

## Frequency and structures of lexical bundles in the German Goethe-Institut website

Amalia Qurrota Ayuni<sup>1\*</sup>, Susi Yuliawati<sup>1</sup>, Dian Ekawati<sup>1</sup>

<sup>1</sup>Universitas Padjadjaran

\*Corresponding Author. Email: [amalia.ayuni@gmail.com](mailto:amalia.ayuni@gmail.com)

### Abstract

Corpus linguistics allows researchers to discover the nature of language use through lexical bundles throughout genres, registers, and language varieties. With the fast-changing development of the internet, the language of websites has become one fascinating variety to investigate. In the contexts of German language studies, Goethe-Institut is a worldwide German cultural institution dedicated to teaching German and propagating German culture, with its website becoming a well-known source for German-related studies. Under these considerations, this research is interested in analyzing the German language patterns in Goethe-Institut website by examining their frequency and structure of lexical bundles. Using a mixed-method approach, the corpus was found to be dominated by lexical bundles within the ranges of three- and four-bundles, with the least quantity of lexical bundles in the range of five. The majority of the four-word lexical bundles on this site fell into the categories of noun, preposition, or verb groups. Meanwhile, the adverb, conjunction, and adjective groups were the fewest to appear in the four-word lexical bundles. The language in the Goethe-Institut was shown to contain semi-formal expressions according to the frequency of use of the prepositions, nouns, verbs, and the active sentence expressions. The utilization of standardized German language and basic vocabulary indicates that this website is designed to be accessible for everyone, including German language learners. The language usage also demonstrates that the Goethe-Institut is especially a user-oriented website with expressions that evoke a 'sense of belonging'.

### Keywords

lexical bundles;  
corpus linguistics;  
website; Goethe-  
Institut

This is an open access article under the [CC-BY-SA](https://creativecommons.org/licenses/by-sa/4.0/) license.



**Citation (APA Style):** Ayuni, A. Q., Yuliawati, S., & Ekawati, D. (2022). Frequency and structures of lexical bundles in the German Goethe-Institut website. *LingTera*, 9(2), 1–12. <https://doi.org/10.21831/lt.v9i2.54523>

## INTRODUCTION

Technological developments in the late 20<sup>th</sup> century had accelerated further with the development of the internet in the 1970s which resulted in cultural changes that shaped today's world. Quoted from Jimenez-Crespo (2021, p. 375), technological developments have a significant impact on modern cultures which is marked by the development of computer technology in the 1980s, the invention of the World Wide Web (WWW) in the 1990s, and the social networking revolution in the early 2000s. Digital technology enables new types of text such as websites. The web is a medium written in HTML (Hyper Text Markup Language), which can display various types of documents such as text, audio, video, images, animation, and multimedia data. Websites are easily accessible via the internet on devices such as mobile phone, laptop, and others. Websites contain large amounts of data, such as texts, pictures, videos, and so on. The verbal component itself has its own characteristics. These characteristics can form language patterns from underlying factors such as cultures, registers, and genres. Websites have various genres with their respective functions such as portal, academic, business, blog, informational website, and so on (Shelly, 2010, p. 67-70).

The language of a website can be an interesting object of research. Large amount of textual data on a website can be investigated in terms of its structures, linguistic forms, and functions by employing corpus study. Corpus linguistics can be defined by its parts, corpus and linguistics. The term ‘corpus’ means data. *Das Korpus* in German can be defined as *eine Sammlung schriftlicher oder gesprochener Äußerungen* or ‘a collection of written and spoken explanations’ (Duden, 2022). Generally, the data in a corpus have gone through digitization process so that it can be read by computers and software. Corpus linguistics can be defined as natural language utterances, their elements and structures, and a theory building based on the analysis of authentic texts, which are summarized in the corpus (Lemnitzer & Zinsmeister, 2015). Similar to this understanding, Cheng (2011) defines corpus linguistics as a study in the field of linguistics that aims to construct the use of language for specific purposes which can produce a theory about a language based on the actual use of it. To conclude, corpus linguistics focused on constructing language uses in a discourse using digitalization process.

The phenomenon of language use in various contexts can be described using corpus approach as an effort to conduct scientific studies. Linguistic studies that focus on various micro- and macro-fields such as semantics, pragmatics, syntax, discourse analysis, and so on can use corpus approach to get a broader view in processing language data. Corpus linguistics can be used to find out how a language is used in a particular situation and find patterns in it. Studies of existing lexical collections involve corpus linguistic approach that deals with large amounts of electronic data and computer tools to investigate the consistent use of language in research articles.

To find language patterns in a discourse, researchers can examine the use of multi-word sequences formed on certain criteria. In contrast to the traditional perspective which regards word learning as the basic building block in language learning and processing, there is empirical evidence in recent study which shows that multi-word sequences can be considered an integral building block in a language (Yuliawati et al., 2021). Multi-word sequences can show the characteristics of the language in a discourse by looking at their frequency because they show the amount to which a sequence of words is saved and utilized as a prefabricated chunk (Biber et al., 2004, p. 37, in Fuster-Márquez, 2017). Several terms that have a similar concept to multi-word sequences are lexicalized sentence stems (Pawley & Syder, 1983), lexical bundles (Scott, 2001), phraseological sequences, phrasing, chunks, lexical bundles (Biber et al., 1999, p. 989) and N-grams (Stubbs, 2005).

