sup> Cf. Elsen (2017, p. 147).

<sup>4</sup> Cf. <https://www.dwds.de/wb/jmdn.%2C%20etw.%20an%20der%20Angel%20haben> or <https://www.dwds.de/wb/den%20Dienst%20versagen> to get an impression of the complexity of the lexicographical description of multi word expression.

The figure below shows a screenshot of the article *Schuss*, where, in the form section, all idiomatic expressions containing this word are listed.

## Schuss, der

Grammatik	Substantiv (Maskulinum) · Genitiv Singular: <b>Schusses</b> · Nominativ Plural: <b>Schüsse</b>
Nebenform	<b>Schuss</b> · Substantiv · Nominativ Plural: <b>Schuss</b> (bei Mengenangabe)
Aussprache	■ [ʃʊs]
Ungültige Schreibung	Schuß
Rechtschreibregeln	§ 2, § 25 (E1)
Wortbildung	mit ›Schuss‹ als Erstglied: ↗ <a href="#">Schussbahn</a> ... <a href="#">35 weitere</a> · mit ›Schuss‹ als Letztglied: ↗ <a href="#">Bauchschuss</a> ... <a href="#">57 weitere</a> · mit ›Schuss‹ als Binnenglied: ↗ <a href="#">Bolzenschussapparat</a> · ↗ <a href="#">Selbstschussanlage</a>
Mehrwortausdrücke	↗ <a href="#">Schuss in den Ofen</a> · ↗ <a href="#">Schuss ins Blaue</a> · ↗ <a href="#">den Schuss nicht gehört haben</a> · ↗ <a href="#">einen Schuss haben</a> · ↗ <a href="#">fern vom Schuss</a> · ↗ <a href="#">fernauf vom Schuss</a> · ↗ <a href="#">im Schuss</a> · ↗ <a href="#">in Schuss</a> · ↗ <a href="#">in Schuss kommen</a> · ↗ <a href="#">nur einen Schuss haben</a> · ↗ <a href="#">sich, etw. im Schuss halten</a> · ↗ <a href="#">sich, etw. in Schuss bringen</a> · ↗ <a href="#">sich, etw. in Schuss halten</a> · ↗ <a href="#">weit vom Schuss</a> · ↗ <a href="#">weitab vom Schuss</a> ... <a href="#">weniger</a>

**Fig. 2:** A screenshot of the article *Schuss* in the DWDS. Clicking on one of the listed MWE leads the user to the resp. article, cf. <https://www.dwds.de/wb/Schuss>.

In the following section we will present several examples of MWE entries in DWDS with the focus on the documentation of their phraseological variation.

### 3. Representation of the phraseological variation in ZDL project: Flexibility of the MWE as a challenge for lexicography

One of the features of MWE is their lexical and semantic stability (Fleischer 1982, p. 41). Variation in the lexical components of MWE is nevertheless a common phenomenon within this segment of the lexicon and cannot be ignored if the goal is to record the full range of recurrent patterns of variation that can be observed in actual language use. There is often more to the picture than meets the lexicographers' intuition (cf. Stumpf 2019, p. 116).

Phraseological variation is quite heterogeneous and comprises alteration of different components within MWE<sup>5</sup>. For example, the expression *am falschen Ende sparen* ('to make savings on the wrong end') can also be used with other nouns: *an der falschen Stelle sparen*, *am falschen Ort sparen*, *am falschen Platz sparen* ('to make savings in the wrong place'). In the expression *hartes Brot* ('hard bread'), which means 'a tough way', the noun can also be used with another adjective, i. e. *schweres Brot* ('heavy bread'). The expression *auf Achse sein* ('to be on the axis') meaning 'to be on journeys' can also be used with a definite article *auf der Achse sein*.

In traditional lexicography, variation is recorded on the basis of the lexicographers' intuition and not cross-checked with evidence from the corpora: many of these descriptions therefore draw an incomplete picture. Variable components are typically put into brackets or are separated from each other by a slash or comma, this leads to very complex forms of headwords that are hard to decode, e. g. from "Deutsche Idiomatik" by Hans Schemann (2011): *die Hand darauf/dadrauf/auf das/ein Versprechen... geben* ('to give one's hand on it / sth. / on a promise ...').

<sup>5</sup> For examples cf. Nunberg/Sag/Wasow (1984), Staffeldt (2009, p. 204).

