LEXICAL FEATURES OF VERBS IN MECHANICAL ENGINEERING TEXTS: A PILOT CORPUS STUDY

IZVORNI ZNANSTVENI RAD / ORIGINAL SCIENTIFIC PAPER

Jelena Hodak
University of Zagreb, Faculty of Mechanical Engineering and Naval Architecture, Zagreb, Croatia
jelena.hodak@fsb.unizg.hr

Abstract

This study examines the role of verbs in mechanical engineering texts, focusing on their frequency of use and typology in the context of tiered vocabulary framework. Verbs are of crucial importance in technical texts as they enable a precise and concise description of processes and actions. Although past linguistic studies have primarily focused on nouns and nominalisations, as markers of key concepts, topics, and terminology in specialised domains, the role of verbs in technical texts has recently attracted increasing attention. This pilot study is based on a corpus of approximately 26,000 words across three genres: academic textbooks, research papers and engineering magazine articles. Frequency was also analysed in the context of other findings from much larger corpus, namely those of Gardner and Davies (2014) and Deng, Liu and Wu (2022). The results show notable differences in the use of verbs in the three genres. Magazine articles have the highest density and lexical diversity of verbs as well as the highest number of domain-specific verbs. The sub-corpus of research papers shows the highest use of Tier 2 verbs, i.e. academic vocabulary. The findings provide valuable insights into verb usage patterns in mechanical engineering texts, offering practical guidance for teaching English for Specific Purposes to mechanical engineering students.

Keywords: academic vocabulary, English for Specific Purposes, frequency list, mechanical engineering, verbs

1. INTRODUCTION

Verbs play a central role not only in the structuring of sentences, but also in the expression of actions, processes and logical relationships that characterise academic discourse. Academic and technical communication in engineering disciplines relies heavily on the precise and efficient use of verbs to convey processes, causality, states and changes. However, this grammatical category has only recently received more scholarly attention. One reason may be that academic texts are known to rely heavily on nouns, which appear more frequently in academic than in other registers of

English (Hyland & Jiang, 2021). The recent development of text analysis programmes has facilitated the increasing number of corpus studies and the production of various types of word frequency lists, often based on two large corpora, the British National Corpus (BNL) and the Corpus of Contemporary American English (COCA). Several notable lists of academic vocabulary have been published, such as the Academic Vocabulary List (AVL) by Gardner and Davies (2014), the Academic Word List (AWL) by Coxhead (2000), the Academic Collocation List (Ackermann & Chen, 2013), the Academic Formulas List (Simpson-Vlach & Ellis, 2010) and others. Research on academic vocabulary is complemented by valuable results from corpus-based research comparing academic with other registers, such as that of Biber et al. (2021). Vocabulary lists are valuable in educational settings as they enable focused curriculum development and facilitate vocabulary acquisition, although each possesses distinct advantages and limitations. The lists also serve as valuable sources for further research on academic vocabulary. While the Academic Word List (AWL) is widely used for teaching general academic vocabulary (Gholaminejad & Sarab, 2021), its reliance on word families hinders word-class analysis, as it does not provide information on word classes (e.g. *power* as a noun versus a verb). Gardner and Davies's Academic Vocabulary List addresses this limitation through its lemma-based, word-class-tagged methodology. However, existing research predominantly focuses on nouns and nominalizations, while verbs remain comparatively underexplored.

In the field of English for Specific Purposes (ESP), understanding the lexical and grammatical behaviour of verbs in the context of a particular discipline is important as it contributes to the development of effective teaching. For example, texts from the field of mechanical engineering present a particular challenge, as both subject-specific terminology and general academic vocabulary are prevalent here. Scholars have also highlighted the usefulness of analysis and creation of frequency lists of formulaic expressions (e.g. collocations) and grammar patterns in language (e.g. *finish* and *suggest* can be followed by a present participle but not a *to*-infinitive) which enrich the instruction of academic vocabulary (Ackermann & Chen, 2013; Ma & Qian, 2020).

In mechanical engineering discourse, verbs are crucial because they reflect the field's focus on physical processes, system modelling and technical precision. Yet, there are few detailed grammatical studies of verbs in this field. Studies such as that by Frels, Onwuegbuzie and Slate (2010) have attempted to categorise the verb types in scholarly texts and propose typologies that can serve as a basis for teaching academic texts. However, lexical choice and syntactic form (e.g. passives, participial phrases) can vary considerably across disciplines. Therefore, an ESP teacher should supplement general sources with domain-specific ones. Ward's (2009) basic word list for undergraduate engineering students is indeed domain-specific, but covers different types of engineering, namely chemical, civil, electrical, industrial and

mechanical engineering, and includes all parts of speech, i.e. the base is one word. This list cannot be used to analyse verbs in particular, as these are sometimes identical in form to nouns and cannot be identified without context.

This paper presents an exploratory corpus-based study on lexical features of verbs in mechanical engineering texts. It focuses on three representative genres – textbooks, research articles and articles from specialised magazines – that deal with some of the foundational topics in mechanical engineering: shafts, axles and gears. The literature and core textbooks consistently include these components as essential elements in both education and practice as they are critical for transmitting motion, torque, and power in a wide range of machinery. The aim of the study is to analyse the frequency of verbs in general and the use of verbs classified as academic or domain-specific according to the classification in the three tiers of vocabulary. The chosen source for the comparison is the Academic Vocabulary List (AVL) (Gardner & Davies, 2014), a resource which will be described in more detail in the following sections of this paper.

The overall aim of the study is to contribute to a better understanding of a specific segment of discourse in the field of mechanical engineering, namely the use of verbs, and to provide LSP teachers with some insights for teaching academic and technical English to mechanical engineering students. As this is a pilot study based on a relatively small corpus, the results should be interpreted cautiously, with future work extending the scope to larger datasets and a broader range of mechanical engineering topics.

2. METHODOLOGY

This pilot study builds upon a similar study by Kereković (2018), adopting the same corpus size per register. However, unlike Kereković (2018), it includes one additional register, engineering magazine articles, as such texts are commonly used in ESP courses for mechanical engineering students to support the learning of technical vocabulary and grammar. In addition, the thematic focus was adjusted. Whereas Kereković examined texts on design in mechanical engineering, the present study focuses on texts dealing with shafts, gears, axles and, to a lesser extent, pipes.