Formulaic language is a multi-word sequence found in a particular discourse with a certain frequency that has meaning and function while lexical bundles are multi-word units that occur often in a register and are defined as bundles of words that indicate a statistical probability to co-occur (Biber et al., 1999, p. 989; Biber et al., 2004). Generally, studies on the use of lexical bundles are carried out in English language discourses. However, in recent years, research has been conducted on lexical bundles in other languages such as Bahasa Indonesia (Febriesti, 2022; Budiwiyanto & Suhardijanto, 2020; Yuliawati et al., 2021; Haq, 2021); Korean (Kim, 2009); Spanish (Butler, 1998; Cortes, 2007; Tracy-Ventura et al., 2007); and Germany (Krekeler, 2020).

Pattern analysis of lexical bundles usage is generally carried out to determine writer's style in the field of academic writing or academic research because it can be used to determine the pattern of language use in teaching, those produced by students, teachers, or even researchers. Examples of studies in lexical bundles analysis in research of academic writing can be found in articles by Yuliawati et al. (2022) or in classroom discourse like Biber et al. (2004). Even so, researchers can analyze patterns of language use in every discourse including regulations (i.e., law) and legal letters (Putri & Suhardijanto, 2022); hotel websites (Fuster-Márquez, 2014); European Union legal text and website (Biel et al., 2019); and academic institution websites (Bernardini, et al., 2010). The use of websites as corpus has become frequent practice in the process of data collection for corpus linguistic studies. This approach is called ‘web-as-corpus’ or the creation and utilization of corpus that can be used for a variety of research goals, including terminology extraction (Castagnoli, 2006) and register analysis (Biber & Kurjian, 2007). The dynamic nature of a website and simplicity of its customization according to the website users’ needs can be used as a research object for language patterns.

Until recently, no research has been conducted that focus on lexical bundles of German websites especially concerning academic websites. Therefore, the current research seeks to address

this gap by analyzing the usage frequencies and structural forms of lexical bundles in one of the academic websites of Germany in order to identify underlying similarities and variations in the frequency and structures of the multi-word phrases used in the discourse. The German version of the Goethe-Institut website was used as the research object in this study. The Goethe-Institut is one of Germany cultural institutions which utilizes a website as their informational resource.

The Goethe-Institut is a non-profit multinational organization and German language educational institution with representations in 98 countries. To expand its impact, Goethe-Institut has localized its website into numerous languages, notably French, Turkish, Korean, and Bahasa Indonesia to achieve higher accessibility. As a cultural institution, Goethe-Institut focuses on teaching German language across the world by providing courses, certifications, and so on. Goethe-Institut's website is suitable to be used as a corpus because Goethe-Institut is one of the German institutions that has strong influence and is popular around the world, including in Indonesia. This website can reflect how German language is used by website owners to communicate and inform their visitors who mainly are German language learners. The researcher examined how lexical bundles are used in this website based on its frequency and morphosyntactic forms according to a theory from Brommer (2018).

## RESEARCH METHOD

This research utilized a mixed-method approach which began with statistical data analysis of the corpus followed by qualitative interpretation. According to Greene et al. (1989, as cited in Creswell, 2011), mixed-method research consists of at least one quantitative component (in this case word collection) and one qualitative component (in this case data interpretation). The approach is used to examine statistical data from the corpus, which are then qualitatively interpreted (Kwary & Arum, 2011, p. 213; Yuliawati, 2014, p. 112). As a result, statistical findings can help researchers to analyze the data. This research utilized corpus-based approach (inductive) to find out the research objectives based on the former theories. The data were collected through the process of corpus building using Sketch Engine on September 26–27, 2022. Besides its function to analyze corpus linguistic data, Sketch Engine can also be utilized as a corpus maker by deploying WebBootCaT (a web-service version of the BootCaT tools). Therefore, researchers can create a corpus and analyze their data in a single software. After being collected, there are 780,435 tokens and 659,371 words in the corpus of German language version of the Goethe-Institut.

The N-gram feature of Sketch Engine was used to seek lexical bundles in the ranges of 3, 4, and 6 words. The researcher determined a minimum frequency to avoid lexical bundles that are written and potentially involve idiosyncrasies of a particular writer(s). For written corpora, the minimum frequency to consider can range from 10 to 40 per million words (Biber et al., 2004; Hyland, 2008). Nevertheless, since Sketch Engine could only extract data with round frequencies, the researcher rounded the normalized frequencies whether they're up or down, as the one conducted by Chen and Baker (2010). The researcher compared lexical bundles in the range of 3 and 4 words after the data had been successfully extracted by Sketch Engine. To analyze the structure, the researcher put emphasize on four-word lexical bundles and took samples by selecting the top 50 sets of data from the lexical bundles with the highest frequency. The structure of lexical bundles in the Goethe-Institut website corpus is examined using a theory by Brommer (2018) who divided morphosyntactic forms of pattern into seven groups, namely *nominalgruppe* (NGr), *verbalgruppe* (VGr), *adjectivgruppe* (AdjGr), *partizipgruppe* (PtGr), *adverbialgruppe* (AdvGr), *konjunkionalgruppe* (KonGr), and *prepositional group* (PrGr). Then, the researcher evaluated the data qualitatively to discover the patterns of the lexical bundles.

