The Better the Description, the Better the Instruction – Passives and Impersonals in the German Language of (Automotive) Engineering

Antje Heine, Greifswald, Lars Schirrmeister, Greifswald,

Matthew O. Anderson, Austin, Texas

ISSN 1470 – 9570

The Better the Description, the Better the Instruction – Passives and Impersonals in the German Language of (Automotive) Engineering

Antje Heine, Greifswald, Lars Schirrmeister, Greifswald,
Matthew O. Anderson, Austin, Texas

Engineering faculty at German universities often observe an absence of German for Specific Purposes (GSP) proficiency among international students, particularly regarding passives and impersonal expressions. This problem partially results from inadequate grammatical description in academic instruction materials, which usually explain these phenomena as transformations of underlying structures using valency-based projectionist grammar models. In these materials, for example, the passive is usually represented as deagentivation and valency-decrease. They present the production of passive clauses as a rather "mechanical" transformation of transitive verbs into different synonymous passive configurations, eliding differences in the latter's distributional properties and communicative functions. Contrary to these findings, studies like Stefanowitsch (2009) and Heine (2016) indicate that native speakers/writers prefer (i) certain passive constructions for specific text types and communicative purposes, and (ii) certain verbs in specific passive constructions. Our investigation draws on GSP research in the field of engineering, using the "Gingko" corpus. It aims to provide a detailed, empirical description of the use of passive/neighboring forms to modify future teaching materials.

1. Introduction

In this paper we discuss both the need for and the present lack of a data- and usage-based linguistic description concerning the language of engineering in academic contexts. We argue that this lack results in the deficient German for Specific Purposes (GSP) proficiencies among international students of engineering that are regularly reported by university lecturers in this field. We examine this problem in a German as a Foreign Language (GFL) context by taking a closer look at a group of grammatical constructions which is often said to be very prominent in the German language of engineering – the passive and its substitute forms. The paper is structured as follows: Section 2 outlines the general problems that led to our investigation, while Section 3 lays out the description of passives and impersonals in the German language of engineering, reviews the most recent GFL studies on the phenomenon and gives a recent example of the presentation of passives and impersonals in instruction materials. We argue that the problem of teaching German passives generally results from partially inadequate linguistic description – in

Section 4 we present two studies which suggest that this description should be reviewed and modified, a finding that led to our investigation using the "Gingko" corpus described in Section 5. Section 6 presents and discusses some of the initial results and outlines the next steps of our investigation. In Section 7, we explain the benefits of our study especially for – but not limited to – GFL teaching.

2. Why a Description of the (German) Language of Engineering?

In recent years, German universities have registered increased enrollments from international students. The numbers rose from 244,775 (including 181,249 *Bildungsauslaender*[1]) in 2010 (DZHW/DAAD 2011: 9) to 358,895 (265,484 Bildungsauslaender) in 2017 (DZHW/DAAD 2018). Bachelor's and master's programs in the engineering fields are currently benefiting from this positive enrollment trend. Recent statistics show that the number of *Bildungsauslaender* in engineering has increased steadily from 41,307 in 2010 (DZHW/DAAD 2011: 19) to 98,274 in 2017 (DZHW/ DAAD 2018: 5) making students of engineering the biggest group among *Bildungsauslaender*, with a total share of 37%. One explanation for increased enrollment might be the shift to English-language programs in some engineering subfields. On the other hand, by taking a closer look at the TU9[2] study programs, we can see that almost all programs in mechanical engineering are taught in German. The strong international reputation of programs in mechanical and electrical engineering is probably very appealing for international students, even if most of them are taught in German. As only very few programs in this field are taught (exclusively) in English at German universities, German language proficiency is crucial for students in order to succeed in them.

To be accepted into a German-speaking bachelor's or master's program, international students are required to pass an officially recognized, standardized language proficiency exam. Two of the most common exams are the "Deutsche Sprachprüfung für den Hochschulzugang – DSH" ('German Language Proficiency Test for the Admission to Higher Education') and TestDaF. Both exams focus particularly on *German as Common Language of Academia* (GCLA) (Rheindorf 2016) – which is considered part of *German*

[1] The term 'Bildungsauslaender' refers to "students of other nationalities who have obtained their higher education entrance qualification outside Germany." (DZHW/DAAD 2018: 2)

[2] 'TU9' is used to refer to the alliance of the nine leading Institutes of Technology in Germany (https://www.tu9.de/en/index.php).

for Specific Purposes (GSP) in the different academic (sub)fields – as well as testing students' more general *German as a Foreign Language* (GFL) proficiencies. To a large extent, university-level German language classes aim to provide training to international students in GFL and GCLA to pass these exams successfully.

Despite the prerequisite of passing the DSH, TestDaF or one of the few other accepted exams, lecturers in German engineering programs often note absent or deficient proficiencies in GFL, GCLA, and GSP among international students.[3] Unfortunately, university-level GSP teaching materials for students of engineering are still extremely scarce. Existing materials commonly focus on specific terminology and less on grammatical peculiarities, and usually replicate grammar instruction from general-purpose GFL or GCLA materials. As existing investigations into the characteristics of GCLA have been conducted by researchers in the liberal arts, GCLA materials also do not necessarily reflect the language of the sciences.

One of the reasons for the (at least partial) inadequacy of current GSP materials and courses is a lack of knowledge concerning the relevant text types and communicative situations which engineering students are expected to navigate on a regular basis. For example, GCLA classes at German universities emphasize text summaries and oral presentations – two important communication forms in liberal arts fields which are considerably less common in engineering study programs. Indeed, more extensive investigations on the linguistic features of engineering language and texts are rare and, in most cases, date back to the 1990s (e.g. Göpferich 1995, Moneiro et al. 1997, and Hanna 2003 – these last two publications are based on an investigation conducted at the Technische Universität Berlin).[4] Due to current trends in print and online publications, linguistic features of engineering texts might have changed in the past and in the present decade.