To explore the lexical features of verbs within this thematic scope, a specially compiled corpus comprising around 26,000 words was created from mechanical engineering texts. To represent the different registers within the field, the corpus includes three genres: a university-level textbook, peer-reviewed research articles and articles from an engineering magazine. All texts were selected based on their focus on common mechanical engineering components – shafts, axles and gears, with the exception of the magazine sub-corpus which also included texts related to pipes. The three sub-corpora were designed to be comparable in both thematic focus and size, to

allow meaningful cross-register analyses of verb usage. The textbook sub-corpus was compiled first and determined the thematic scope and approximate token count for the remaining two sub-corpora, which were then selected to align with these parameters.

Textbook: Excerpts from the *Springer Handbook of Mechanical Engineering* (Grote & Hefazi, 2021) were selected, particularly chapters dealing with shafts, gears and axles. The textbook sections totalled 8,298 words, setting the approximate target size of the other two sub-corpora. This textbook was chosen due to its credibility as a publication by a renowned academic publisher and its function as a comprehensive handbook. Handbooks typically provide overviews of a field, in contrast to highly specialized or niche research papers.

Research articles: Two articles focusing on gears and shafts were selected to match the thematic focus of the study and to align the sub-corpus size with the textbook. Both articles were from mechanical engineering journals indexed in major scientific databases (Chowdhury & Yedavalli, 2016; Feng et al., 2019). The total word count of the research texts was 9,163.

Magazine articles: The third sub-corpus comprises 20 articles published in the UK-based engineering magazine *The Engineer*. Articles were selected based on their thematic relevance and technical content, using the keywords: gears, shafts, axles and pipes, with only those containing technical descriptions included. The search was extended to include the keyword pipes in order to meet the target word count for this sub-corpus, while adhering to the criterion of technical content. This sub-corpus contains 8,386 words, of which 1,375 words are from texts related to pipes.

All texts were manually cleaned of figures, tables, formulae and references to ensure consistency. The corpus was analysed using the online corpus manager and text analysis software *Sketch Engine* (Kilgarriff et al., 2004). The lemmatised corpus for verbs was downloaded in Excel format and each word form was manually checked for part of speech and grammatical form (active, passive, participle, infinitive). Only verbs functioning as lexical verbs or auxiliary verbs (*be, have, do*) in the active or passive voice were retained for analysis.

Both quantitative and qualitative methods of corpus analysis were used in the study to analyse the frequency and typology of verbs in the three text types. The following parameters guided the analysis: 1) verb frequency; 2) tiered vocabulary framework.

Verb frequency was analysed within each of the three genres – textbook (TXB), research paper (PAP) and magazine article (MAG) – and compared with each other to highlight the variations across different registers. To contextualize these findings against larger corpora, the frequency results were also compared with the top 20 verbs listed in three versions of the Academic Vocabulary List (AVL):

1. Gardner and Davies' (2014) Academic Vocabulary List (AVL 2014),
2. Gardner and Davies' (2014) Academic Vocabulary List – Science and Technology domain (AVL 2014 Sci/Tech), and
3. Deng, Liu, and Wu's (2022) AVL – Science and Technology domain (AVL 2019 Sci/Tech).

The tiered vocabulary framework was examined with a focus on domain-specific verbs in the AVL, particularly those classified as Tier 3 vocabulary. The different AVLs are described in greater detail in the following section.

The Gardner and Davies' Academic Vocabulary List (AVL) published in 2014 comprises 554 verbs drawn from a 120-million-word academic sub-corpus of the 425 million word Corpus of Contemporary American English (COCA; Davies, 2010). To qualify as academic, a verb had to meet several criteria: 1. occur at least 50% more frequently in the academic sub-corpus of COCA than would otherwise be expected (i.e. a frequency ratio of 1.50 or higher per million words); 2. have a balanced distribution across the nine domains, with a Juilland's D measure of at least 0.80; 3. achieve at least 20% of the expected frequency in at least seven of the nine domains; 4. not exceed three times the expected frequency in any single domain (Gardner & Davies, 2014).

The COCA is available online at https://www.english-corpora.org/coca/. Two Academic Vocabulary Lists were downloaded from the website as Microsoft Excel files – the core-AVL and the allwords-AVL. The AVL 2014 top 20 verbs were taken from the core-AVL which contains the top 3,000 lemmas across all academic domains. They are listed individually (i.e. not in word families) and labelled according to word class (noun, verb, adjective, etc.). The AVL 2014 Sci/Tech top 20 verbs were also extracted from the core-AVL, which provides discipline-specific data for nine academic domains: Humanities, Philosophy/Religion, History, Law/Political Science, Education, Social Science, Business, Science/Technology, and Medicine. The Science/Technology sub-corpus comprised 22,777,656 words, of which 13,363,151 were from journals and 9,414,505 from magazines. Journals included *Bioscience*, *Environment*, *Mechanical Engineering*, *Physics Today* and *PSA Journal*, while magazines included *Science News*, *Astronomy* and *Technology Review* (Gardner & Davies, 2014).

The COCA Academic corpus used by Gardner and Davies (2014) was largely based on magazine and newspaper articles. Over time, the database was updated by replacing these with journal articles, resulting in the 2019 version being based solely on journal articles, while maintaining a token count of around 120 million. Deng, Liu and Wu (2022) used the updated version to conduct both qualitative and quantitative analyses across nine academic disciplines producing new verb rankings for each. Their Sci/Tech sub-corpus contained 17,454,120 words from 187 journals, yielding a verb ranking that was used in this study for comparison (AVL 2019).

Finally, for the tiered vocabulary framework analysis, presented in section 3.2., the allwords-AVL was used. It consists of around 20,000 most frequent lemmas from the COCA-Academic corpus. This list allows for the identification of domain-specific words, categorised according to the criterion that a word has at least three times the expected frequency in a specific domain (based on the size of this domain in COCA-Academic). In these cases, the word cannot be part of the core academic list.