## FINDINGS AND DISCUSSION

In this study, the lexical bundles were examined based on its frequency as this study followed a method by Biber et al. (1999) to examine the lexical bundles which is solely based on frequency. As a research object, the Goethe-Institut website corpus contains 916 documents, 780,435 tokens, and 659,371 words. This research employed frequency limit of occurrence with a normalized frequency to identify lexical clusters in the corpus. Since corpus linguistic research is conservative, it is

necessary to calculate the frequency of occurrence of lexical bundles ranging from 10, 20 to 40 times per million words in five distinct texts. The calculation is illustrated below:

$$20/1,000,000 \times 780,435 = 15.6087$$

Since Sketch Engine can only extract data with integer frequencies, the normalized frequency constant is set to the closest number by rounding, in this case, sixteen (16). The rounded-up frequency is known as the normalized frequency (Chen & Baker, 2010). The total number of lexical bundles in the corpus with the minimum normalization frequency can be seen in Table 1.

**Table 1. Distributions of Lexical Bundles in 3,4,5,6 ranges**

No.	Lexical Bundles	Total Data	Total Normalized Frequency	% of Normalized Frequency
1.	3-words	516	13,556	74.22%
2.	4-words	120	2,564	14.04%
3.	5-words	65	1,289	7.06%
4.	6-words	44	856	4.69%
	Total	745	18,265	100.00%

With the majority of data (516), three-words lexical bundles are the most frequent to occur. The amount of three-words lexical bundles differ significantly, approximately four times than that of the four-words lexical bundles which only consists of 120 data. Then, there are the five- and six-words lexical bundles which have modest percentage difference of 2.37%. According to prior study by Putri and Suhardijanto (2022), the fewer the words that make up a lexical bundle, the higher their level of usage are in a corpus. To discover about lexical bundles further in the corpus, see the table in Appendix 2 which displays the top 50 lexical bundles in three- and four-words ranges.

Appendix 2 table shows that three-words lexical bundles are mainly composed of three words strings such as “*Kinder und Jugendliche*” (Child and Teens), “*auch in der...*” (also in the...), and “*Leben in Deutschland*” (live in Germany). These findings are similar with a study of lexical bundles in academic English and Bahasa Indonesia by Yuliawati et al. (2022). When the 50 data with the highest frequency from each word range are compared, it can be seen that the most frequently used words in the three- and four-words lexical bundles are ‘*Deutsch*’, ‘*Fremdsprache*’, and ‘*Welt*’. Overlapping lexical bundles such as “*für Deutsch als Fremdsprache*” (for German as foreign language), “*und Deutsch als Fremdsprache*” (and German as foreign language), and “*Bereich Deutsch als Fremdsprache*” (Area of German as foreign language) also occurs several times in the corpus. These findings indicate that the Goethe-Institut are focused on the topic of German as foreign language.

The lexical bundles in Goethe-Institut website can also be examined by its structure. The morphosyntactic form of patterns from Brommer (2018) was originally adopted from the theory of syntactic categories by Eisenberg (2014). This theory can be found in the book *Sprachliche Muster*. There are some adjustments made by Brommer. In particular, it only takes complex constituent categories such as *Nominalgruppe* (NGr), *Verbalgruppe* (VGr), *Präpositionalgruppe* (PrGr), and does not include word class or part-of-speech categories. Brommer divided morphosyntactic or structural form of patterns into seven types namely *Nominalgruppe* (NGr) or ‘Noun group’, *Verbalgruppe* (VGr) or ‘verbal group’, *Adjektivgruppe* (AdjGr) or ‘adjective group’, *Partizipgruppe* (PtGr) or ‘participle group’, *Adverbgruppe* (AdvGr) or ‘adverb group’, *Konjunktionalgruppe* (KonGr) or ‘conjunctive group’, and *Präpositionalgruppe* (PrGr) or ‘prepositional group’. When compared to lexical bundles with a span of three and five words, the structure of four-word word clusters can provide a more precise and efficient interpretation of structure and function for analysis (Hyland, 2008). Therefore, the analysis of lexical bundles structure in the Goethe-Institut corpus only focused on the four-words lexical bundles.

Based on table in Appendix 2 that contains lexical bundles and their structural forms, it can be seen that the frequency of occurrences of lexical bundles from *Nominagruppen* (38%) and

*Präpositionalgruppen* (30%) were the most commonly found categories in the corpus. Meanwhile, the conjunction, adverb, and adjective group are the structure form with the least usage. In addition, the researcher did not find any lexical bundles of the *Partizipgruppe* structure. To understand the structure of lexical bundles in the corpus of the Goethe-Institut website, the researcher interpreted the structure of some lexical bundles data in each structural form.