Additionally, even recent publications do not seem to base their recommendations for teaching GSP to students of engineering on data- and/or usage-based investigations, but instead solely on the (inter)subjective experiences of a limited number of teachers and

3 An example of a survey at the Institut für Automobiltechnik Dresden (IAD) can be accessed at: https://tu-dresden.de/bu/verkehr/iad/lvm/studium/sprachkompetenz.

4 A good example for this outdatedness is the fact that in the portfolio of one of the most important publishers in German as a Foreign Language, the Stauffenburg Verlag, the only GSP publication for scientists and engineers, "Baustein Mathematik für Naturwissenschaftler und Ingenieure", dates back to 1992!

lecturers. Besides this, most contributions concerned with the language(s) of sciences and especially engineering focus almost exclusively on lexicological aspects of the field-specific terminology, and not on grammatical structures (cf. Kärchner-Ober 2018). The latter is justified by the observation that syntactic features were less diverse in texts in engineering and the sciences (Borgwaldt & Sieradz 2018: 69), which implies that they might also pose less of an acquisition problem. Comparable statements can be found in the publication of Buhlmann & Fearns (2018: 34f.). Positions like this might be one of the reasons for still using explanations such as those found in Fig. 1.

However, in most publications on the topic, German passives and neighboring impersonal substitute constructions[5] are usually considered a very prominent group of expressions in both GCLA and GSP.

3. Impersonals and Passives in GFL

To our knowledge there are no recent corpus-based studies investigating the use of impersonal expressions in the German language of engineering. Existing studies –usually based on the individual experiences of teachers and lecturers – often highlight the much higher frequency of impersonal expressions within written GSP for engineering than in GFL (and even in GCLA), as a typical peculiarity. This is due to the omission of first-person singular and plural personal pronouns in written texts in this field. The most frequently used grammatical constructions for impersonal expressions in these texts are passives as in (1a) and (1b), middle constructions with non-causative *sich*+*lassen*+Infinitive ('let'+REFL+Infinitive)[6] as in (1c), constructions with *sein*+*zu*-Infinitive ('be+to'-Infinitive) as in (1d), the use of the indefinite pronoun *man* ('one') in active constructions as in (1e), and light verb constructions (1f).

(1) a. Das Projekt *wurde* *beendet*.
[NP$_{Patient}$NOM 'become' 3.Ps.Sg.PST 'complete' PST PTCP]
'The project was terminated.'

b. Die Untersuchung *ist* *abgeschlossen*.
[NP$_{Patient}$NOM 'be' 3.Ps.Sg.PRS 'conclude' PST PTCP]
'The study is concluded.'

[5] The term *neighboring impersonal substitute constructions* refers to different forms that, to a large extent, fulfill comparable communicative functions.

[6] Subsequent examples of the *sich*+*lassen*+Infinitve refer to the non-causative form – as in the example (1c) – and not to the causative form, as in "Peter lässt sich vom Friseur die Haare schneiden" ('Peter lets the barber cut his (Peter's) hair').

c. Die Temperatur *lässt* *sich* stufenlos *regeln*.
[$NP_{Patient}$NOM ‘let’ 3.Ps.Sg. PRS REFLPRO AKK ADV ‘control’ INF]
‘The temperature may be regulated continuously.’
d. Die Ergebnisse *sind* in der Abbildung *zu* *sehen*.
[$NP_{Patient}$NOM ‘be’ 3.Ps.Pl. PRS PP_{local} ‘to’ ‘see’ INF]
‘The results can be seen in the figure.’
e. *Man* *vermutet,* dass … .
[‘one’ INDEFPRO ‘suppose’ 3.Ps.Sg. ‘that’ …]
‘One assumes that...’
f. Der erste Prüfstand *ging* Ende letzten Jahres *in* *Betrieb*.
[$NP_{Patient}$ NOM ‘go’ 3.Ps.Sg.PST NP DAT NP GEN PREP NP AKK]
‘The first test bench went into operation at the end of last year.’

The German passives with *werden* (‘become’) and *sein* (‘be’), impersonal expressions with *man*, and light verb constructions can appear in combination with modal verbs. Constructions with *sich*+*lassen* and *sein*+*zu*-Infinitive include modality without explicitly using a modal verb. Instructors usually see the German passives and substitute forms – and their combination with modality – as especially challenging for students in developing academic language competence in GFL and GSP. One of the reasons for this might be the variety of alternative forms used for the expression of impersonality (as in 1a–f). GFL grammar instruction materials regularly feature lists and/or tables presenting the different passive configurations and their substitutes in combination with modal verbs or their inherent modal meaning (cf. Steinmetz & Dintera 2014: 64, Buscha & Szita 2011: 76). Unfortunately, empirical investigations concerning the usage constraints or distributional differences of passives and neighboring from a GFL perspective are still missing.