3. RESULTS

The results are presented in accordance with the key parameters of the study: 1) verb frequency and 2) tiered vocabulary framework.

3.1. Verb frequency

The corpus, which comprises 25,847 tokens (running words), was distributed across the three text types as follows: 8,298 tokens in the textbook sub-corpus, 9,163 in the two research papers, and 8,386 in the magazine articles. Lexical verbs are reported in terms of tokens (all word forms, i.e., inflected forms such as *run, runs, ran*) to capture their overall frequency and distribution within each sub-corpus. Verb lemmas, on the other hand, are used to assess lexical diversity, as grouping inflected forms into a single lemma allows a clearer picture of how many distinct verbs occur in each text type. In the present study lexical verbs in all word forms accounted for 7.68% in the textbook sub-corpus (TXB), 8.84% in the research papers (PAP) and 11.01% in the magazine articles (MAG). Verb lemmas in TXB, PAP and MAG accounted for 23.7%, 24.94% and 29.25% of the total number of verbs in each text type, respectively (Table 1). The magazine sub-corpus had the highest density of verbs, suggesting a more varied narrative style.

Table 1
Lexical Verbs in Three Types of Texts

	TXB	PAP	MAG
TOKENS	8298	9163	8386
VERBS - ALL FORMS (% in corpus)	637 (7.68%)	810 (8.84%)	923 (11.01%)
VERBS – LEMMAS (% in verbs - all forms)	151 (23.7%)	202 (24.94%)	(29.25%)

3.1.1. Primary verbs – be, have, do as auxiliary verbs

Primary verbs *be*, *have* and *do* were analysed separately according to their role in a verb phrase, i.e. as main verbs and auxiliary verbs. The frequency of the primary

auxiliary verbs was uniform overall and accounted for 3% in each of the three text types. However, the frequency of the three auxiliary verbs varied considerably, as can be seen in Figure 1.

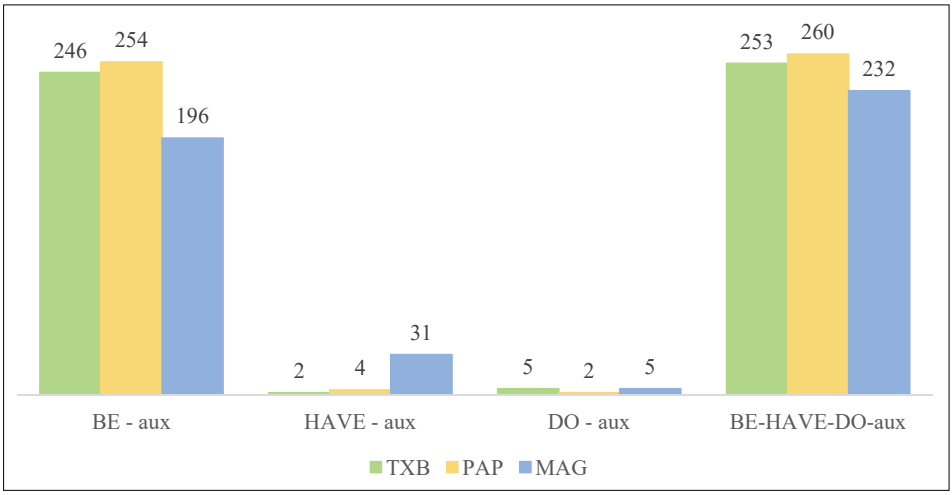

Figure 1
Raw Frequency of Primary Auxiliary Verbs in Three Types of Texts: Textbook, Research Paper and Magazine.

The verb *be* was by far the most frequent verb across all three text types, although it occurred least frequently in MAG. The function of the auxiliary verb *be* is to mark the progressive aspect and the passive voice. Since passive constructions occur more frequently in academic prose, this could explain the comparatively lower frequency of the auxiliary verb *be* in MAG (Fig. 1). The auxiliary verb *do* occurred with similarly low frequency in all three text types, while *have* occurred significantly more frequently in MAG than in TXB and PAP (Fig. 1). This may be attributed to the fact that the auxiliary verb *have* is used exclusively to mark the perfect tense, which occurs more frequently in magazine articles that report on events or activities.

3.1.2. *Primary verbs – be, have, do as main verbs*

The analysis of the primary verbs that function as main verbs showed that the copular *be* occurred with the lowest frequency in TXB (90), compared to PAP (179) and MAG (175). In contrast, the frequencies of *have* and *do* as main verbs remained the same in all three text types. Specifically, *have* appeared 20 times in TXB, 17 times in PAP and 22 times in MAG, while *do* appeared only once as a main verb in the entire corpus (Figure 2).

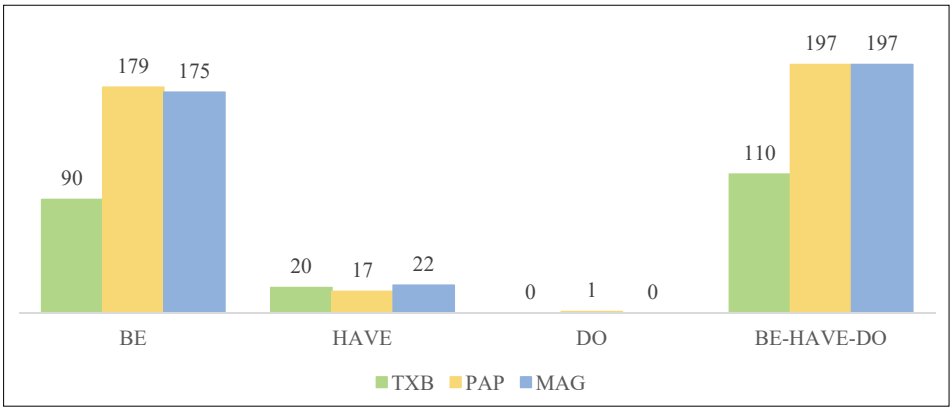

Figure 2
Raw Frequency of Primary Verbs as Main Verbs in Three Types of Texts: Textbook, Research Paper And Magazine.