Most of the lexical bundles that fall into the *Präpositionalgruppen* category begin with the preposition *für*, *von*, *in*, *auf*, and *aus*. Generally, the preposition is followed by a noun or adverb. The lexical bundles “*für Deutsch als Fremdsprache*” (for German as a foreign language) is the four-words lexical bundle with the highest frequency in the corpus. This bundle begins with the preposition *für* (for) and continues with the noun *Deutsch*, conjunctions *als* (as) and *Fremdsprache* (foreign language). The prepositions themselves can be defined as additional words belonging to a word class called *Funktionswörter* (task words) (Helbig & Buscha, 1996, p. 401, in Khairani, 2013). Therefore, the structure of this lexical bundle is included in *Präpositionalgruppen* because the prepositions act as the core of the bundle. This can be seen in the sample text which shows that the bundle “*für Deutsch als Fremdsprache*” is generally located in the middle of the sentence to explain the previous statement.

In the lexical bundle “*im Bereich Deutsch als*” (in the field of German as), there is a preposition *im* which is a combination of the preposition *in* with the noun *der Bereich* and affected with the dative case. The word *im* serves to explain the previous statement which includes dealing with the ‘German field as a thing’. In the taken sample text, the entire lexical bundles ends with the word *fremdsprache* or foreign language. So, ‘German field’ in this bundle can be interpreted as the German field as a foreign language. Therefore, this data can be categorized in a prepositional group that is followed by a noun.

Furthermore, the *Nominalgruppen* category in the corpus mostly corresponds to the institution name Goethe-Institut, *Deutschland* (German), *sprache* (language), and Language (language) in the word cluster “*der Zentrale des Goethe-Instituts*”. It means “the main center/office of the Goethe-Institut” which came from the full text “*Bei allen anderen Fragen steht Ihnen der Fachbereich Bildende Kunst in der Zentrale des Goethe-Instituts gerne zur Verfügung*” which means “If you have any other questions, the Department of Fine Arts at the Goethe-Institut headquarters will be happy to help.”. It can be seen that the bundle can be followed either before or after the preposition *in*. The bundle is used to refer to an entity, namely the Goethe-Institut. Another *Nominalgruppen* category, namely “*Mein Weg nach Deutschland*” (My Way to Germany) is one of the names for the rubric or section on the website that discusses information about life in Germany. As such, this bundle has a fairly high frequency. The structure of the lexical bundle “*Deutschland Mein Weg nach*” consists of the pronoun *mein*, noun *Weg*, adverbial *nach*, and noun *Deutschland*.

Conjunction group or *Konjunktionsgruppe* can be found in lexical bundle “*und kann jederzeit widerrufen*” (and can be revoked at any time) as shown in the sample concordance “*Die Anmeldung ist freiwillig und kann jederzeit widerrufen werden*” (Registration is voluntary and can be revoked at any time). This lexical bundle can be included in the ‘boilerplate’ or standard text strings that are often found on websites. An example of boilerplate text is a legal document on a website that contains privacy settings for the website visitor. The conjunction ‘*und*’ (and) is followed by modal verb ‘*kann*’ (can), *jederzeit* (any time), and verb *widerrufen* (revoke). *Konjunktionsgruppe* can also be found in lexical bundle “*dass meine E-Mail-Adresse für*” (that my email address for). The conjunction ‘*und*’ falls into the conjunction type as connectors, namely *koordinierte Konjugationen*.

Similarly, the following lexical bundle is also a boilerplate text, i.e., in “*Mit meiner Anmeldung stimme ich zu, dass meine E-Mail-Adresse für die Zusendung des abonnierten E-Mail-Newsletter verwendet wird*” (with my registration, I agree that my e-mail address will be used to send the subscribed email newsletter). The conjunction *dass* usually connects subordinate clauses with main clauses in which verbs such as *behaupten* (assert), *bestätigen* (confirm), *denken* (thinking), *glauben* (believe), and *hoffen* (hope) (Duden, 2022). In contrast with the previous lexical bundle in *Konjunktionsgruppe*, the conjunction *dass* is *subordinierende Konjunktionen* or conjunctions as subordinate. Along with this theory, this lexical bundle is also used to connect between

subordinate clause as in “*meine E-Mail-Adresse für...*” and main clause “*Mit meiner Anmeldung stimme ich...*”

*Verbengruppe* in the top 50 lexical bundles from Goethe-Institut website mostly begin with verbs and followed by pronouns. In the lexical bundle “*Haben Sie gut aufgepasst*” (have you pay attention) and “*können Sie anschließend selbst*” (then you can do it yourself), the writer used modal verb followed by pronoun *Sie* (salutation to one or more person in formal form). The difference between these two lexical bundles is that the first lexical bundle have a structure of verb + pronoun + adjective + verb in *perfekt* form, while the second lexical bundle is formed with modal verb + pronoun + adverb + adverb.