Few recent studies investigate the passive(s) in a GFL context. Bordag & Sieradz (2012) use an early version of the FALKO corpus (Reznicek, Lüdeling, et al. 2012) to investigate and describe, among other aspects, the interesting fact that GFL learners tend to acquire the *sein* passive before the *werden* passive, a finding that Abbot-Smith & Behrens (2006) had already made earlier regarding German L1 acquisition. Raeisi Dastenaei (2017) describes positive and negative transfer phenomena for Iranian GFL learners acquiring *werden* and *sein* passives in different tenses. Christante & Schimke (2018) investigate the processing of passive clauses by L1 German and L2 German children and find very similar developments. Kaiser & Peyer (2011) observe that, using a reading task, no significant differences can be found between the understanding of active and transferred passive clauses by L2 learners of German with French, Italian, and English L1 backgrounds across all levels of the Common European Framework of Reference for

Languages (CEFR). Hinrichs (1999) and Chen (2008) investigate German *werden* and *sein* passives from a contrastive perspective, the former in contrast to Spanish and the latter to Chinese. Even though many interesting aspects found in these studies will potentially improve the teaching of German passives, none of the investigations refers to or critically reviews the grammatical description current GFL instruction materials are based on.

Steinhoff (2011) concentrates on the pedagogical aspects of how instructional materials deal with passives as a teaching problem. Her investigation is based on an analysis of 26 GFL textbooks and concludes that materials (at least until 2009) usually do not distinguish whether the use of passive forms is preferred over active ones in specific text types, communicative situations, and/or to address certain communicative intentions. She also criticizes the order in which passives are taught in GFL and the type of exercises textbooks use to introduce and practice passives, which consist of transformations of active into passive sentences and vice versa. This type of transformation task can even be found in the standard exams such as DSH (FaDaF 2012). Therefore, active-passive transformations seem to be not only a pedagogical aid for passive training in GFL but also one of the end-goals of the instruction. Steinhoff explicitly expresses doubts about the effectiveness and even the plausibility of using structural transformation exercises, as native speakers/writers would not naturally take the 'detour' via the active version to formulate propositions in the passive voice (Steinhoff 2011: 54). If that is the case, then why should GFL-learners do so?

We strongly agree with Steinhoff's findings – our own analyses of actual instructional materials (from 2008 to 2017) revealed minimal progress in exercise and training variants. Only very few of the most recent materials include even a limited amount of information about certain passive configurations in specific text types.

In contrast to Steinhoff's study, which does not question the linguistic description of the phenomenon which instructional materials are based on, we argue that problems in both instruction and training at least partially result from inadequate grammatical description. In traditional valency-based projectionist grammar description, the German passives are usually represented as a type of deagentivation and valency-decrease. Accordingly, instructional materials present the production of passive clauses as a rather "mechanical"

transformation of potentially every verb (with a passive form) into different "synonymous" passive configurations without any usage restrictions (as shown in fig. 1).[7] As we will show in the following sections, the position of the traditional description does not hold from a quantitative data-based perspective and thus needs to be corrected.

Regel: Für Passiv mit Modalverb gibt es drei mögliche Ersatzformen
'Rule: The passive with modal verb has three possible substitutes'

Grammatische Form 'grammatical form'	**Beispielsätze** 'example sentences'
Passiv mit Modalverb 'passive with modal verb' (werden+can/must+PastPart)	Bei der Addition *können* die Summanden *vertauscht werden.*
Ersatzformen 'substitutes'	
Verwendung von *man* 'use of *man*' ('one' in active structures)	Bei der Addition *kann man* die Summanden vertauschen. / *Man kann* bei der Addition die Summanden vertauschen.
Adjektiv mit Endung *-bar* 'adjectivation with *–bar*'	Bei der Addition sind die Summanden *vertauschbar.*
sich lassen + Infinitiv 'sich + lassen + Infinitive	Bei der Addition *lassen sich* die Summanden vertauschen.
Die Bedeutung bleibt immer gleich. 'The meaning is always the same'	

Fig. 1: The German passive with modality and its substitutes in an instruction textbook for students of engineering (Steinmetz & Dintera 2014: 64).[8]

4. Recent Data-based Linguistic Description

In a study using the LIMAS corpus (roughly 1 million tokens, compiled from written German from the 1970s), Stefanowitsch (2009) investigates (filler) verbs that appear in the German passive *sein*-construction with a modal infinitive [$NP_{Patient/Theme}$ *sein* + *zu-Infinitiv*, NP 'be' + 'to'-Infinitive] (as seen in (1d)) and compares them to the appearance of the verbs in the neighboring active construction [NP_{Agent} *müssen VP(*Infinitiv*)*, NP + 'must' VP(Infinitive)] as in *Du musst die Maschine reinigen* ('You have to/must clean the machine').

(2) "a. Die Maschinenteile sind zu reinigen.
b *The machine parts are to clean.
c. The machine parts are to be cleaned. […].
d. You are to clean the machine parts." (Stefanowitsch 2009: 578)

[7] Steinmetz & Dintera (2014) is the only GFL textbook for students of engineering to date.

[8] Translations were provided by the authors of this paper.

The traditional valency-based description would predict a roughly equal distribution of a given filler verb over the two constructions in relation to its overall frequency in the corpus. Using distinctive collexeme analys (Gries & Stefanowitsch 2004), Stefanowitsch's study shows that there is no equal distribution of the majority of the filler verbs in the infinitive slots between the two constructions.[9] On the contrary, he observes that verbs belonging to the category "verbs of perception and mental activity" – like *ansehen* ('to look (at sth.)'), *verstehen* ('to understand'), *annehmen* ('to suppose') etc. –, and verbs that he associates with the Authority-Frame (Stefanowitsch 2009: 585),[10] appear significantly more often than expected in the first, but not in the second, construction relative to their overall frequency and the frequency of the two constructions.

As part of a smaller pre-study, Heine (2016) investigates the distribution of filler verbs in the passive *sein*+*zu*-Inifinitive construction and the *sich*+*lassen*+Infinitive "middle" construction. Due to the aforementioned restrictions on the use of personal forms in scientific GSP, these impersonal constructions are very frequent in technical texts, for example in the subfield of mechanical engineering. For her study, she uses a small corpus of 223 scientific articles (from 1998 to 2014) from the two most circulated German journals for automotive engineering, with a total of around 600,000 tokens.

