

# A Resource of German Light Verb Constructions Along with Possible Alternative Formulations

Ralf Laue and Martin Läuter

University of Applied Sciences Zwickau, Department of Computer Science  
Kornmarkt 1, 08056 Zwickau, Germany  
ralf.laue@fh-zwickau.de

## Abstract

In this paper we present a freely accessible resource covering German light verb constructions from the domain of administrative and business language (such as “Bericht erstatten” (to report)) together with their associated verb frames (in particular objects and governing prepositions). These constructions, which in many cases represent overly complex and difficult-to-read language, are joined within the same resource by simple verbs, also together with their respective verb frames.

This not only provides a type of thesaurus for light verb constructions that can assist in text simplification, but the documented verb frames can also support the automatic generation of grammatically correct suggestions for simpler textual alternatives.

## 1 Introduction

The aim of text simplification methods is to avoid complicated or uncommon words and phrases as well as to replace unnecessarily complex lexical or syntactic units by simpler constructs. The present contribution focuses on the latter topic, specifically on verb-noun collocations in German language that can be replaced by single verbs to improve text comprehensibility. To this end, we introduce a resource that contains a mapping between multi-word expressions and possible paraphrases. We limit our focus to the domain of administrative and business language, as this area is particularly often criticised for unnecessarily complicated formulations.

Research on simplified German has been particularly driven forward in recent years by lawmakers demanding barrier-free accessibility (in Germany: *Barrierefreie Informationstechnik-Verordnung*, *BITV 2.0* and *Barrierefreiheitsstärkungsgesetz*, *BFGS*).

However, independent of the needs of specific population groups, there is a demand for simple and understandable communication. Deficiencies

in this area are often criticised in the communication between authorities and citizens. [Ballod \(2020\)](#) writes in the preface to a book on administrative language: “Many people find texts from organisations and institutions, especially legal and administrative texts, difficult to read or even incomprehensible.”<sup>1</sup>

This statement is supported by the results of a survey conducted in 2008/09 by the Allensbach Institute, a German public opinion research institute, in which 86% of respondents stated that they had difficulties understanding written communications from authorities ([Eichhoff-Cyrus et al., 2009](#)).

[Eckhard and Friedrich \(2024\)](#) undertook an explorative interview-based study among civil servants. They point out that “... interviewees noted the importance of simple and plain administrative language that makes complex administrative procedures and regulations understandable.” Moreover, a study by [Fink et al. \(2021\)](#) demonstrated that clearly comprehensible official texts can positively influence democratic participation rights. In this study documents in German language related to public consultations on power lines were examined. The authors of the study concluded that “if the description of a power line is simplified between two consultation rounds, the number of contributions mentioning that power line increases.” In this paper, simplification is defined as readability, measured by Lix readability scores. ([Anderson, 1983](#))

In the calculation of these (and similar) readability scores, only surface features of the text, such as word and sentence lengths, are considered. Readability is therefore (as it is well known) not the same as understandability, but it is certainly an important tool for promoting understandability.

---

<sup>1</sup>All quoted texts from German-language books were translated to English by the authors of this article.

## 2 Light Verb Constructions

A phenomenon that is often associated with bureaucratic communication in German are light verb constructions and nominalisations.

This means that the meaning of a verbal phrase is either an interplay of the phrase's members, or a new meaning of its own. Examples are "Bericht erstatten" (to report) or "Planung durchführen" (to plan sth.).

This phenomenon has been studied under various names such as "support verb construction", "light verb construction", "function verb construction" or (for German language) "Funktionsverbgefüge", "verblasste Verben", etc. The term "light verb" refers to the fact that the semantics of such a verb is "either void or reduced to a small set of semantic features that are relevant for very large subclasses of verbs" (Langer, 2004).

There is no generally accepted definition of the term "light verb construction" (LVC) or the term "Funktionsverbgefüge" in German (van Pottelberge, 2008; Harm, 2021). A common point of existing definitions is their referring to a collocation where a verb – the light verb – is a main verb that has lost most of its concrete lexical semantics. It is not mainly the verb but its collocate that describes the action.