The pilot study by Kereković (2018), which compared textbook and journal texts on mechanical engineering design, served as a methodological model for the present study, particularly regarding corpus size. Kereković (2018) reported similar results for the verbs *be* and *have* functioning as main verbs. Interestingly, in the same study, the verb *do*, functioning as a main verb, occurred nine times in the textbook corpus. Considering the small but comparable size of the two corpora, this is a notable difference, which may be attributed to the different topics of the corpora.

3.1.3. Lexical verbs

An analysis of the frequency of lexical verbs in the three text types revealed considerable variation. By selecting verbs that occurred at least ten times in a particular text type, a list of 43 verbs was created. The threshold of ten occurrences was chosen as a practical and interpretative cut-off. For a corpus of approximately 26,000 tokens, verbs occurring fewer than ten times were treated as low-frequency. Including them would considerably increase the length of the results section without adding analytical value. Of the 43 verbs, only *be*, *use*, *have*, *operate* and *require* were among the most frequent verbs in all three text types, while *make*, *give*, *develop*, *show* and *design* occurred in two text types. Table 2 presents the full list, including the number of occurrences and the relative frequencies in each sub-corpus, calculated as percentages of total verb lemmas, illustrating their prominence within the overall verb usage in the register.

Table 2

Verbs Occurring at Least 10 Times, With Their Relative Frequencies (% of Verb Lemmas) in Each Text Type

	TXB		PAP		MAG		
VERB	N	%	N	%	N	%	TOTAL N
be	90	59.60%	179	88.61%	175	64.81%	444
use	55	36.42%	35	17.33%	48	17.78%	138
have	20	13.25%	17	8.42%	22	8.15%	59
make	18	11.92%	-	-	33	12.22%	51
give	21	13.91%	14	6.93%	-	-	35
calculate	29	19.21%	-	-	-	-	29
show	13	8.61%	14	6.93%	-	-	27
develop	-	-	11	5.45%	15	5.56%	26
determine	25	16.56%	-	-	-	-	25
obtain	-	-	23	11.39%	-	-	23
operate	5	3.31%	7	3.47%	11	4.07%	23
follow	22	14.57%	-	-	-	-	22
require	8	5.30%	3	1.49%	11	4.07%	22
allow	-	-	-	-	20	7.41%	20
consider	-	-	19	9.41%	-	-	19
bend	19	12.58%	-	-	-	-	19
say	-	-	-	-	18	6.67%	18
increase	-	-	-	-	18	6.67%	18
couple	-	-	17	8.42%	-	-	17
apply	16	10.60%	-	-	-	-	16
define	16	10.60%	-	-	-	-	16
reduce	-	-	-	-	16	5.93%	16
spline	-	-	-	-	16	5.93%	16
provide	-	-	-	-	15	5.56%	15
depend	14	9.27%	-	-	-	-	14
update	-	-	14	6.93%	-	-	14
design	4	2.65%	-	-	10	3.70%	14
call	13	8.61%	-	-	-	-	13
occur	13	8.61%	-	-	-	-	13
rotate	13	8.61%	-	-	-	-	13
manufacture	-	-	-	-	13	4.81%	13
model	-	-	13	6.44%	-	-	13
fit	-	-	-	-	12	4.44%	12
introduce	-	-	12	5.94%	-	-	12
propose	-	-	12	5.94%	-	-	12
equal	11	7.28%	-	-	-	-	11
influence	11	7.28%	-	-	-	-	11
take	11	7.28%	-	-	-	-	11
mesh	-	-	11	5.45%	-	-	11
represent	-	-	11	5.45%	-	-	11
fix	10	6.62%	-	-	-	-	10
connect	-	-	10	4.95%	-	-	10
duplicate	-	-	-	-	10	3.70%	10

Looking at the corpus as a whole, 418 individual lemmas were found, of which only 49 (11.72 %) occurred in all three text types, 107 (25.6 %) in two and 262 (62.68 %) in a single sub-corpus (Table 3). In other words, almost two thirds of all verb lemmas in the corpus occurred in only one text type. This indicates a relatively high degree of variability in the use of verbs in the different text types. The largest overlap was between PAP and MAG (51 common lemmas), followed by 34 common verb lemmas between TXB and MAG. Somewhat unexpectedly, the smallest overlap was between TXB and PAP, with only 22 common lemmas (Table 3).

Table 3

Overlap of Verb Lemmas Between Different Text Types

Lemmas that appear in		# of verb lemmas	Σ
single text type	only TXB	46	262
	only PAP	80	
	only MAG	136	
two text types	TXB & PAP	22	107
	PAP & MAG	51	
	MAG & TXB	34	
three text types	TXB & PAP & MAG	49	49
TOTAL verb lemmas in the overall corpus			418

The list of verbs occurring in more than one text type is provided in Table 4.

Table 4

The List of the Verbs That Appear in More Than One Text Type

ALL THREE TEXTS	TXB & PAP	TXB & MAG	PAP & MAG
achieve; act; adjust; allow; apply; base; be; become; build; calculate; cause; choose; compare; consider; define; depend; describe; determine; discuss; divide; drive; expect; explain; find; fix; follow; get; give; have; include; increase; lead; make; obtain; occur; operate; perform; prevent; produce; provide; remain; require; set; show; subject; take; turn; use; vary	*assume; avoid; coincide; combine; control; correspond; demonstrate; engage; execute; express; form; hold; ignore; implement; influence; list; note; pass; relate; represent; result; transform*	*begin; call; change; check; compensate; create; cut; design; disrupt; distribute; exceed; exist; handle; induce; know; leave; limit; load; mean; mesh; mix; pair; play; protect; reduce; rotate; save; say; see; separate; support; tend; transport; work*	*add; affect; alter; appear; arise; attach; bring; carry; characterise; collect; come; comprise; consist; couple; decrease; deliver; develop; display; employ; enable; establish; estimate; evolve; experience; extend; focus; generate; govern; identify; incorporate; install; integrate; introduce; involve; maintain; match; measure; modify; mount; pose; remove; replace; run; simulate; smooth; speed; split; state; study; test; transmit*

The magazine sub-corpus contained the highest number of verb occurrences, i.e. verbs that occur in any of the forms (e.g. *be*, *was*, *been*), and also the largest number of verbs that occur only once. Of the 923 verb occurrences identified in MAG, 144 (15.6%) verbs occurred only once, compared to 88 (10.86%) in PAP and 69 (10.83%) in TXB. One possible explanation for this greater heterogeneity of the MAG sub-corpus in this respect is its composition: it comprises numerous shorter texts written by different authors, which introduces greater lexical variety through individual writing styles and vocabulary choices.