She finds indications that some verbs in this text type tend to be notably more frequent in one specific impersonal construction than in the neighboring ones. To examine *sein*+*zu*-Inf, she takes a sample from a quarter of the corpus that includes 74 instances of *sein*+*zu*-Inf (Heine 2016: 304). Within these 74 instances, she finds 42 different verbs, which is not surprising. 29 of these verbs only appear in one instance of *sein*+*zu*-Inf, while the remaining 13 appear in 45 instances. What stands out in the data set is the fact that each one of the three verbs – *berücksichtigen* ('to consider'), *erkennen* ('to see, to recognize'), and *erwarten* ('to expect') – appears in 6 instances of the construction, that is, in almost a quarter of all instances. In the 373 instances of the *sich*+*lassen*+Inf construction in the corpus only *erkennen* appears 3 times, *berücksichtigen* and *erwarten* do not appear at all. Again, the traditional, valency-based grammar description used in GFL would predict an

[9] In the following, when we use the term *construction* we refer to it as defined by Goldberg within her *Construction Grammar* approach: "Any linguistic pattern is recognized as a construction as long as some aspect of its form or function is not strictly predictable from its component parts or from other constructions recognized to exist. In addition, patterns are stored as constructions even if they are fully predictable as long as they occur with sufficient frequency." (2006: 5)

[10] Following Ziem (2014), we adopt the frame-semantic convention of marking frame names using the Courier New font type.

equal distribution of the verbs in the neighboring construction relative to their overall frequency and the frequency of the two constructions (approx. 296 *sein*+*zu*-Inf vs. 373 *sich*+*lassen*+Inf).

Like Stefanowitsch (2009), Heine attempts to group all the verbs found in the *sein*+*zu*-Inf construction in terms of meaning, but does not explicitly use a frame semantic description. Instead, she tries to group them inductively by the communicative actions they perform. She states that all instances of the construction with its filler verbs can be divided into 6 groups by concrete communicative functions: (i) to communicate observations, results, and conclusions; (ii) to refer to conditions, causes, and constraints; (iii) to express goals, intentions, necessities; (iv) to name mathematical/physical activities, (v) to refer to a topic, and (vi) to refer to a figure. The first two of these categories are the ones with the highest number of different filler verbs used in the *sein*+*zu*-Inf construction.

All of these observations indicate that the scientific language of engineering is highly conventionalized and that authors apply specific patterns to produce their texts. The results of the pre-study have inspired new questions for a much more thorough and detailed investigation. The main goal of our current investigation into the GFL and GSP context is to produce a linguistic description that could be "translated" into more assertive statements like (hypothetically): "To communicate observations, results, and conclusions in engineering texts learners should use the construction *sein*+*zu*-Inf with one of the filler verbs a, b, c, d, e, or f but not g, h, I, j, etc." One of the main questions remains: which verbs or verb groups are licensed for specific constructions and which ones are not? It is important to state here that, in our opinion, statements like this should not be mistaken for explicit instructions provided to GFL/GSP learners, but are part of the background knowledge that instructors should have when teaching the passive(s).

5. The Gingko Corpus

Heine's pre-study served as an application for a research project grant with the German Research Foundation (DFG), which was approved in the middle of 2016 for a project duration of 3 years. Since the project's inception in 2017, we have been building "Gingko" ("**G**eschriebenes **ing**enieurwissenschaftliches **Ko**rpus," or 'Written Corpus for the Language of Engineering') at the University of Greifswald in collaboration with the Karlsruhe Institute of Technology (KIT). By the end of 2018, Gingko will consist of approximately 5 million tokens culled from 2,500 articles in the 2007-2016 volumes of

two scientific journals: Automobiltechnische Zeitschrift (ATZ) – Journal for Automotive Engineering –, and Moterentechnische Zeitschrift (MTZ) – Journal for Motor Engineering.[11] These two journals are the most widely-circulated in the German language in the subfield of automotive engineering and reflect the standard language usage for scientific articles within this field. We consider Gingko to be representative, given that all of the articles included were composed by field-experts from universities, research institutions, and companies. The majority of the articles were co-authored, and for most of them one (co-)author typically publishes no more than one article per year in either of the two journals, both of which produce 11 issues annually. As such, Gingko accounts for a wide dispersion of texts that reflect standard language usages within the field.

The tokens in Gingko are lemmatized and POS annotated (with the Stuttgart Tübingen Tag Set – STTS) using the TreeTagger. The current error ratio of the POS annotation is as low as 5%. Sentence annotation has also been added. For search queries, we use the browser-based platform ANNIS, which is described in Krause & Zeldes (2016). We plan to make Gingko publicly accessible after the project has been completed in April 2020.

Gingko is part of the broader research project, "Muster in der Sprache der Ingenieurwissenschaften – eine korpusbasierte konstruktionsgrammatische Analyse" ('A Corpus-Based Construction Grammar Approach to the Analysis of Patterns in the Language of Engineering'). As part of the project, we aim to describe the aforementioned impersonal forms – including modality – with a focus on the three constructions *sein*+*zu*-Infinitive, *müssen/können*+*werden*+Past Participle, and *sich*+*lassen*+Infinitive. First, we will measure the frequency of each of the three constructions in the corpus followed by measuring the overall frequency of lexical verbs. In the next step, the appearances of the most frequent lexical verbs in the three constructions will be compared using collexeme analysis.