This article does not aim to discuss peculiarities of existing definitions, nuances between support verb constructions, function verb constructions, nominalisation constructions, etc. On the contrary – we want to introduce a resource that also includes collocations which are not typically regarded as LVC, for example, "Kopie anfertigen" (to make a copy). While this is a normal verb phrase consisting of a verb and an object, the object specifies the action more precisely than the verb, and the term can easily be rephrased as "kopieren" (to copy). We therefore want to introduce a collection of verb-noun collocations that could potentially be replaced by a simple verb to improve readability.

Such collocations have long been discussed as often unnecessary, hindering understanding and stylistically unappealing. There is no shortage of recommendations that such constructions be replaced by "strong" verbs whenever possible.

Wolf Schneider, the long-time director of the Hamburg Journalism School writes in one of his books about good style in professional German texts: "Similarly, when it comes to verbs... several groups should be viewed with suspicion: a) Ac-

tivity words that do not dare to stand alone on the plan but prefer to marry a noun." (Schneider, 1984, p. 45)

Germany's Federal Office of Administration published (Bundesverwaltungsamt, 2002) recommendations for administrative language with the goals of clarity, citizen-friendliness, and fostering cooperation between authorities and private individuals. One of the recommendations read: "In general, you should express an event using simple and meaningful verbs. When you force an event or activity into a nominal description (noun and meaningless verb), the sentence statement becomes an empty formula." The German Federal Ministry of Justice published the "Handbuch der Rechtsförmlichkeit" (Manual of Legal Formalities) with recommendations for the drafting of legal regulations. It states: "Prefer verbs! Avoid nouns!" (Bundesministerium der Justiz der Bundesrepublik Deutschland, 2008, S. 63) In the same manner, the Language Service of the Swiss Federal Chancellery recommends on LVCs: "If these constructions do not have their own meaning, you can replace them." (Schweizerische Eidgenossenschaft, Bundeskanzlei, Zentrale Sprachdienste, Sektion Deutsch, 2008)

One might argue that all these recommendations are merely based on stylistic and aesthetic perceptions and possibly have nothing to do with the actual readability and comprehensibility of texts. However, that is not the case. In (Wolfer, 2017), it was found using eye-tracking that nominalisations are processed significantly more slowly during reading than other nouns, and that reformulated texts, in which nominalisations are replaced by verbal structures, are read more quickly. This means that, at least for LVCs such as "Meldung erstatten" (to report) or "Mitteilung machen" (to make a notification / to inform), there is strong evidence supporting the replacement of these with verbal structures.

## 3 Collections of LVCs

German LVCs have been the subject of research for many years (von Polenz, 1963). This was also noted by Näf and Duffner (2006), who went on to argue: "Nevertheless, this research has mainly been concerned with questions of distinction and definition, and as a result has overlooked the empirical gathering of linguistic data." Although concrete proposals have been made in (Hanks et al., 2006) on how LVCs could be represented in a monolin-

gual dictionary, we are not aware of any initiatives that have implemented this on a large scale.

At least, Kamber (2008) examined the most frequently occurring German light verbs based on the SPIEGEL corpus (5 million words). The aim of this study was to improve word lists for university-level foreign language education, not to create a comprehensive overview of German LVCs. Therefore, Kamber limited the collection to capturing LVCs associated with the 10 most frequent light verbs in the corpus.

The goal of creating a more extensive list of LVCs was undertaken by Bruker (2013), who searched for such constructions in the TIGER corpus (900,000 tokens of German newspaper text) and the DWDS corpus (72 billion tokens), supplemented by findings of additional occurrences in the literature with the aim of expanding the Hagen German Lexicon (Hartrumpf et al., 2003). In this study, 2,500 LVCs were identified.

Our resource contains 4,700 LVCs, including 785 nominalisations with verbs such as “durchführen” (to carry out). The collection is deliberately limited to LVCs that typically appear in business or administrative contexts (i. e., we discarded LVCs such as “to give a kiss”). Those text types are the ones where simplification is often particularly necessary. An advantage of the present collection is that not only the LVCs were collected, but also the corresponding verb frames are documented, which can be helpful for the automatic generation of suggestions for alternative formulations. Furthermore, in the majority of cases, there are references to alternative formulations with the same meaning.