In the ZDL project, we apply another method for MWE documentation, which avoids brackets and other punctuation symbols for highlighting variation. According to the complexity of the variation(s) on the part of the headword, we treat these either as a single entry or we create separate entries.

For example, in the case of moderate variation such as an alteration of an adjective, two (or more) separate entries are created, one for each variant:



**Fig. 3:** phraseological variants represented in two dictionary entries: screenshot of the article *hartes Brot* (<https://www.dwds.de/wb/hartes%20Brot>) and the article *schweres Brot* (<https://www.dwds.de/wb/schweres%20Brot>) – moderate variation of the head form.

Further cases of moderate variation of the headword, such as the alteration of prepositions or the extension by one word with no alteration in meaning, are dealt within one single entry, e. g. (see Fig. 4): *keinen Dunst haben* – *keinen blassen Dunst haben* ‘to have no (pale) haze’ meaning ‘to have not the faintest idea’.

## keinen Dunst haben

Grammatik Mehrwortausdruck  
 Nebenform **keinen blassen Dunst haben**  
 Aussprache   
 Bestandteile [Dunst](#)

### Bedeutung

**umgangssprachlich** (jmd. **hat keinen Dunst** von etw.), (jmd. **hat keinen blassen Dunst** von etw.)  
 Synonym zu [keine Ahnung haben](#)

#### BEISPIELE:

Das Schlimme an solchen Vögeln ist ja, dass der Experte aus dem Theater, Film und TV-Bereich weiß, dass das, was die sich da zusammen darstellen, nichts mit Schauspiel zu tun hat und man leicht an ihrem ergebnislosen, blassen Spiel erkennt, dass sie sich selbst spielen und von Rollenarbeit **keinen Dunst haben**. [Ohauerha, 16.07.2015, aufgerufen am 15.09.2018]

**Fig. 4:** phraseological variants represented in one dictionary entry: screenshot of the article *keinen Dunst haben* (<https://www.dwds.de/wb/keinen%20Dunst%20haben>) with the moderate extension of an otherwise stable headword.

More complex are the cases of variation that affect the integrity of the headword itself, with (example 3) or without (example 1) change in meaning. Frequently, the variation is limited to a small and arbitrary set of lexical units (example 2):

- (1) *sich verkaufen, laufen, gehen, weggehen ... wie geschnitten Brot*  
 refl. pronoun to sell, to run, to go, to go away ... as cut bread  
 ‘sell like hot cakes’

- (2) *sich zum Deppen, Narren, Horst, Affen ... machen*  
 refl. pronoun to the moron, fool, Horst, monkey ... make  
 ‘to make a fool of oneself’
- (3) *am Drücker sitzen, sein, bleiben*  
 at pusher sit, be, stay  
 ‘to be/stay in power (of sth.)’

In short, there are many types of variation. A flexible way to document the full range is necessary. In the following section, we will introduce one of the dictionary entry models that we have been developing for common variation types as the high alteration of the verb in verbal MWE (1, 3). Its function is to capture both the most frequent forms of the MWE as well as other less common variants and modifications, as obtained from our corpus.

#### 4. Hub-node model for groups of MWE entries

As mentioned above, some MWE families are too complex to be described in one or two dictionary entries. In particular, it concerns the MWE with high levels of verb variation. Therefore, we created an entry model, which we call hub-node, to document such cases. It is aimed at representing the relation of the stable (or core) and variable idiom components. A depiction of the model with one example is displayed in Figure 5.