In the next phase we intend to categorize the lexical filler verbs used in the constructions in terms of their meaning and communicative function. This will be followed by the description of the correlations between verb categories, constructional preferences, and communicative function. In the following section we present our initial findings.

[11] We are conducting our queries with the currently available part of Gingko (covering the issues from 2011-2016) with a total of around 3 million tokens.

5. Presentation, Discussion, and Interpretation of Initial Results[12]

As mentioned above, we think that the ways in which passives and their "substitutes" are typically presented and practiced in GFL teaching – especially preparatory GFL courses for studying at German universities – are insufficient because they do not reflect the actual usage preferences and restrictions of the passive and/or impersonal constructions. For example, the rules presented to learners through in-class instruction and instructional materials suggest that the use of passives in combination with modality could be interchanged with various so-called substitutes without considering which form would be the most adequate in which usage context function. This appears to explain the fact that in most grammar and language textbooks for the CEFR levels B2 and C1 and especially materials for preparatory GFL courses, we find transformation exercises where learners are forced to exchange a form of the passive with a structurally different substitute which had the "same meaning" (cf. Fig. 1).

The results given in Stefanowitsch (2009) and Heine (2016) show that neither a balanced distribution of the different passives and neighboring substitute constructions nor a balanced use of all verbs within these constructions can be found. In other words, we are convinced that, a) specific constructions are not randomly used or distributed in texts, and b) there are interrelations between the use of a specific construction and a specific verb (or group of verbs). In the following section, we will provide data-based evidence for a) and b) within the language of automotive engineering.

To test a) and b), the constructions shown in Tab. 1 were searched in Gingko.

Construction	Count of Instances
werden + Past Part. (without modal verb)	39928
Modal verb + *werden* + Past Part.	20123
sein + *zu* + Infinitive	7950
sich + *lassen* + Infinitive	3036
sein + V-*bar*	2162
man + V	1519

Tab. 1: Number of instances of 6 passive and substitute constructions

[12] We would like to thank our dear colleague Marlene Rummel for her work on the data analysis, and for the many good ideas and constructive discussions we shared on the statistic modeling within our project.

Even though it is clear that the constructions are not randomly interchangeable, what can be observed is that there is no balanced distribution. If a) is not correct, then there should be a roughly equal distribution of *sein*+*zu*-Inf, *sich*+*lassen*+Inf and *sein*+V-bar, which is not the case.

Collexeme analysis (cf. Stefanowitsch & Gries 2003, Gries 2015) is used to calculate the usage preference of certain lexical verbs (VV) within certain constructions (C), as expressed in b). In short, we determine whether a Verb VV_1 in a construction C appears randomly or not. For this calculation we need the values given in Table 2:

	Lexical verb (VV_1)	Other lexical verbs (VV_n)	Total
Construction C	VV_1 in C	VV_n in C	
Other constructions	VV_1 outside C	VV_n outside C	
Total			

Tab. 2: Cross tabulation model for collexeme analysis

If we take the lexical verb *realisieren* ('to put into practice') within the construction *sich*+*lassen*+Inf as an example, the cross table (Table 3) looks as follows:

	Realisieren	Other lexical verbs	Total
sich+*lassen*+Infinitive	120	2916	3036
Other structures	821	183772	184593
Total	941	186688	187629

Tab. 3: Cross table for *realisieren* in *sich*+*lassen*+Infinitive

Whether or not a lexical verb is a collexeme of a construction can be determined by its significance value, which is calculated using the Fisher-Yates exact test (FYE). This means that for verbs with an extremely low value, the null hypothesis that "the verb is equally distributed over all constructions in question" is rejected. We conducted this analysis for three of the constructions mentioned above, namely i) modal verb+*werden*+Past Participle, ii) *sein*+*zu*+Infinitive, and iii) *sich*+*lassen*+Infinitive. The results were groups of verbs with p-values of ≤ 0.05, ≤ 0.01, and ≤ 0.001. We interpret the p-values in terms of the rejection of the null hypothesis as significant (≤ 0.05), very significant (≤ 0.01), and highly significant (≤ 0.001). So, the lower the p-value, the lower the probability of a balanced or random distribution. In other words, the verb VV_1 occurs more frequently in the construction C than by chance.

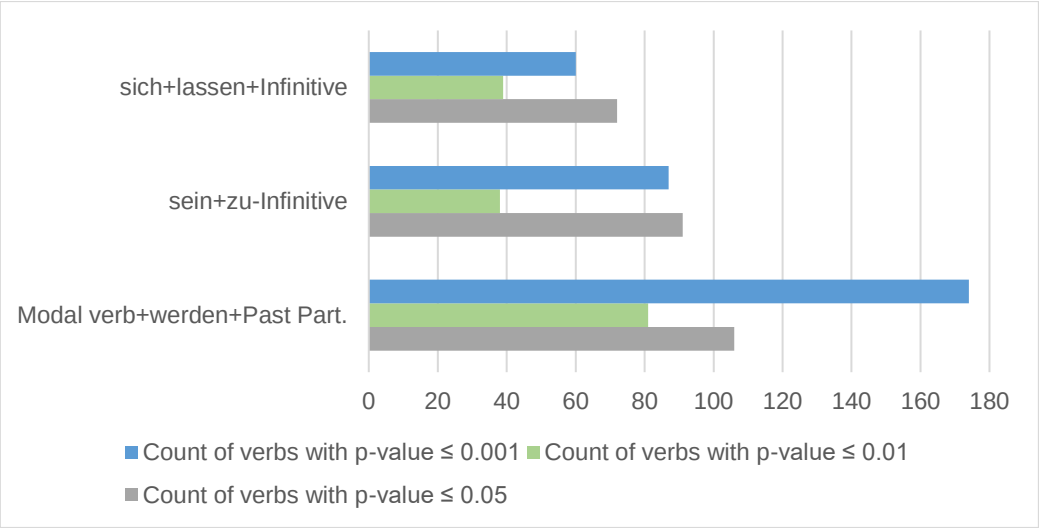

Fig. 2: Verb groups (based on significance) over the three constructions

Fig. 2 shows how many lexical verbs per group were found in the three constructions compared. Hence, we can state that in the investigated part of the language of engineering, represented by Gingko, 171 lexical verbs appear significantly more frequently than expected within the construction *sich*+*lassen*+Infinitive. We also found 216 significant verbs for the construction *sein*+*zu*-Infinitive, and 361 for the modal verb+*werden*+Past Participle.