One of the aims of this study was to investigate the extent to which the results of the three different genres align with those obtained from much larger corpora. The top 20 academic verbs were identified in the three variants of the Academic Vocabulary List: 1) the 2014 Gardner and Davies' Academic Vocabulary List (AVL 2014); 2) the 2014 AVL Science and Technology domain (AVL 2014 Sci/Tech); and 3) the Deng, Liu and Wu's (2022) ranking for Science and Technology domain based on the updated version of the AVL (AVL 2019 Sci/Tech).

It should be noted here that the ranking of verbs in the AVL 2014 was based on the lemma frequency in 120 million words in the COCA Academic corpus, while Deng, Liu, and Wu's (2022) ranking is based on the normalised frequency per million words (PMW). The comparison between the three AVL-based rankings and the raw verb frequency data from this corpus is shown in Tables 5–7. To facilitate interpretation, the colour gradient in table cells darkens progressively with increasing values. A perfect match would therefore show a gradual transition from the darkest shade at the table's top to the lightest at the bottom. As expected, the results correspond more closely with the discipline-specific (Sci/Tech) rankings than the general AVL ranking. Interestingly, the improvement in agreement is most notable in the AVL 2019 MAG sub-corpus, even though the AVL 2019 corpus comprises only academic journal articles, while 41% of the AVL 2014 Sci/Tech corpus are magazine and newspaper articles (Tables 5-7).

Table 5

Comparison of the 2014 Core-AVL Academic Top 20 Verbs and Their Raw Frequencies Across the Three Text Types (TXB = Textbook, PAP = Research Paper, MAG = Magazine)

	VERB	TXB	PAP	MAG
1	*provide*	1	4	15
2	*include*	6	6	8
3	*develop*	0	11	15
4	*suggest*	0	0	0
5	*require*	8	2	11
6	*report*	0	0	1
7	*base*	2	8	4
8	*describe*	2	4	1
9	*indicate*	0	5	0

10	*produce*	4	3	4
11	*identify*	0	2	2
12	*support*	1	0	1
13	*increase*	5	4	12
14	*note*	3	2	0
15	*represent*	1	11	0
16	*determine*	23	5	8
17	*occur*	13	2	3
18	*present*	0	5	0
19	*reduce*	3	0	15
20	*involve*	0	4	2

Note: Darker shading highlights higher raw frequencies. The colour distribution shows limited overlap of frequency ranking, as several high-frequency verbs in this corpus (*determine, represent, occur, reduce*) rank much lower in the AVL general academic verb ranking, while typical academic verbs (*suggest, report, indicate, support*) occur less frequently in this corpus.

Table 6

Comparison of the 2014 Core-AVL – Sci/Tech Top 20 Verbs and Their Raw Frequencies Across the Three Text Types (TXB = Textbook, PAP = Research Paper, MAG = Magazine)

	VERB	TXB	PAP	MAG
1	*provide*	1	4	15
2	*include*	6	6	8
3	*develop*	0	11	15
4	*produce*	4	3	4
5	*require*	8	2	11
6	*reduce*	3	0	15
7	*increase*	5	4	12
8	*suggest*	0	0	0
9	*design*	4	10	1
10	*form*	4	3	0
11	*base*	2	8	4
12	*occur*	13	2	3
13	*determine*	23	5	8
14	*contain*	1	0	0
15	*observe*	0	7	0
16	*test*	0	1	2
17	*indicate*	0	5	0
18	*support*	1	0	1
19	*identify*	0	2	2
20	*describe*	2	4	1

Note: Darker shading highlights higher raw frequencies. Compared to the 2014 AVL general academic verb ranking (Table 5), alignment improves with its Sci/Tech list, particularly for verbs such as *determine, occur, reduce,* and *increase.*

Table 7

Comparison of the 2019 AVL – Sci/Tech Top 20 Verbs and Their Raw Frequencies Across the Three Text Types (TXB = Textbook, PAP = Research Paper, MAG = Magazine)

	VERB	TXB	PAP	MAG
1	*provide*	1	4	15
2	*include*	6	6	8
3	*develop*	0	11	15
4	*require*	8	2	11
5	*increase*	5	4	12
6	*produce*	4	3	4
7	*reduce*	3	0	15
8	*base*	2	8	4
9	*determine*	23	5	8
10	*compare*	4	6	4
11	*indicate*	0	5	0
12	*suggest*	0	0	0
13	*occur*	13	2	3
14	*represent*	1	11	0
15	*identify*	0	2	2
16	*describe*	2	4	1
17	*design*	4	10	1
18	*report*	0	0	1
19	*contain*	1	0	0
20	*observe*	0	7	0

Note: Darker shading highlights higher raw frequencies. The 2019 AVL – Sci/Tech list shows the strongest correspondence with the present corpus, particularly for *determine*, which is highly frequent in the textbook sub-corpus, and for *require*, *increase*, and *reduce* in the magazine sub-corpus, all showing better alignment than in the 2014 AVL – Sci/Tech ranking (Table 6).

It should be noted that these results are based on raw frequency counts and percentages within a relatively small corpus. Normalized values were not used, as the corpus size is insufficient for meaningful normalization. Therefore, the findings should be interpreted as indicative rather than fully generalisable.

3.2. Vocabulary tier distribution

The *tiered vocabulary framework* is a commonly applied framework in education that categorises words according to their frequency of use, their degree of complexity and their contextual range. It serves as a practical tool for educators to determine which vocabulary should be emphasised in the classroom, particularly to improve learners' academic performance and language acquisition. Tier 1 includes basic, everyday words that are typically acquired through natural language exposure. Tier 2 comprises

high-frequency academic vocabulary that occurs across a variety of subjects, while Tier 3 consists of specialised, low-frequency terms that are mainly relevant in specific disciplines or subject areas. In the context of mechanical engineering, these would be technical terms.