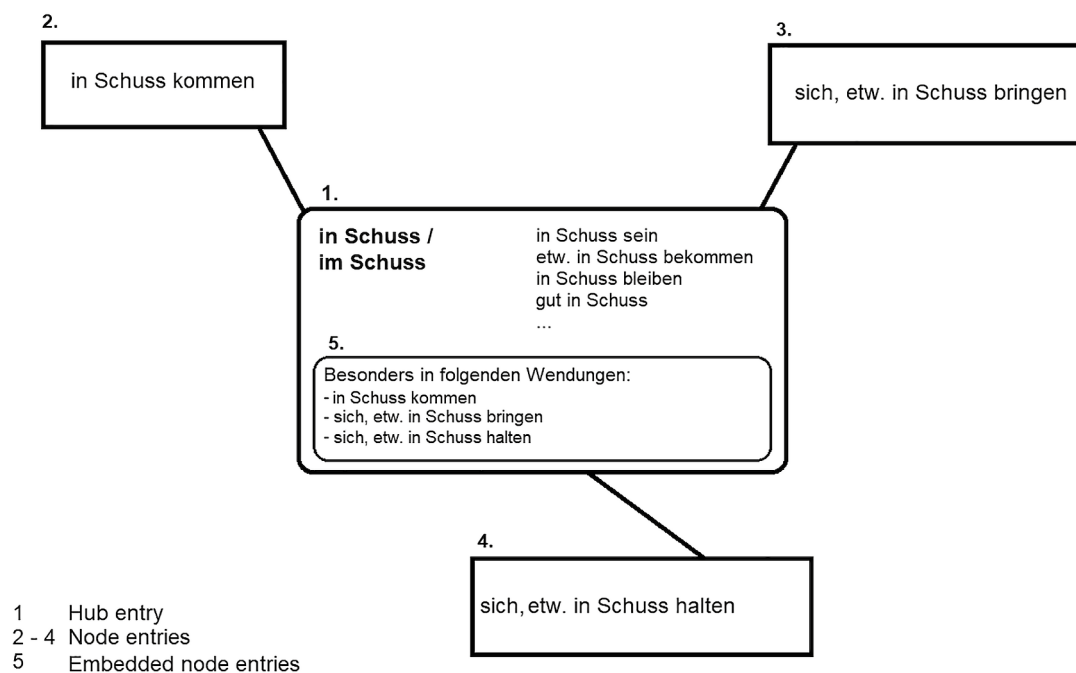


Fig. 5: Hub-node entry model

**Hub entry** In general, the hub entry such as *in Schuss/ im Schuss* (Fig. 5) meaning ‘in a good state’ serves as a central point in a (small) network of related entries. In particular, it serves as a container for less frequent variants of the idioms – variants which are typically not catered for in traditional dictionaries but are nonetheless present in corpora and might raise a need to consult the dictionary, e.g. *in Schuss sein* (‘to be in a good state’), *etw. in Schuss bekommen* (‘to get sth. in a good state’), *in Schuss bleiben* (‘to stay in a good state’) or *gut in Schuss* (‘in a good state’).

The main distinction of the hub entry from other MWE entry types is its lemma. In comparison to the usual MWE entries where the lemma is typically equal to an idiom (or to its canonical form known from other dictionaries) the hub entry is always titled with a phrase, which is the idiomatic nucleus or core metaphoric element of a number of the related idioms.

On the one hand, the reduction of the MWE lemma to its idiomatic nucleus allows the lexicographer to embed as many MWE variants as necessary in the dictionary entry; on the other hand, the idiomatic nucleus with its short and often underspecified form might look unfamiliar to the users. To eliminate this risk we decided to document the most common idioms, which the lemma of the hub entry is part of, as separate entries and describe them as node entries.

**Node entry** The primary function of the node entry is the lexicographic record of the salient and most common form of a verb idiom, which exhibits a high verb variation, under a distinct, easily recognisable lemma, in Figure 5: *sich, etw. in Schuss bringen* ('to bring oneself, sth. in a good state'), *etw. in Schuss halten* ('to keep sth. in a good state'), *in Schuss kommen* ('to come into a good state'). Node entries reduce the risk of the user not finding the commonly used forms of MWE.

Node entries follow the same micro-structure as the hub entry. They are fully-fledged entries of their own and can be consulted independently. Furthermore, they are projected into the hub entry with which they are associated.

Node entries are linked to the hub entry by a specialized link, which allows for the automatic projection of the content into the hub entry.

The node and the hub entry complement each other: the hub entry comprises the information not documented within the node entry and the node entry compensates for the short, often underspecified lemma of the hub entry.

With regard to its complexity, we consider the headword of a hub entry to be in between that of a single word headword and full idiomatic expression.

**Synoptic entry** As mentioned above, the content of the node entries is projected into the hub entry and forms a synopsis of them in it. In the DWDS-dictionary, it is positioned below the sense block. The synoptic section is headed by the phrase "Besonders in den folgenden Wendungen" ("mainly in the following expressions"). The function of this synopsis is a user-friendly representation of the common idiom forms in the hub entry and their easy accessibility of them through it.