#### 4 Replacing LVCs by Verbal Constructions

In many cases, LVCs may be substituted with simple verbs to improve readability. However, this can not always be done, and certainly not easily in an automatic way.

In some cases, a clear recommendation can be made: A=“eine Erhöhung vornehmen” (to make an increase) should probably always be replaced by B= “etw. erhöhen” (to increase sth.). However, there are numerous less clear-cut cases. Such cases occur, for example, when a LVC either has no synonymous simple verb or multiple ones. An example of the first case is “Inventur durchführen” (to conduct an inventory), for which there is simply no corresponding verb. An example of an LVC with

multiple possible meanings is discussed in (Bruker, 2013): A=“Order erteilen” can be interpreted as meaning “to place an order” or “to give an instruction”. Depending on the meaning, A can either be replaced by “etw. bestellen” (to order sth.) or “jmd. anweisen” (to instruct so.).

At this point, an advantage of our resource become evident, because it not only includes light verb constructions (LVCs) and their single-verb paraphrases but also provides syntactic verb frames for both cases. For example, in the case of the LVC “Order erteilen” (to place an order), the resource might indicate that when accompanied by “bei+Dative”, the meaning “to order sth.” is preferred. On the other hand, the meaning “to give an instruction” should be chosen when the person being instructed is named as the pure dative object.

In other cases, as observed by Storrer (2007), an LVC “tends to develop specific subsenses or form characteristic collocations. Not all of them can be expressed by the base verb.” For instance, A=“Beschwerde einlegen” (to complain) and B=“beschweren” could be regarded as synonyms (and in fact, the word lists presented in this article do so). In informal texts, it may be appropriate to replace occurrences of A with B. However, in official language, it can make a difference whether we use A (someone files an appeal, possibly also implying that this appeal has been submitted in accordance with the deadline and formal requirements) or B (someone merely complains).

Furthermore, the possibility of substitution can also depend on the presence of other sentence components. A=“Frage stellen” (to ask a question) can in many instances be replaced by B=“fragen” (to ask). However, Kabatnik (2023) (as well as Storrer (2007)) examined this LVC more closely and identified numerous cases where a simple substitution is not advisable. This applies, for example, to formulations like “zwei Fragen stellen” (to ask two questions) or “eine wichtige Frage stellen” (to ask an important question), where the noun “Frage” is modified by an adjective.

Finally, a recommendation to replace an LVC can depend on the geographical origin of the intended audience: While A=“auf die Tagesordnung setzen” can be replaced by B=“traktandieren” in Switzerland, this would not work in Germany.

## 5 Our Resource and Possible Uses

An earlier version of the resource presented in this paper was developed within a project aimed at automatically analysing German-language business process models, semi-formal diagrams written in visual languages like BPMN (Object Management Group, 2009), which document the temporal sequence of business activities. The tasks in these diagrams are labelled with short phrases. In the context of automatic analysis, our goal was to identify the activity and object(s) based on the task labels. In the case of a label “Ablehnung aussprechen” (to express a refusal), it would not be sufficient to determine that the verb is “aussprechen” and the object is “Ablehnung”; rather, it must be recognised that the activity is “to refuse”. The details regarding the creation and structure of our resource can be found in (Laue et al., 2024), but the approach will be briefly outlined here:

From a corpus containing 6,711 business process models, we generated a list of words that have been tagged as verbs by the Hanover Tagger (Wartena, 2019) together with the words in their neighbourhood and manually inspected this list to identify all those cases where the identified verb alone would not be sufficient to describe the activity properly. To this list, we added LVCs from the already mentioned collection of Bruker if they are relevant to the business domain (based on the evaluation by the first author and another researcher) as well as all phrases from a collection of LVCs in the business domain published in (Marušić, 2015) for which 10 business reports from large German corporate groups had been analysed. Furthermore business-related verbs from (Schumacher, 1986), a compendium of German verbs, have been added. For the paper at hand, we additionally included 255 nominalisations of the type “Zählung durchführen” (to carry out counting) as these frequently occur in official texts and should be simplified.