Another form of *Verbengruppe* can be found in the lexical bundles “*kann jederzeit widerrufen werden*” (can be revoked at any time) and “*ist freiwillig und kann*” (is optional and can). The lexical bundle “*kann jederzeit widerrufen werden*” has a structure of modal verb + adverb + verb + verb. Different from the other data, the lexical bundle “*ist freiwillig und kann*” consists of verb + adjective + conjunction + modal verb. This data positioned its modal verb in the end of the bundle.

At last, the least used structural forms of lexical bundles, namely the adjective group can be seen in the lexical bundle “*freiwillig und kann jederzeit*” (voluntarily and can at any time). This lexical bundle is also a boilerplate text used in the login page of the Goethe-Institut website. The full sentence containing this lexical bundle is “*Die Anmeldung ist **freiwillig und kann jederzeit** widerrufen werden*” (Registration is voluntary and can be revoked at any time). The structure of this lexical bundles is adjective + conjunction+ modal verb+ adverb.

In the Goethe-Institut website, the researcher rarely found passive expressions. Constructions of passive structure with temporal tenses *perfekt* (perfect), *plusquamperfekt* (past continuous), and *futur* (future) such as “*wird davon ausgegangen das*” (it is assumed that), and “*wie in bild dargestellt*” (as shown in picture) were not found in this corpus. The verbs used by the writer(s) were mainly the *vollverben*, which can form a predicate on its own and be conjugated along with the subject or pronouns in the sentence. There were only ordinary prepositions that function as positional markers or connectors, such as *auf*, *über*, *in*, and *für*. There were no other types of prepositions such as *postpositionen* (*halber*, *hinaus*), *zirkumpositionen* (*um... herum*), and *ambipositionen* (*wegen*, *nach*). The use of prepositions is very often found in groups of words other than prepositional groups, such as noun groups and conjugation groups. Lastly, the pronoun *meine* (mine) was utilized in several lexical bundles, it was included in the possessive pronoun, which indicates possession. The website aims to give its users ‘a sense of belonging’ and to encourage or make them comfortable by being a member of the Goethe-Institut.

There are several differences and similarities compared to the previous research from Krakeler (2018) who investigated lexical bundles in dissertations of machine engineering majors in Germany. In the study, he combined the types of noun groups with preposition group into a new type namely *Nominal-und Präpositionalgruppen*. These two types of bundles are combined because in German language prepositions are frequently used with the noun phrases. As a result, nouns can be found in prepositional groups in both the middle and final positions, and vice versa. In contrast to the previous research, the researcher in this study decided not to merge the two groups to learn about the features and structure of the noun and prepositional groups. Similar to the current study, as the percentage shown in Table 3, the *Nominalgruppen and Präpositionalgruppen* are the most common structure type in Goethe-Institut website.

This phenomenon can be explained by writer’s tendency to use certain lexical bundles as connectors between one explanation and another or as an explanation for the previous statement. If the lexical bundles are positioned in front of the lexical bundles, they are usually used to clarify the previous sentences. As shown from the concordance lines “*Begriffe und Redewendungen parat haben und ins Langzeitgedächtnis abspeichern können. Haben Sie gut aufgepasst? Hier finden Sie ein abwechslungsreiches Arbeitsblatt zu unserem Podcast.*” (Have you paid close attention? Here is a varied worksheet for our podcast.) which have terms and phrases at hand and be able to save them in long-term memory. The lexical bundle “*hier finden sie ein*” (here you find a) still connect to the previous explanation about self -language learning instruction using podcast from Goethe-Institut.

This function can also be found in lexical bundles with the highest occurrences in the machine engineering corpus, namely “*in bezug auf die*” (in relation to the). This lexical bundle is also classified as a *Präpositionalgruppe* and can be used to refer to the previous statement. Another example of lexical bundle which have similar functions is “*für Deutsch als Fremdsprache*” (for German as foreign language).

In the mechanical engineering dissertation corpus, the use of impersonal expressions is generally accompanied by the use of technical language. In German, it is called *Fachsprache* or the language used in a particular field. They also use more passive sentences, which makes their writing more objective and formal. Meanwhile, the Goethe-Institut website usually uses a semi-formal expression with standardized German. The use of casual language can be seen in the tendency to use active expressions and pronouns such as *meine* which indicates the subject of the sentence. Since the vocabulary used in the top 50 lexical bundles is based on standard German, it can be argued that it is easy for German language learners to understand. The majority of the words in the top 50 lexical bundles are related to education and are found on pages that provide information and news about language learning. Reflecting the organization's work as a cultural and language institution, the language in the Goethe-Institut website is designed to be user-centered. German language learner or anybody who wants to know more about Germany and has basic German proficiency will most likely be able to successfully understand and explore this website while also learning more German language through its contents.

## CONCLUSION

Research on lexical bundles can help the development of language studies by revealing patterns of language used in the community. Through this research which explored the Goethe Institut website, it can be seen that this cultural institution (and also the center for German language learning or *Deutsch Als Fremdsprache*) reflects the use of standard German that can be understood by the users. Several lexical bundles that have frequent occurrences are used in similar discourse. Therefore, the use of this repeated bundles can make it easier for users (learners) to familiarize themselves with the language. The frequency of lexical bundles in the corpus is dominated by three- and four-words lexical bundles. Regarding the structure of four-words lexical bundles, the top 50 lexical bundles in the corpus were mostly divided into noun groups, prepositional groups, and verb groups and are mainly functioned to clarify previous statement.