As already mentioned, the COCA-Academic allwords-AVL makes it possible to extract domain-specific words (Tier 3). The criterion for inclusion in this particular list is that the word occurs at least three times as frequently in one of the nine domains of COCA-Academic (e.g. Law, Medicine or Business) as would otherwise be expected given the domain's size. According to this type of verb categorisation and based on the COCA-Academic ranking, there were only two Tier 3 verbs in TXB and PAP and, somewhat surprisingly, four in the magazine sub-corpus. The verbs categorised as Tier 3 and Tier 2 in this study are listed in Tables 8 and 9 respectively.

Table 8

COCA-Academic Domain-Specific (Tier 3) Verbs Found in the Three Types of Texts (TXB = Textbook, PAP = Research Paper, MAG = Magazine)

TXB (2 verb lemmas)	PAP (2 verb lemmas)	MAG (4 verb lemmas)
cool, mate	power; simulate	download; optimise; retrofit; simulate

Table 9

COCA-Academic Vocabulary List (Tier 2) Verbs Found in the Three Types of Texts (TXB = Textbook, PAP = Research Papers, MAG = Magazine)

TXB (59 of 151 verb lemmas; 39.07%)	PAP (109 of 202 verb lemmas; 53.96%)	MAG (81 of 270 verb lemmas; 30%)
achieve; apply; assume; base; calculate; characterize; classify; coincide; compare; compensate; contain; control; convert; correspond; counteract; define; demonstrate; depend; describe; design; determine; deviate; differ; discuss; distribute; dominate; engage; enlarge; exceed; express; form; fulfill; group; include; increase; influence; interpret; note; obtain; occur; perform; produce; provide; reduce; relate; represent; require; resolve; result; simplify; specify; subject; summarize; superimpose; tend; transform; turn; vary; verify	*achieve; affect; alter; analyze; apply; approximate; arise; assume; attempt; base; calculate; characterise; coincide; compare; compose; comprise; consist; constitute; constrain; contribute; control; convey; correspond; couple; deem; define; demonstrate; denote; depend; derive; describe; detect; determine; develop; diminish; discuss; display; distinguish; employ; enable; engage; establish; estimate; evolve; exist; experience; express; extend; focus; form; formulate; generate; govern; highlight; identify; illustrate; include; incorporate; increase; indicate; induce; influence; initiate; integrate; interact; isolate; justify; limit; locate; maintain; measure; model; modify; neglect; normalize; note; observe; obtain; occur; organise; outline; perform; permit; possess; precede; predict; present; produce; promote; propagate; propose; provide; recognize; relate; rely; render; represent; require; result; simplify; state; subject; summarise; support; term; test; transform; transmit; vary*	*account; accumulate; achieve; adapt; affect; alter; apply; arise; base; calculate; characterise; compare; compensate; comprise; consist; construct; couple; define; depend; describe; determine; develop; devise; discuss; display; distribute; employ; enable; encounter; ensure; establish; estimate; evolve; exceed; exhibit; exist; experience; extend; extrapolate; focus; function; generate; govern; identify; improve; include; incorporate; increase; induce; integrate; intersect; involve; limit; maintain; map; measure; minimise; modify; obtain; occur; perform; produce; provide; quantify; range; reduce; replicate; report; require; span; standardise; state; strive; subject; support; suppress; tend; test; transmit; vary; view*

It is interesting to note that in the present corpus, most domain-specific verbs (Tier 3) occurred in the MAG sub-corpus, and not in TXB or PAP, which one would expect given the more technical and disciplinary orientation. As far as the distribution of Tier 2 verbs across the three text types is concerned, the proportion was highest in the PAP sub-corpus (53.96%), while it was lowest in MAG sub-corpus (30%). However, the fact that almost a third of Tier 2 verbs occurred in the MAG sub-corpus suggests that magazine articles may be a valuable resource for learners in the early stages of academic study. The distribution of verbs across the three tiers is shown in Figure 3.

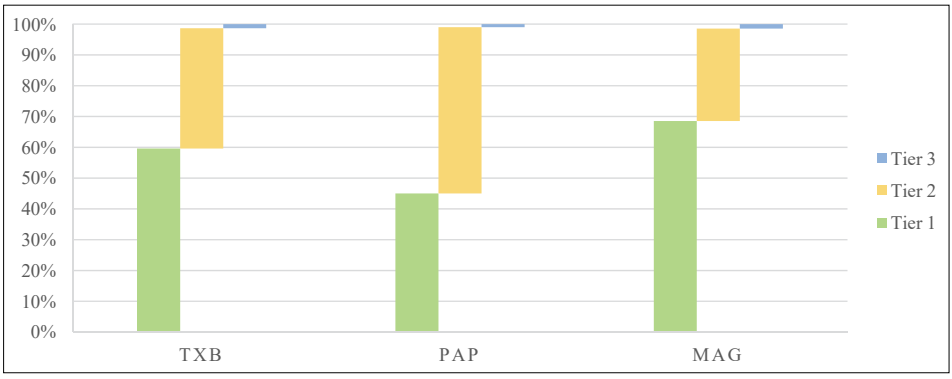

Figure 3
The Proportion of Verb Lemmas Categorised by Tiers in Each Type of Text (TXB = Textbook, PAP = Research Paper, MAG = Magazine)

4. DISCUSSION

One limitation of the present study is the relatively small size of the corpus, which comprises just over 26,000 running words. However, in contrast to the general nature of AVL, this corpus is narrowly focused on a niche topic in mechanical engineering – gears, shafts and axles – components essential for transmitting motion and power in mechanical systems. Due to the limited availability of adequate magazine articles on these components, approximately 1,000 words of the magazine sub-corpus cover the topic of pipes, which represents an additional limitation of the study.