In the following, we demonstrate the hub-node-model with the example of the hub entry *in Schuss/ im Schuss* ('in a good state') and one of the node entries *sich, etw. in Schuss bringen* ('to bring oneself, sth. into a good state') as well as the synoptic entry.

### Hub entry with the synoptic entry

**in Schuss / im Schuss**

Grammatik: Mehrwortausdruck  
Bestandteile: > Schuss

**Bedeutung** ZD-Vollartikel

umgangssprachlich in Ordnung, in (sehr) gutem, brauchbarem Zustand; in (sehr) guter Verfassung

KOLLOKATIONEN:  
als Prädikativ: in, im Schuss sein  
als Adverbialbestimmung: etw. in Schuss bekommen; etw. in Schuss haben; gut, körperlich in, im Schuss (sein, bleiben)

BEISPIELE:  
Auf der Fahrt geht es vorbei an verschlafenen Dörfern, an Reisigütern, von denen manche gut **in Schuss** sind; andere bröckeln vor sich hin und scheinen nur mehr von Katzen bewohnt zu sein. [Süddeutsche Zeitung, 01.05.2014]  
[...]

**Besonders in folgenden Wendungen**

**in Schuss kommen**

- umgangssprachlich (jmd. **kommt in Schuss**) eine gute körperliche Verfassung, die volle Leistungsstärke erreichen

BEISPIELE:  
Auf die deutschen Rodler und Bobfahrer ist Verlass, je näher die Olympischen Spiele in Peking rücken. Drei Wochen sind es noch bis dahin. Auch die Biathleten scheinen rechtzeitig **in Schuss zu kommen**. [Berliner Morgenpost, 16.01.2002]  
... 5 weitere Belege

- umgangssprachlich (etw. **kommt in Schuss**) (wieder) gebrauchsfertig, nutzbar sein; einen guten Zustand erreichen; sich positiv, dynamisch entwickeln

BEISPIELE:  
[...] Bekleidungsstücke wie Blazer, Strickjacken, Hosen [...] können über Nacht an die frische Luft gehängt werden, um wieder **in Schuss zu kommen**. Die feuchte Luft bringt sie in Form, Sitzfalten verschwinden und allfällige Gerüche werden neutralisiert. [Thüringer Zeitung, 13.12.2001]  
... 4 weitere Belege

**sich, etw. im Schuss halten**

- umgangssprachlich (jmd. **hält etw. in Schuss**) etw. zur Erhaltung eines guten, funktionsfähigen Zustandes mit den erforderlichen Maßnahmen behandeln; eine technische Anlage, ein Fahrzeug o. Ä. **warten**; Pflanzen, einen Garten o. Ä. **pflügen**

BEISPIELE:  
Wer ein Haus hat, ein Auto oder teure Kleidung, der muss sich drum kümmern. Man muss Dinge pflügen, warten, **im Schuss halten**. [Thüringer Zeitung, 05.02.2002]  
... 6 weitere Belege

- umgangssprachlich (jmd. **hält sich in Schuss**) durch Training, Gymnastik o. Ä. seine körperliche Leistungsfähigkeit erhalten

BEISPIELE:  
[R.] ist 65 Jahre alt und **hält sich** seit seiner Rente mit Sport **in Schuss**: »Für die Beweglichkeit und um ein paar Kilos zu verlieren, wie er sagt. [Neue Osnabrücker Zeitung, 13.02.2016]  
... 3 weitere Belege

**sich, etw. in Schuss bringen**

[...]

### Node entries

**in Schuss kommen**

Grammatik: Mehrwortausdruck  
Bestandteile: > Schuss > kommen

**Bedeutungsübersicht** ZD-Vollartikel

- umgangssprachlich (jmd. **kommt in Schuss**) eine gute körperliche Verfassung, die volle Leistungsstärke erreichen
- umgangssprachlich (etw. **kommt in Schuss**) (wieder) gebrauchsfertig, nutzbar sein; einen guten Zustand erreichen; sich positiv, dynamisch entwickeln

**Bedeutungen** ZD-Vollartikel

- umgangssprachlich (jmd. **kommt in Schuss**) eine gute körperliche Verfassung, die volle Leistungsstärke erreichen

[...]