Tab. 4 depicts these findings by listing all the lexical verbs we found for *sich*+*lassen*+Infinitive with a p-value of ≤ 0.001 (highly significant). The significance value gives an impression of the relationship between construction and lexical verbs. This means that writers prefer certain verbs in specific constructions and not in neighboring ones. The significance testing (Fig. 2 and Tab. 4) already points to a particular conclusion for the GFL/GSP teaching context: if instructors and teaching materials intend to teach a realistic use of passives and substitute forms in the language of German engineering, then transformation exercises (at least in the manner in which they are used in current instructional materials) appear less useful. We are convinced that there are reasons for the distributions we have found in Gingko thus far, and we suspect that the reasons are probably found in the interface of the formal distribution, meaning, and communicative function of verbs and constructions.

abbilden	einteilen	interpretieren	umsetzen
ableiten	erahnen	kombinieren	untergliedern
ablesen	erkennen	loesen	unterteilen
abschaetzen	erklaeren	manoevrieren	verallgemeinern
anpassen	ermitteln	nachbilden	verarbeiten
approximieren	erreichen	nachpruefen	verbessern
ausnutzen	erschliessen	nutzen	verdichten
berechnen	erweitern	realisieren	verdrehen
bestimmen	erzielen	reduzieren	verwirklichen
darstellen	festhalten	sagen	vorlesen
durchführen	festmachen	senken	weiterverwenden
einbinden	feststellen	simulieren	zurueckfuehren
einfallen	herleiten	sparen	zusammenfassen
einsparen	identifizieren	steigern	
einstellen	integrieren	umklappen	

Tab. 4: All lexical verbs with $p \leq 0.001$ (highly significant) in *sich*+*lassen*+Infinitive

What the p-values do not provide is information about the direction of associations between verbs as lexical items and grammatical constructions. One of the important questions in a GFL acquisition context would be whether specific verbs condition the use of specific constructions or specific constructions condition the use of specific verbs. The fact that both conditions are not the same is shown by Gries's (2013) study on the English collocation "of course." In his paper, he provides statistical evidence for the fact that "course" is strongly associated with "of" but "of" not as strongly with "course." In other words: "[…] *course* is a better cue to *of* than vice versa" (Gries 2013: 144).[13] A value that can show associations in both (above mentioned) directions is ΔP (Delta P). Following Gries, we think that ΔP cannot only be used to analyze collocations, but also lexical verbs in more complex grammatical constructions. Statistical analysis of the data using association measures is one of the next steps in our project – another would be to calculate the relative entropy H_{rel}. This is "a measure of uncertainty that approximates 1 as distributions become more even […] and that approximates 0 as distributions become more uneven and, thus, more predictable […]." (Gries & Ellis 2015: 8).

As the initial and expected results have thus far conformed to formal distributions, we will have to further describe at least the two missing aspects of the above-mentioned interface, a) the semantic and b) the communicative/pragmatic side of the use of the

[13] What Gries (2013) analyses for English also applies for other languages. Lehr (1998) finds that German collocations can also be directed which means that one component of a collocation might depend more on the other than vice versa.

lexical verbs in combination with the analyzed grammatical constructions, in order to derive useful didactic/pedagogical implications to improve future instruction.

6. Conclusion

Thus far, our data suggest that the traditional description of the grammatical phenomena of passives and impersonal forms in GFL instruction materials (as shown in Fig. 1) are inadequate because they do not take into account actual distributional characteristics. A precise analysis of the distributions of lexical filler verbs – at least in the three neighboring constructions we investigate – shows that native writers exhibit statistically significant preferences for the use of specific verbs in given constructions, a finding that fundamentally differs from the traditional description. Consequently, we believe that the grammatical instruction provided to learners of GFL, GCLA, and GSP needs to be modified. That is because for learners "it needs to be highlighted which meanings are most typically construed by which construction and, what the most common lexical items are in each construction" (Ellis, Römer & O'Donnell 2016: 300). In order to have an adequate description (beyond the formal aspects), a detailed account of the meaning and communicative function of filler verbs and the constructions as a whole is needed to specify the characteristics of licensing a verb as filler for a given construction. Based on that description, future materials should consider proposals such as those made by Handwerker (2008) to include information about which constructions a verb can be used in for the learner's "lexical database." A proposal for possible *chunks* provided to learners in the form of enhanced input sequences (as laid out in Handwerker & Madlener 2009) might be the best option to show and teach these forms, especially if the optimized distributional frequencies of items in the input (as shown in Madlener 2016) are taken into account to raise the possible salience of target structures for learners.

There are several positive results that extend beyond the main objectives of our investigation. First, starting in 2020, Gingko will be a publicly accessible tool for investigations of German for Specific Purposes in the subfield of automotive engineering. Second, for the design of our investigation we had to construct an investigation "pipeline" – as Alexander Ziem and Hans Boas might call it – which could serve as a standardized procedure for future investigations, not just for GSP but also for GFL in general. Third, the results of our investigation might be integrated into the *German Constructicon* framework developed by Lasch (2016) and Ziem & Boas (2017).