The analysis of the primary verbs is in line with the results of extensive corpus studies. According to Biber et al. (2021), *be* is the most common verb in academic prose, a finding that was confirmed in this study. The results regarding the use of primary verbs as auxiliary verbs are also consistent with those reported in large corpus studies, such as Biber et al. (2021). Furthermore, the results are comparable to those of Kereković (2018), whose study on a similarly sized corpus (approx. 8,000 words per sub-corpus) focused on the topic of engineering design.

One of the research questions addressed in this study concerned the similarity between the three sub-corpora in terms of verb selection and frequency. Interestingly, the highest degree of overlap in the use of verbs was found between the research paper and magazine article sub-corpora. On closer inspection, this can be attributed to the fact that magazine articles often report on recent research findings. Comparing the results of this study with those based on large corpora such as the AVL emphasises the value of discipline-specific word lists – particularly verb lists – in facilitating the development of targeted and effective LSP curricula. The results suggest that magazine articles specialising in a particular discipline contain a considerable amount of academic and may even contain domain-specific vocabulary, making them a valuable resource for developing instructional materials and introducing learners to discipline-specific language. Given their density of Tier 2 verbs, research papers, on the other hand, seem best suited for use in more advanced stages of LSP teaching. Finally, regarding the diversity of vocabulary, more specifically the diversity of verbs, it seems that textbooks should not be the preferred choice as a source for teaching material and that using magazine texts and research papers as source texts will prepare students to use textbooks competently in the field of mechanical engineering.

5. CONCLUSION

The results of this exploratory, corpus-based study provide an insight into the selection and use of verbs in a niche topic of mechanical engineering, which is represented in three text types: textbooks, research papers and engineering magazine articles. The aim of the study was to investigate the pedagogical value of different text types for teaching in the context of Language for Specific Purposes (LSP) and to inform teachers and curriculum designers. The Academic Vocabulary List (AVL) was used as a reference to situate the results within a broader body of research based on much larger corpora. The AVL is based on a corpus of 120 million words taken from texts in various academic disciplines. This study contributes to our understanding of verb use in mechanical engineering discourse by identifying verb choice preferences in textbooks, research papers and magazine articles specialising in this field.

These findings have several pedagogical implications for the teaching of English for Specific Purposes (ESP). Firstly, the significant variation in verb use between text genres emphasises the need for genre-aware language teaching. In other words, sources for vocabulary instruction should be selected according to student's needs and level of proficiency. Secondly, while general academic vocabulary lists such as the AVL are useful for selecting the most frequent vocabulary, they should be complemented by domain-specific corpora in ESP teaching. It should also be noted that although vocabulary lists may be helpful for the development of curricula or teaching materials,

they should not be regarded as definitive or universally applicable, but should be used as a guideline, as language is a dynamic entity and any instructional material should be customised to the subject area and the needs of the learners. Finally, the findings underscore the importance of further targeted corpus-based research that looks more closely at vocabulary, syntactic patterns (e.g. passive constructions, verb complementation) or the development of verb use at different language proficiency levels.

References

Ackermann, K., & Chen, Y. (2013). Developing the academic collocation list (ACL) – A corpus-driven and expert-judged approach. *Journal of English for Academic Purposes, 12*(4), 235–247. https://doi.org/10.1016/j.jeap.2013.08.002

Biber, D., Johansson, S., Leech, G. N., Conrad, S., & Finegan, E. (2021). *Grammar of spoken and written English* (2nd ed.). John Benjamins Publishing Company.

Coxhead, A. (2000). A New Academic Word List. *TESOL Quarterly, 34*(2), 213–238. https://doi.org/10.2307/3587951

Davies, M. (2010). The Corpus of Contemporary American English as the first reliable monitor corpus of English. *Literary and Linguistic Computing, 25*(4), 447–464. https://doi.org/10.1093/llc/fqq018

Deng, Y., Liu, D., & Wu, S. (2022). Academic English verbs across disciplines: A corpus study and its implications. *Ampersand, 9*, 100093. https://doi.org/10.1016/j.amper.2022.100093

Frels, R. K., Onwuegbuzie, A. J., & Slate, J. R. (2010). Editorial: A typology of verbs for scholarly writing. *Research in the Schools, 17*(1), xx–xxxi.

Gardner, D., & Davies, M. (2014). A new academic vocabulary list. *Applied Linguistics, 35*(3), 305–327. https://doi.org/10.1093/applin/amt015

Gholaminejad, R. and Sarab, M. R. A. (2021). A Comparison of the Academic Word List and the Academic Vocabulary List: Should the AVL Replace the AWL?. TEFLIN Journal, 32(1). https://doi.org/10.15639/teflinjournal.v32i1/167-182

Hyland, K., & Jiang, F. (2021). Academic Naming: Changing Patterns of Noun Use in Research Writing. *Journal of English Linguistics*, *49*(3), 255–282. https://doi.org/10.1177/00754242211019080

Kereković, S. (2018). Typologies of verbs used in mechanical engineering texts. In P. Guerrero Medina, R. Torre Alonso, & R. Vea Esparza (Eds.), *Verbs, clauses and constructions: Functional and typological approaches* (pp. 3–20). Cambridge Scholars Publishing.

Kilgarriff, A., Rychlý, P., Smrž, P., & Tugwell, D. (2004). *The Sketch Engine*. In G. Williams & S. Vessier (Eds.), *Proceedings of the 11th EURALEX International Congress* (pp. 105–116). Université de Bretagne Sud. https://www.sketchengine.eu/

Ma, H., & Qian, M. (2020). The creation and evaluation of a grammar pattern list for the most frequent academic verbs. *English for Specific Purposes, 58*, 155–169. https://doi.org/10.1016/j.esp.2020.01.002

Simpson-Vlach, R., & Ellis, N. C. (2010). An Academic Formulas List: New Methods in Phraseology Research. *Applied Linguistics, 31*(4), 487–512. https://doi.org/10.1093/applin/amp058

Ward, J. (2009). A basic engineering English word list for less proficient foundation engineering undergraduates. *English for Specific Purposes*, *28*(3), 170–182. https://doi.org/10.1016/j.esp.2009.04.001

Corpus references

Chowdhury, S., & Yedavalli, R. K. (2016). Dynamics of belt-pulley-shaft systems. *Mechanism and Machine Theory, 98*, 199–215. https://doi.org/10.1016/j.mechmachtheory.2015.11.011

Feng, K., Gu, F., Ball, A. D., & Ma, L. (2019). Vibration-based updating of wear prediction for spur gears. *Wear, 426–427*, 1410–1415. https://doi.org/10.1016/j.wear.2019.01.017

Grote, K. H., & Hefazi, H. (2021). *Springer handbook of mechanical engineering* (2nd ed.). Springer.