**sich, etw. in Schuss halten**

Grammatik: Mehrwortausdruck  
Häufigkeit: sich, etw. im Schuss halten  
Bestandteile: > Schuss > halten

**Bedeutungsübersicht** ZD-Vollartikel

- umgangssprachlich (jmd. **hält etw. in Schuss**) etw. zur Erhaltung eines guten, funktionsfähigen Zustandes mit den erforderlichen Maßnahmen behandeln; eine technische Anlage, ein Fahrzeug o. Ä. **warten**; Pflanzen, einen Garten o. Ä. **pflügen**
- umgangssprachlich (jmd. **hält sich in Schuss**) durch Training, Gymnastik o. Ä. seine körperliche Leistungsfähigkeit erhalten

**Bedeutungen** ZD-Vollartikel

- umgangssprachlich (jmd. **hält etw. in Schuss**) etw. zur Erhaltung eines guten, funktionsfähigen Zustandes mit den erforderlichen Maßnahmen behandeln; eine technische Anlage, ein Fahrzeug o. Ä. **warten**; Pflanzen, einen Garten o. Ä. **pflügen**

[...]

**sich, etw. in Schuss bringen**

Grammatik: Mehrwortausdruck  
Bestandteile: > Schuss > bringen

**Bedeutungsübersicht** ZD-Vollartikel

- umgangssprachlich (jmd. **bringt etw. in Schuss**) etw. Vernachlässigtes, nicht mehr funktionstüchtiges reparieren, erneuern, renovieren
- umgangssprachlich (jmd. **bringt sich in Schuss**) den eigenen Körper trainieren, zur vollen Leistungsfähigkeit bringen; sich (wieder) fit machen

**Bedeutungen** ZD-Vollartikel

- umgangssprachlich (jmd. **bringt etw. in Schuss**) etw. Vernachlässigtes, nicht mehr funktionstüchtiges reparieren, erneuern, renovieren

[...]

**Fig. 6:** Hub entry *in Schuss/ im Schuss* (<https://www.dwds.de/wb/in%20Schuss>) and the respective node entries *in Schuss kommen* (<https://www.dwds.de/wb/in%20Schuss%20kommen>), *sich, etw. in Schuss halten* (<https://www.dwds.de/wb/sich%2C%20etw.%20in%20Schuss%20halten>), *sich, etw. in Schuss bringen* (<https://www.dwds.de/wb/sich%2C%20etw.%20in%20Schuss%20bringen>).

The model was successfully applied to other idiom groups, such as: hub entry *im/ in den Keller* ('extremely low / at the very low point') with the node entries *im Keller sein* ('to be at rock bottom' about prices, state of mood), *in den Keller sinken* ('to fall on the lowest level'), *in den Keller gehen* ('to drop to the lowest level'); *Wind in den Segeln* ('wind in the sails' meaning '(new) motivation'), *Wind in die Segel blasen* ('to blow wind in the sails'), *Wind in den Segeln sein* ('to be wind in the sails').

Additionally, it should be mentioned, that the model is only applicable in those cases, where the nucleus is a multi-word expression itself.

More generally, the hub-node-entry model is applicable in cases where a) the headword of the hub entry can be described independently of the nodes, i. e. it is a (complex, multi-word) lexical unit with its own semantic core such as *bittere Pille* ('bitter pill' meaning 'sth. unpleasant') and b) there is a set of sense-related node entries which is neither too large nor too small.

It has to be decided on a case-by-case basis and only after a comprehensive consultation of the corpora (and other dictionaries) which way of treating a particular MWE is the most appropriate one. However, we see some potential of the hub-node-model beyond the example that has been presented in this paper.



## 5. Conclusion and future work

In this article, we have introduced a model that helps to cope with the rich variation of use that comes along with many idiomatic expressions. Our model caters for both the less frequent but nonetheless relevant variants of an idiom (via hub entries) as well as with frequent variants (via node entries).

We try to follow a hybrid approach to lexicographically cover formal variation of various kinds of multi-word units:

- 1) most multi-word units are described in entries that are separate from the entries of their meaning-bearing components
- 2) the entries for the multi-word units are linked to the articles of their components and can be referenced from these component entries (see Fig. 2)
- 3) sets of multi-word entries with related meaning will be connected by lexical semantic cross-references in the same way as this is done in single word entries (e. g. synonymous senses, antonymous senses, associated senses, etc.)
- 4) multi-word expressions that share the same nucleus and exhibit a wide spectrum of variation can be described within the hub-node entry.