As a result, we provide a freely available resource<sup>2</sup>, which contains verb frames for 4,700 LVCs and 3,600 simple verbs. All entries refer to the domain of administration and business. Each such expression (be it by an LVC or by a simple verb) is represented by an entry in a JSON file, and each entry consists of a syntactic verb frame with the verb infinitive occupying the head slot. In accordance with the terminology of (GermaNet), we refer to the structure of head and labelled slots

<sup>2</sup><https://github.com/bflowtoolbox/VerbframesDE>

as a syntactic verb frame, and we largely follow GermaNet for attaching labels to the slots, e. g. *vfin* for the finite verb, *NN* for the subject (subject), *AN* (accusative object), etc. The full description of the data format used can be found in (Laue et al., 2024), as well as in the documentation on GitHub. We will demonstrate the structure by means of an example:

“etw. geben” (=to give sth.) as a simple verb has a *NN* (subject) slot in addition to *AN* (accusative object) and *DN* (dative object). In contrast, “Alarm geben” (=to alarm/alert) has a *NN*, but its *AN* slot is filled with “Alarm”.

In some cases, it is helpful to mark a slot mandatory to be filled. While we cannot know in advance *which* object is used in an utterance, we want a verb frame only to match *if* there is an object present. Analogously a slot can be marked *not* to be filled, or *forbidden*. To give an example: “warten” (=to service sth.) has a mandatory *AN* slot, while for “warten” (=to wait) the *AN* slot is marked as “forbidden”. This allows for the consideration of different sentence structures as synonyms in the first case compared to the second. The entries in the JSON file for those verb slots look like this: 

```
{ "vfin": "warten", "optional": [ "NN", "auf+A" ], "forbidden": [ "AN" ], "synsetIds": [ 1088 ] }, { "vfin": "warten", "optional": [ "NN" ], "mandatory": [ "AN" ], "synsetIds": [ 1548 ] }
```

 To represent mappings within groups of synonymous expressions, we grouped the syntactic verb frames into sets of verb frames that are considered semantically equivalent, so-called synsets (represented by the *synsetIds* in the JSON file). Each verb frame belongs to at least one such synset, but some are assigned to more than one synset (e. g. “einführen” can have the meaning of “insert into”, but also “introduce sth./so.”).

This enables the retrieval of a range of alternative formulations from the same synset for a given complex expression. For the phrase “Auskunft einholen” (to obtain information), alternatives such as “fragen”, “erfragen”, etc. can be identified, alongside others like “Frage stellen”, all leading to the same synset.

The present resource comes without software, but contains enough linguistic information to be useful for a variety of automated processes, from phrase detection to example generation and verification of other systems. We are convinced that our resource can be a helpful tool, particularly in supporting the simplification of texts by editors.

## 6 Limitations

Besides the aforementioned lack of accompanying software, the greatest limitation lies in the restriction of the vocabulary to the domains of business and administration. Yet, given that texts from these domains often feature overly complex language, we believe the resource can still be helpful. Furthermore, the authors are aware that a resource like this can never be claimed to be complete; we intend to continue improving the entries in the future.

## Acknowledgments

Our research project was funded by the Deutsche Forschungsgemeinschaft – project 445156547.