The Engineer. Mark Allen Group. https://www.theengineer.co.uk/. Accessed February. 2024

Elesa. (2018, November 9). New Standard Components – modular Racks and Spur gears from Elesa. *The Engineer*. https://www.theengineer.co.uk/content/product/new-standard-components-modular-racks-and-spur-gears-from-elesa

JBJ Techniques. (2018, September 5). Bearing supported spline shafts. *The Engineer*. https://www.theengineer.co.uk/content/product/bearing-supported-spline-shafts

JBJ Techniques. (2019, June 21). Planetary Gearboxes designed for rugged use and dependability. *The Engineer*. https://www.theengineer.co.uk/content/product/planetary-gearboxes-designed-for-rugged-use-and-dependability

JBJ Techniques. (2022, August 30). Mechanical Power Transmission ~ Driver to Driven. *The Engineer*. https://www.theengineer.co.uk/content/product/mechanical-power-transmission-driver-to-driven

JBJ Techniques. (2022, July 7). Complete selection of PLANETARY GEARBOXES which include rotating shaft, in-line and right angle output. *The Engineer*. https://www.theengineer.co.uk/content/product/complete-selection-of-planetary-gearboxes-which-include-rotating-shaft-in-line-and-right-angle-output

Knight, H. (2019, July 30). Volkswagen goes greener with new manual gearbox. *The Engineer*. https://www.theengineer.co.uk/content/news/volkswagen-goes-greener-with-new-manual-gearbox

norelem UK. (2020, March 6). norelem hits Top Gear with Bevel Gears and Worm Drives. *The Engineer*. https://www.theengineer.co.uk/content/product/norelem-hits-top-gear-with-bevel-gears-and-worm-drives

norelem UK. (2020, September 23). Splined Shafts and Hubs from norelem are the Torque of the Town. *The Engineer*. https://www.theengineer.co.uk/content/product/splined-shafts-and-hubs-from-norelem-are-the-torque-of-the-town

The Engineer (2006, January 19). *Another fine mesh*. https://www.theengineer.co.uk/content/news/another-fine-mesh

The Engineer. (2008, August 6). *Flexible spring shafts from Rino*. https://www.theengineer.co.uk/content/news/flexible-spring-shafts-from-rino

The Engineer. (2007, March 13). *Full control.* https://www.theengineer.co.uk/content/news/full-control

The Engineer. (2008, December 19). *Gearboxes stay cool.* https://www.theengineer.co.uk/content/news/gearboxes-stay-cool

The Engineer. (2009, June 15). *Hansen gear units speed up Corus line.* https://www.theengineer.co.uk/content/news/hansen-gear-units-speed-up-corus-line

The Engineer. (2008, January 30). *High tolerance worm gears are profiled at Metav.* https://www.theengineer.co.uk/content/news/high-tolerance-worm-gears-are-profiled-at-metav

The Engineer. (2006, July 14). *Model pipes up.* https://www.theengineer.co.uk/content/news/model-pipes-up

The Engineer. (2009, October 23). *New Hypoid style gearbox for quiet 90 degree operation.* https://www.theengineer.co.uk/content/news/new-hypoid-style-gearbox-for-quiet-90-degree-operation

The Engineer. (2009, April 6). *Precision line shafts offer convenience and performance.* https://www.theengineer.co.uk/content/news/precision-line-shafts-offer-convenience-and-performance

The Engineer. (2008, June 23). *Renold gears.* https://www.theengineer.co.uk/content/news/renold-gears

The Engineer. (2005, November 28). *Twist control.* https://www.theengineer.co.uk/content/news/twist-control/

The Engineer. (2008, April 28). *Wright flow technologies.* https://www.theengineer.co.uk/content/news/wright-flow-technologies

GLAGOLSKE ZNAČAJKE U TEKSTOVIMA IZ PODRUČJA STROJARSTVA: PILOT ISTRAŽIVANJE

Ovo istraživanje analizira značajke glagola u tekstovima iz područja strojarstva, s posebnim naglaskom na njihovu učestalost i tipologiju u okviru triju razina vokabulara. Glagoli imaju ključnu ulogu u tehničkim tekstovima jer omogućuju precizno i sažeto opisivanje procesa i radnji. Iako su se prethodna jezična istraživanja uglavnom usmjeravala na imenice i nominalizacije kao pokazatelje temeljnih pojmova, tema i terminologije u stručnim područjima, uloga glagola u tehničkim tekstovima sve je češće predmet proučavanja. Ova pilot-studija temelji se na korpusu od približno 26.000 riječi u tri potkorpusa: akademski udžbenik, znanstveni radovi i popularno-znanstveni članci iz strojarstva, po uzoru na istraživanje Kereković (2018). Učestalost je također analizirana u kontekstu većih korpusa, konkretno radova Gardner i Davies (2014) te Deng, Liu i Wu (2022). Rezultati pokazuju značajne razlike u uporabi glagola među trima žanrovima: časopisi imaju najveću gustoću i leksičku raznolikost glagola, kao i najviše glagola iz treće razine vokabulara (stručni vokabular, tehnička terminologija), dok znanstveni radovi broje najviše glagola iz druge razine, odnosno akademskog vokabulara. Rezultati doprinose boljem razumijevanju glagolskih značajki u tekstovima iz područja strojarstva, što može unaprijediti razvoj nastave engleskog jezika struke za studente strojarstva.

Ključne riječi: akademski vokabular, čestotni popis riječi, engleski za posebne svrhe, glagoli, strojarstvo