Above all, however, we are convinced that more adequate grammatical description is crucial for improving in-class instruction. In other words, the better the description, the better the instruction.

References

Abbot-Smith, Kirsten; Behrens, Heike (2006) How Known Constructions Influence the Acquisition of Other Constructions: The German Passive and Future Constructions. *Cognitive Science* 30, 995–1026.

Bordag, Denisa; Sieradz, Magdalena (2012) Erwerb von Perfekt und Passiv bei DaF-Lernern: Eine Korpusstudie. *GFL – German as a Foreign Language* 1/2012, 1-26.

Borgwaldt, Susanne; Sieradz, Magdalena (2018) Grammatische Eigenschaften von Fachsprachen. In: Jörg Roche; Sandra Drumm (eds.) *Berufs-, Fach- und Wissenschaftssprachen*. Tübingen: Narr, 64-71.

Buhlmann, Rosemarie; Fearns, Anneliese (2018) *Handbuch des fach- und berufsbezogenen Deutschunterrichts DaF, DaZ, CLIL*. Berlin: Frank & Timme.

Buscha, Anne; Szita Szilvia (2011) *B Grammatik. Übungsgrammatik Deutsch als Fremdsprache. Sprachniveau B1 – B2*. Leipzig: Schubert.

Chen, Shing-lung (2008) Entwicklung eines neuen Forschungsmodells für das Passiv anhand deutsch-chinesischer und chinesisch-deutscher Gebrauchsanleitungen in der Informationstechnik. *InfoDaF* 35/1, 43-60.

Christante, Valentina; Schimke, Sarah (2018) Die Verarbeitung von Passivsätzen und OVS-Sätzen im kindlichen L2-Erwerb. In: Sarah Schimke; Holger Hopp (eds.) *Sprachverarbeitung im Zweitspracherwerb*. Berlin/Boston: De Gruyter.

DAAD/HIS-Institut für Hochschulforschung (eds.) (2011) *Wissenschaft weltoffen 2011. Daten und Fakten zur Internationalität von Studium und Forschung in Deutschland*, Bielefeld. https://www.wbv.de/openaccess/themenbereiche/hochschule-und-wissenschaft/shop/detail/name/_/0/1/7004002jw/facet/7004002jw///////nb/0/category/1168.html (last visited 30.4.2018).

DZHW-Deutsches Zentrum für Hochschul- und Wissenschaftsforschung/DAAD (eds.) (2018) *Wissenschaft weltoffen Kompakt 2018. Facts and Figures on the International Nature of Study and Research in Germany*. http://www.wissenschaftweltoffen.de/kompakt/wwo2018_kompakt_en.pdf. (last visited 30.4.2018).

Ellis, Nick C.; Römer, Ute; O'Donnell, Mathew B. (2016) *Usage-based Approaches to Language Acquisition and Processing: Cognitive and Corpus Investigations of Construction Grammar*. Hoboken, NJ: Wiley-Blackwell.

FaDaF (2012) *Deutsche Sprachprüfung für den Hochschulzugang. DSH. Handbuch für Prüferinnen und Prüfer zur DSH-Erstellung und -Durchführung nach der DSH-MPO vom 17.11.2011*. Göttingen: FaDaF.

Goldberg, Adele E. (2006) *Constructions at work. The nature of generalizations in language*. Oxford: University Press.

Göpferich, Susanne (1995) *Textsorten in Naturwissenschaften und Technik. Pragmatische Typologie – Kontrastierung – Translation*. Tübingen: Narr.

Gries, Stefan Th. (2013) 50-something years of work on collocations: what is or should be next … *International Journal of Corpus Linguistics* 18/1, 137-65.

Gries, Stefan Th. (2015) More (old and new) misunderstandings of collostructional analysis: on Schmid & Küchenhoff (2013). *Cognitive Linguistics* 26/3, 505-36.

Gries, Stefan Th.; Ellis, Nick C. (2015) Statistical Measures for Usage-Based Linguistics. *Language Learning* 65 (Supplement 1), 1-28.

Gries, Stefan Th.; Stefanowitsch, Anatol (2004) Extending collostructional analysis: a corpusbased perspective on 'alternations'. *International Journal of Corpus Linguistics* 9/1, 97-129.

Handwerker, Brigitte (2008) Lernbasis Lexikon. — Das Verb und die Lizenz zu konstruieren. In: Christian Fandrych; Ingo Thonhauser (eds.) *Fertigkeiten und Kompetenzen - separiert oder integriert? Zur Neubewertung der Fertigkeiten und Kompetenzen im Fremdsprachenunterricht*. Wien: Praesens-Verlag, 35-54.

Handwerker, Brigitte; Madlener Karin (2009) *Multimedia-Chunks für Deutsch als Fremdsprache: Theoretischer Hintergrund und Prototyp einer multimedialen Lernumgebung*. Baltmannsweiler: Schneider.

Hanna, Ortrun (2003) *Wissensvermittlung durch Sprache und Bild. Sprachliche Strukturen in der ingenieurwissenschaftlichen Hochschulkommunikation*. Frankfurt a. M.: Peter Lang.

Heine, Antje (2016) "Aber die müssen doch so eine Prüfung machen" – zu sprachlichen Anforderungen und Voraussetzungen internationaler Studierender in den Ingenieurwissenschaften. In: Eglé Kontutyte; Vaiva Zeimantiene (eds.) *Sprache in der Wissenschaft*. Frankfurt a. M.: Peter Lang, 295-311.