The next step for this approach will be to apply the hub-node entry model to other types of variation and modification. As future work, we plan to extend our work to other idiom patterns such as s. o. PREP ART N V where N is an open list of nouns, e. g. *jmdm. auf ART NN gehen* ('to get on someone's nerves (with something)) with an arbitrary range of lexical fillings for NN (*Nerven, Zeiger, Keks, Senkel ...*).

Among the central goals of the hub-node model is to give a much more comprehensive account of variation and modification of idioms in general language dictionaries in comparison to traditional dictionaries.

## References

### Dictionaries

Duden (2020): Duden 11. Redewendungen und sprichwörtliche Redensarten. Wörterbuch der deutschen Idiomatik. 5th edition. Berlin.

DUW = Deutsches Universalwörterbuch: Das umfassende Bedeutungswörterbuch der deutschen Gegenwartssprache. Hrsg. von der Dudenredaktion. 9th edition. Berlin 2019.

Schemann, H. (2011): Deutsche Idiomatik. Wörterbuch der deutschen Redensarten. Berlin/New York.

WDG = Wörterbuch der deutschen Gegenwartssprache. Hrsg. von Ruth Klappenbach und Wolfgang Steinitz. 6 volumes. Berlin, pp. 1964–1977.

WDW = WAHRIG Wörterbuch der deutschen Sprache. Hg. von Renate Wahrig-Burfeind. 27th edition. München 2018.

## Other publications

- Burger, H. (2015): *Phraseologie. Eine Einführung am Beispiel des Deutschen*. 5th edition. Berlin.
- Elsen, H. (2017): Wortgruppenlexeme zwischen Wortbildung und Phraseologie. In: *Yearbook of Phraseology* 8 (1), pp. 145–160.
- Fleischer, Wolfgang (1982): *Phraseologie der deutschen Gegenwartssprache*. Leipzig.
- Hüning, M./Schlücker, B. (2015): „Multi-word expressions“. In: Müller, P./Ohnheiser, I/Olsen, S./Rainer, F. (eds.): *Word-formation. An international handbook of the languages of Europe*. Berlin/New York, pp. 450–467.
- Nunberg, G./Sag, I. A./Wasow T. (1994): Idioms. In: *Language* 70 (3), pp. 491–538.
- Staffeldt, S. (2011): Die phraseologische Konstruktionsfamilie [X Präp *Hand* Verb]. In: *Zeitschrift für germanistische Linguistik* 39 (2), pp. 188–216.
- Stumpf, S. (2019): Phraseografie und Korpusanalyse. In: *Linguistik online* 96 (3/19), pp. 115–131. <https://bop.unibe.ch/linguistik-online/article/view/5523> (last access: 29-07-2021).

## Contact information

### **Maria Ermakova**

Zentrum für digitale Lexikographie der deutschen Sprache  
Berlin-Brandenburgische Akademie der Wissenschaften  
ermakova@bbaw.de

### **Alexander Geyken**

Zentrum für digitale Lexikographie der deutschen Sprache  
Berlin-Brandenburgische Akademie der Wissenschaften  
geyken@bbaw.de

### **Lothar Lemnitzer**

Zentrum für digitale Lexikographie der deutschen Sprache  
Berlin-Brandenburgische Akademie der Wissenschaften  
lemnitzer@bbaw.de

### **Bernhard Roll**

Zentrum für digitale Lexikographie der deutschen Sprache  
Berlin-Brandenburgische Akademie der Wissenschaften  
roll@bbaw.de

## Acknowledgements

The work in the project that has been funded by the German Ministry of Education and Research, (BMBF, FKZ: 01UG1902A). Special thanks goes to our colleagues who are currently working on dictionary entries for multi-word expressions and with whom we had lively discussions that finally led to the model presented here: Ralf Osterwinter, Lisa Palmes, Katrin Siebel, Stefanie Reckenthäler, Juliane Nau. We also thank Dmitrij Dobrovol'skij, Kathrin Steyer, Alexander Ziem and Carmen Mellado Blanco with whom we had the pleasure to discuss some of these issues in the Berlin-Workshop on MWE on Oct. 14<sup>th</sup>, 2021.