The standard German language and the vocabulary usage in this website can be potentially understood by everyone, especially German language learners who want to know more about Goethe-Institut. When compared to prior studies of different research object, this study shows a significant difference particularly in the expressions employed, such as passive-active forms, impersonal and semi-formal expressions, and the usage of technical languages. As a contribution to corpus linguistics studies, this research can be utilized as a recommendation for website content writers, particularly for German language-based websites, specifically in academic genres such as the Goethe-Institut. This study could be used by translators of websites from German to Bahasa Indonesia to design user-focused language styles, as is done by the Goethe-Institut. Furthermore, while this research is limited to sampled data consisting of 50 lexical bundles, it is very likely to be continued and further developed.

## REFERENCES

- Bernardini, S., Ferraresi, A., & Gaspari, F. (2010). Institutional Academic English in the European context: A web-as-corpus approach to comparing native and non-native language. In *Professional English in the European Context: The EHEA Challenge* (pp. 27–53).
- Biber, D., Conrad, S., & Cortes, V. (2004). If you look at ...: *Lexical bundles* in university teaching and textbooks. *Applied Linguistics*, 25(3), 371–405.  
<https://doi.org/10.1093/APPLIN/25.3.371>



- Biber, D., Johansson, S., Leech, G., Conrad, S., & Finegan, E. (1999). *Longman grammar of spoken and written English*. Harlow: Longman. 109–132.
- Biel, Ł., Koźbiał, D., & Wasilewska, K. (2019). The formulaicity of translations across EU institutional genres: A corpus-driven analysis of lexical bundles in translated and non-translated language. *Translation Spaces*, 8(1), 67–92. <https://doi.org/10.1075/TS.00013.BIE>
- Brommer, S. (2018). *Sprachliche muster*. de Gruyter.
- Budiwiyanto, A., & Suhardijanto, T. (2020). Indonesian lexical bundles in research articles: Frequency, structure, and function. *Indonesian Journal of Applied Linguistics*, 10(2), 292–303. <https://doi.org/10.17509/IJAL.V10I2.28592>
- Butler, C. S. (1998). Collocational frameworks in Spanish. *International Journal of Corpus Linguistics*, 3, 1-32.
- Castagnoli, S. (2006). Using the web as a source of LSP corpora in the terminology classroom. In Baroni, M. & Bernardini, S. (eds.) 159–172.
- Cortes, V. (2004). Lexical bundles in published and student disciplinary writing: Examples from history and biology. *English for specific purposes*, 23, 397–423.
- Chen, Y. H., & Baker, P. (2010). Lexical bundles in L1 and L2 academic writing. *Language Learning & Technology*, 14, 30–49.
- Cheng, W. (2011). *Exploring corpus linguistics*. Routledge.
- Cortes, V. (2007). A comparative analysis of lexical bundles in academic history writing in English and Spanish. *Corpora*, 3(1), 43–57.
- Creswell, J. W. & Vicki L. P. C. (2011). *Designing and conducting mixed methods research* (3rd ed.). SAGE Publication.
- Duden. (2022, November 10). *Rechtschreibung, Bedeutung, Definition, Herkunft*. <https://www.duden.de/>
- Eisenberg, P. (2014). Grundriss der deutschen Grammatik. In *Grundriss der deutschen Grammatik*. J.B. Metzler. <https://doi.org/10.1007/978-3-476-00743-8>
- Márquez, M. F. (2014). Lexical bundles and phrase frames in the language of hotel websites. *English Text Construction*, 7(1), 84–121. <https://doi.org/10.1075/ETC.7.1.04FUS/CITE/REFWORKS>
- Goethe Institut. (2022, September 26). *Sprache. Kultur. Deutschland*. <https://www.goethe.de/de/index.html>
- Helbig, G. & Buscha J. (1996). *Deutsche grammatik ein handbuch für den ausländer*. Langenscheidt.
- Hyland, K. (2008). As can be seen: Lexical bundles and disciplinary variation. *English for Specific Purposes*, 27(1), 4–21. <https://doi.org/10.1016/J.ESP.2007.06.001>
- Jiménez-Crespo, M. A. (2020). *Localization*. The Routledge Handbook of Translation and Globalization, 375–390. <https://doi.org/10.4324/9781003121848-31>