Hinrichs, Beatrix (1999) Passivstrukturen im Spanischen und im Deutschen. Ein kontrastiver Vergleich. *InfoDaF* 26(4), 348-54.

Kaiser, Irmtraud; Peyer, Elisabeth (2011) *Grammatikalische Schwierigkeiten beim Lesen in Deutsch als Fremdsprache – eine empirische Untersuchung*. Baltmannsweiler: Schneider.

Kärchner-Ober, Renate (2018) Ingenieurwissenschaftliche Fachsprache. In: Jörg Roche; Sandra Drumm (eds.) *Berufs-, Fach- und Wissenschaftssprachen*. Tübingen: Narr, 218-27.

Krause, Thomas; Zeldes, Amir (2016) ANNIS3: A new architecture for generic corpus query and visualization. *Digital Scholarship in the Humanities* 31(1), 118-39.

Lasch, Alexander (2016) *Nonagentive Konstruktionen des Deutschen*. Berlin/Boston: De Gruyter.

Lehr, Andrea (1998) Kollokationen in Langenscheidts Großwörterbuch Deutsch als Fremdsprache. In: Herbert Ernst Wiegand (ed.) *Perspektiven der pädagogischen Lexikographie des Deutschen. Untersuchungen anhand von „Langenscheidts Großwörterbuch Deutsch als Fremdsprache"*. Tübingen: Niemeyer, 256-81.

Madlener, Karin (2016) Input optimization. In: Heike Behrens; Stefan Pfänder (eds.) *Experience Counts: Frequency Effects in Language*. Berlin/Boston: Walter de Gruyter, 133-73.

Monteiro, Maria; Rieger, Simone; Skiba, Romuald; Steinmüller, Ulrich (eds.) (1997) *Deutsch als Fremdsprache: Fachsprache im Ingenieurstudium*. Berlin: Iko-Verlag.

Raeisi Dastenaei, Rana (2017) ‚Ein Brief wird erhalten!': Passivfehler iranischer DaF-Studierender. *Zeitschrift für Interkulturellen Fremdsprachenunterricht* 22(2), 169-78.

Reznicek, Marc; Lüdeling, Anke; Krummes, Cedric; Schwantuschke, Franziska; Walter, Maik; Schmidt, Karin; Hirschmann, Hagen; Andreas, Torsten (2012) *Das Falko-Handbuch. Korpusaufbau und Annotationen Version 2.01*. https://www.linguistik.hu-

berlin.de/de/institut/professuren/korpuslinguistik/forschung/falko/FalkoHandbuchV2/view (last visited 05-22-2018).

Rheindorf, Markus (2016) Die Figurativität der allgemeinen Wissenschaftssprache des Deutschen. *Linguistik online*, 76/2, 177-95.

Stefanowitsch, Anatol; Gries, Stefan Th. (2003) Collostructions: Investigating the interaction of words and constructions. *International Journal of Corpus Linguistics* 8/2, 209-43.

Stefanowitsch, Anatol (2009) Bedeutung und Gebrauch in der Konstruktionsgrammatik. Wie kompositionell sind modale Infinitive im Deutschen? *Zeitschrift für Germanistische Linguistik* 37/3, 565-92.

Steinhoff, Annette (2011) *Lernmedien Deutsch als Fremdsprache und die Vermittlung des Passivs*. München: Iudicium.

Steinmetz, Maria; Dintera, Heiner (2014) *Deutsch für Ingenieure. Ein DaF-Lehrwerk für Studierende ingenieurwissenschaftlicher Fächer*. Wiesbaden: Springer Vieweg.

Ziem, Alexander (2014) Von der Kasusgrammatik zum FrameNet: Frames, Konstruktionen und die Idee eines Konstruktikons. In: Alexander Lasch; Alexander Ziem (eds.) *Grammatik als Netzwerk von Konstruktionen. Sprachwissen im Fokus der Konstruktionsgrammatik*. Berlin: De Gruyter, 263-90.

Ziem, Alexander; Hans C. Boas (2017) Towards a Construction for German. In: *Proceedings of the AAAI 2017 Spring Symposium on Computational Construction Grammar and Natural Language Understanding*. Technical Report SS-17-02, Stanford University, 274-77.

TU9 (2006) TU9 German Institutes of Technology Website. https://www.tu9.de/en/index.php (last visited 10.7.2018).

Institut für Automobiltechnik Dresden – IAD (2016) Zur Sprachkompetenz internationaler Studierender. https://tu-dresden.de/bu/verkehr/iad/lvm/studium/sprachkompetenz (last visited 10.7.2018).

Author Biographies

Antje Heine is Junior Professor for German as a Foreign Language at the Department of German Philology at Greifswald University, Germany. Before coming to Greifswald, she was researcher at the Universities in Wuppertal, Helsinki and Leipzig. Her research interests include phraseology, Construction Grammar, lexicography, and corpus linguistics (for GFL).

Lars Schirrmeister is Assistant Researcher and teaching faculty member at the Department of German Philology at Greifswald University, Germany. His main research interests are Construction Grammar and its application to SLA, grammatical description for SLA and GFL, non-agentive Constructions of German, and grammatical gender acquisition in GFL.

Matthew O. Anderson is a doctoral candidate and Assistant Instructor of German in the Department of Germanic Studies at the University of Texas at Austin. His dissertation explores the intermedial and international nexes of visuality, genre, and reading in historical German youth adventure literature, but he has also presented on L2 reading activities in the German language classroom.

Key words

Corpus Linguistics, Construction Grammar, German for Specific Purposes, German language of engineering, German passive, collexeme analysis.