## References

- Jonathan Anderson. 1983. Lix and Rix: Variations on a little-known readability index. *Journal of Reading*, 26(6):490–496.
- Matthias Ballod. 2020. *Klar-text in Organisationen: Ein Ratgeber zur Optimierung administrativer Informationen*. Springer Fachmedien, Wiesbaden.
- Astrid Bruker. 2013. *Funktionsverbgefüge im Deutschen*. Bachelor+Master Publ.
- Bundesministerium der Justiz der Bundesrepublik Deutschland. 2008. *Bekanntmachung des Handbuchs der Rechtsförmlichkeit*. *Bundesanzeiger*, 60(160a).
- Bundesverwaltungsamt. 2002. *Arbeitshandbuch Bürgernahe Verwaltungssprache*.
- Steffen Eckhard and Laurin Friedrich. 2024. How spoken administrative language affects citizen satisfaction. *Journal of Public Administration Research and Theory*, 34(1):122–135.
- Karin M. Eichhoff-Cyrus, Gerd Antos, and Rüdiger Schulz, editors. 2009. *Wie denken die Deutschen über die Rechts- und Verwaltungssprache?* Gesellschaft für deutsche Sprache e. V., Wiesbaden.
- Simon Fink, Eva Ruffing, Tobias Burst, and Sara Katharina Chinnow. 2021. Less complex language, more participation: how consultation documents shape participatory patterns. *Interest Groups & Advocacy*, 10(3):199–220.
- GermaNet. Verb frames, accessed 17 June 2025.
- Patrick Hanks, Anne Urbschat, and Elke Gehweiler. 2006. German light verb constructions in corpora and dictionaries. *International Journal of Lexicography*, 19(4):439–457.
- Volker Harm. 2021. *Funktionsverbgefüge des Deutschen: Untersuchungen zu einer Kategorie zwischen Lexikon und Grammatik*. De Gruyter.
- Sven Hartrumpf, Hermann Helbig, and Rainer Osswald. 2003. The Semantically Based Computer Lexicon HaGenLex – Structure and Technological Environment. *Traitement automatique des langues*, 44(2):81–105.
- Susanne Kabatnik. 2023. "Das ist doch etwas Totes" – Funktionsverbgefüge in der digitalen Sprachkritik. *Journal für Medienlinguistik*, 5(1):78–112.
- Alain Kamber. 2008. *Funktionsverbgefüge – empirisch: Eine korpusbasierte Untersuchung zu den nominalen Prädikaten des Deutschen*. Dissertation, Université de Neuchâtel, Tübingen.
- Stefan Langer. 2004. A formal specification of support verb constructions. In Stefan Langer and Daniel Schnorbusch, editors, *Semantik im Lexikon*, pages 179–201. Gunter Narr Verlag, Tübingen.
- Ralf Laue, Kristin Kutzner, and Martin Läter. 2024. Collecting activities and states in German business process models. In *Business Process Management Workshops*, pages 466–477. Springer Nature.
- Borislav Marušić. 2015. *Funktionsverbgefüge in deutscher Konzernsprache*. Phd thesis, University of Osijek, Croatia.
- Anton Näf and Rolf Duffner. 2006. *Korpuslinguistik im Zeitalter der Textdatenbanken. Vorwort der Herausgeber*. *Linguistik Online*, 28(3):7–22.
- Object Management Group. 2009. *Business process model and notation (BPMN)*.
- Wolf Schneider. 1984. *Deutsch für Profis*. Goldmann.
- Helmut Schumacher, editor. 1986. *Verben in Feldern*. Walter de Gruyter.
- Schweizerische Eidgenossenschaft, Bundeskanzlei, Zentrale Sprachdienste, Sektion Deutsch. 2008. *Behördenbriefe persönlich – sachgerecht – verständlich*.
- Angelika Storrer. 2007. Corpus-based investigations on German support verb constructions. In Christiane Fellbaum, editor, *Idioms and Collocations: Corpus-based Linguistic and Lexicographic Studies*, pages 164–187. Continuum.
- Jeroen van Pottelberge. 2008. Funktionsverbgefüge und verwandte Erscheinungen, in: *Phraseologie, Volume 1*, pages 436–444. De Gruyter Mouton.
- Peter von Polenz. 1963. Funktionsverben im heutigen Deutsch, volume 5 of *Wirkendes Wort: Beiheft*. Pädagogischer Verlag Schwann, Düsseldorf.
- Christian Wartena. 2019. A probabilistic morphology model for German lemmatization. In *15th Conference on Natural Language Processing*.
- Sascha Wolfer. 2017. *Verstehen und Verständlichkeit juristisch-fachsprachlicher Texte*, volume 7 of *Korpuslinguistik und interdisziplinäre Perspektiven auf Sprache*. Narr Francke Attempto, Tübingen.