- Khairani. (2013). *Preposisi bahasa Jerman mit dan bei dalam majalah nadi (2009) dan padanannya dalam bahasa Indonesia*. Universitas Negeri Yogyakarta.
- Kim, Y. J. (2009). Korean lexical bundles in conversation and academic texts. *Corpora*, 4(2), 135–165. <https://doi.org/10.3366/E1749503209000288>
- Krekeler, C. (2020). Schreiben im studium: Eine korpuslinguistische untersuchung zum sprachgebrauch in den studiengängen maschinenbau und betriebswirtschaft. *Zeitschrift Für Interkulturellen Fremdsprachenunterricht*, 25(2). <https://zif.tu-journals.ulb-tu-darmstadt.de/article/id/3303/>
- Kwary, D.A., & Arum, K.W.A. (2011). Lincoln’s vs. Obama’s presidencies: Diachronic corpus based analysis of the adjectival collocates of [man] and [woman] in the American English. *ReVEL*, 9 (17), 211-225.
- Lemmitzer, L., & Zinsmeister, H. (2015). *Korpuslinguistik* (3rd ed.). Gunter Narr Verlag. <https://www.narr.de/korpuslinguistik-16886-2/>
- Pawley, A., & Syder, F. H. (1983). Two puzzles for linguistic theory: Nativelike selection and nativelike fluency. *Language and Communication*, 191–226. <https://doi.org/10.4324/9781315836027-12>
- Putri, M. T. D. (2022). Struktur dan fungsi gugus leksikal dalam teks peraturan perundang-undangan, surat resmi, dan surat bisnis (Structure and function of lexical bundles in legal texts, formal letters, and business letters). *Kandai*, 18(1), 1–21. <https://doi.org/10.26499/JK.V18I1.3322>
- Scott, M. (2001). Comparing corpora and identifying key words, collocations, and frequency distributions through the Word Smith Tools suite of computer programs. In *Small corpus studies and ELT*. John Benjamins Publishing Company. <https://doi.org/10.1075/sc1.5.07sc>
- Shelly, G. B., & Misty, E. V. (2010) *Discovering computers 2010*. Cengage Learning.
- Stubbs, M. (2005). *The most natural thing in the world: Quantitative data on multi-word sequences in English*. Paper Presented at Phraseology.
- Tracy-Ventura, N., Cortes, V., & Biber, D. (2007). Lexical bundles in Spanish speech and writing. In G. Parodi (Ed.), *Working with Spanish corpora* (pp. 354–375). Continuum.
- Yuliawati, S. (2014). Analisis berbasis korpus: Kolokasi kata-kata bermakna “Perempuan” dalam media sunda (majalah manglé, 2012 – 2013). *Ranah: Jurnal Kajian Bahasa*, 3(2), 107–123. <https://doi.org/10.26499/RNH.V3I2.42>
- Yuliawati, S., Ekawati, D., & Mawarrani, R. E. (2021). Investigating lexical bundles in the corpora of English and Indonesian research articles with the sketch engine. *Jurnal Sosioteknologi*, 20(2), 188–200. <https://doi.org/10.5614/SOSTEK.ITBJ.2021.20.2.5>

## Appendix 1. Distributions of Lexical Bundles in 3 and 4 range

No.	3- Words	Normalized Frequency	4- Words	Normalized Frequency
1	<i>Deutsch als Fremdsprache</i>	279	<i>für Deutsch als Fremdsprache</i>	75
2	<i>der deutschen Sprache</i>	125	<i>von Deutsch als Fremdsprache</i>	44
3	<i>Schülerinnen und Schüler</i>	111	<i>im Bereich Deutsch als</i>	35
4	<i>Hier finden Sie</i>	109	<i>aus der ganzen Welt</i>	35
5	<i>an der Universität</i>	107	<i>in den letzten Jahren</i>	32
6	<i>für Deutsch als</i>	97	<i>der Zentrale des Goethe-Instituts</i>	31
7	<i>in Deutschland und</i>	81	<i>Mein Weg nach Deutschland</i>	31
8	<i>die deutsche Sprache</i>	80	<i>Content and Language Integrated</i>	30
9	<i>aus aller Welt</i>	79	<i>and Language Integrated Learning</i>	30
10	<i>in den USA</i>	77	<i>auf der ganzen Welt</i>	29
11	<i>der ganzen Welt</i>	75	<i>Deutsch als Fremdsprache und</i>	29
12	<i>im Bereich der</i>	72	<i>die Schülerinnen und Schüler</i>	27
13	<i>der deutschen Minderheit</i>	69	<i>in der Zentrale des</i>	27
14	<i>in der Regel</i>	69	<i>Bereich Deutsch als Fremdsprache</i>	27
15	<i>The Big Pond</i>	68	<i>Sie für die Bearbeitung</i>	25
16	<i>im Rahmen der</i>	66	<i>Mit dem Lösungsblatt können</i>	24
17	<i>und in der</i>	61	<i>dem Lösungsblatt können Sie</i>	24
18	<i>in den letzten</i>	60	<i>selbst die Korrektur vornehmen</i>	24
19	<i>in Bezug auf</i>	59	<i>Rechnen Sie für die</i>	24
20	<i>in Zusammenarbeit mit</i>	58	<i>ein Lernerfolg ist somit</i>	24
21	<i>Deutsch als Fremd</i>	55	<i>anschließend selbst die Korrektur</i>	24
22	<i>wie z. B.</i>	54	<i>und Deutsch als Zweitsprache</i>	24
23	<i>im Rahmen des</i>	54	<i>Sie anschließend selbst die</i>	24
24	<i>mit dem Thema</i>	53	<i>Haben Sie gut aufgepasst</i>	24
25	<i>von Deutsch als</i>	53	<i>für die Bearbeitung ca.</i>	24
26	<i>in der Schule</i>	53	<i>können Sie anschließend selbst</i>	24
27	<i>Deutsch als Zweitsprache</i>	53	<i>Lernerfolg ist somit garantiert</i>	24

28	<i>Art und Weise</i>	52	<i>Lösungsblatt können</i>	24
29	<i>der Europäischen Union</i>	52	<i>Sie anschließend in Zusammenarbeit mit dem</i>	24
30	<i>in Kooperation mit</i>	51	<i>mit der deutschen Sprache</i>	24
31	<i>des Goethe-Instituts in</i>	51	<i>und kann jederzeit widerrufen</i>	23
32	<i>die in der</i>	50	<i>dass meine E-Mail-Adresse für</i>	23
33	<i>Emerging International Voices</i>	48	<i>das Leben in Deutschland</i>	23
34	<i>Kinder und Jugendliche</i>	47	<i>Anmeldung ist freiwillig und</i>	23
35	<i>auch in der</i>	47	<i>Mit meiner Anmeldung stimme</i>	23
36	<i>Leben in Deutschland</i>	47	<i>E-Mail-Adresse für die Zusendung</i>	23
37	<i>Lehren und Lernen</i>	47	<i>des abonnierten E-Mail-Newsletter verwendet</i>	23
38	<i>Goethe Institutee. V.</i>	46	<i>die Zusendung des abonnierten</i>	23
39	<i>den letzten Jahren</i>	45	<i>Anmeldung stimme ich zu</i>	23
40	<i>in der Lage</i>	44	<i>freiwillig und kann jederzeit</i>	23
41	<i>des Goethe-Instituts und</i>	43	<i>für Kinder und Jugendliche</i>	23
42	<i>und Deutsch als</i>	42	<i>für die Zusendung des abonnierten E-Mail-Newsletter verwendet</i>	23
43	<i>Studium in Deutschland</i>	42	<i>wird</i>	23
44	<i>für den Unterricht</i>	42	<i>Zusendung des abonnierten E-Mail-Newsletter</i>	23
45	<i>in der Fremdsprache</i>	41	<i>Hier finden Sie ein</i>	23
46	<i>sich mit der</i>	41	<i>Die Anmeldung ist freiwillig</i>	23
47	<i>in der deutschen</i>	41	<i>an der University of</i>	23
48	<i>sich in der</i>	40	<i>kann jederzeit widerrufen werden</i>	23
49	<i>und vor allem</i>	39	<i>ist freiwillig und kann</i>	23
50	<i>das Leben in</i>	39	<i>meiner Anmeldung stimme ich</i>	23

---

**Appendix 2. Structural Form of Lexical Bundles in Goethe Institute Germany Website**

<b>Structural Form</b>	<b>Types</b>	<b>%</b>	<b>Lexical Bundles</b>
<i>Adverbialgruppen</i>	3	6%	<i>anschließend selbst die Korrektur Hier finden Sie ein</i>
<i>Adjektivegruppen</i>	1	2%	<i>freiwillig und kann jederzeit</i>
<i>Konjunktionsgruppe</i>	4	8%	<i>und kann jederzeit widerrufen dass meine E-Mail-Adresse für and Language Integrated Learning und Deutsch als Zweitsprache</i>
<i>Nominalgruppe</i>	19	38%	<i>im Bereich Deutsch als selbst die Korrektur vornehmen der Zentrale des Goethe-Instituts Mein Weg nach Deutschland Deutsch als Fremdsprache und Content and Language Integrated die Schülerinnen und Schüler Bereich Deutsch als Fremdsprache dem Lösungsblatt können Sie ein Lernerfolg ist somit Lernerfolg ist somit garantiert Lösungsblatt können Sie anschließend das Leben in Deutschland Anmeldung ist freiwillig und E-Mail-Adresse für die Zusendung die Zusendung des abonnierten Anmeldung stimme ich zu Zusendung des abonnierten E-Mail-Newsletter Die Anmeldung ist freiwillig meiner Anmeldung stimme ich von Deutsch als Fremdsprache für Deutsch als Fremdsprache aus der ganzen Welt in den letzten Jahren auf der ganzen Welt in der Zentrale des Sie für die Bearbeitung Mit dem Lösungsblatt können für die Bearbeitung ca. in Zusammenarbeit mit dem mit der deutschen Sprache Mit meiner Anmeldung stimme für Kinder und Jugendliche für die Zusendung des an der University of Rechnen Sie für die Sie anschließend selbst die Haben Sie gut aufgepasst des abonnierten E-Mail-Newsletter verwendet abonnierten E-Mail-Newsletter verwendet wird können Sie anschließend selbst kann jederzeit widerrufen werden ist freiwillig und kann</i>
<i>Präpositionalgruppen</i>	15	30%	
<i>Verbengruppen</i>	7	14%	
<i>Total</i>	50	100%	