# Disciplinary Variation in Engineering Research Article Abstracts: Comparing Generative-AI Descriptions with Corpus-Based Evidence

**Yuka ISHIKAWA**
Nagoya Institute of Technology
Gokiso, Nagoya, Japan
ishikawayuka.jp@gmail.com

**Abstract**

This study investigates disciplinary variation in the use of tense, aspect, and voice in research article abstracts from three engineering fields—chemical engineering, civil engineering, and computer science—and examines the extent to which AI-generated descriptions accurately capture such variation. First, ChatGPT was prompted in three independent sessions to identify characteristic grammatical patterns in abstracts from the three fields. The observations were then compared with corpus-based findings from the Engineering Research Abstract Corpus (ERAC), comprising 3,000 abstracts published in 2025–2026. The corpus was tagged using the Multi-Dimensional Analysis Tagger and analyzed using AntConc, with particular attention to present and past tenses, perfect aspect, passive voice, and active constructions with *we*. The results showed substantial disciplinary differences. Computer science abstracts were characterized by frequent use of the present tense and active constructions with *we*, largely confirming the AI-generated descriptions. In contrast, corpus evidence revealed that chemical engineering abstracts also strongly favored the present tense and active voice, contrary to AI descriptions emphasizing past-tense passives. Civil engineering abstracts used the past tense and passive voice most frequently and rarely employed *we* as the subject. These findings demonstrate that AI does not always capture discipline-specific linguistic patterns accurately and suggest that generative-AI advice should be supplemented with corpus evidence in ESP writing instruction.

**Keywords:** research article abstracts, ESP, Multi-Dimensional Analysis Tagger

## 1. Introduction

Writing an effective research article (RA) abstract is a critical yet challenging task for graduate students in science and engineering (Caplan & Johns, 2022; Glasman-Deal, 2010; Huckin, 2001; Swales & Feak, 2009). Although previous studies have extensively examined the macro-structures of RAs, variation across engineering subdisciplines remains underexplored, particularly with respect to patterns of linguistic co-occurrence.

Grounded in English for Specific Purposes (ESP), this study investigates the lexico-grammatical characteristics of RA abstracts across three engineering disciplines: chemical engineering, civil engineering, and computer science. Within ESP, the engineering domain has often been treated as a single, monolithic category labelled simply "Engineering." However, engineering encompasses a wide range of disciplines, and recent corpus-based studies have revealed distinct patterns of language use across its various sub-disciplines (e.g., Hyland, 2005; Ishikawa, 2024).

Within the field of ESP, it is widely recognized that tense, aspect, and voice are among the most challenging linguistic features for EFL students (Apse & Farneste, 2020; Fuchs & Werner, 2020; Hinkel, 2002, 2004; Swales & Feak, 2012). Even when performing the same rhetorical function—such as reporting research findings—students must choose among markedly different grammatical structures. These include active present forms like "*we find*," active past forms like "*we found*," and passive variations like "*it was found*" and "*it has been found*."

While generative AI is increasingly used to support academic writing, including research paper writing, it remains unclear whether it adequately reflects the linguistic nuances that distinguish these sub-disciplines. This study examines the extent to which AI-generated guidance reflects disciplinary variation in language use across these fields by comparing AI-generated descriptions with corpus-based findings on the linguistic features of engineering RA abstracts. We first use AI to investigate how such lexico-grammatical choices are distributed across engineering disciplines, and subsequently verify the findings through corpus analysis. We also examine how the obtained insights can be directly applied to the writing of research abstracts in the field of engineering.

The study addresses the following three research questions:

RQ1: How are choices of tense, aspect, and voice distributed across engineering disciplines in AI-generated descriptions?

RQ2: To what extent are AI-generated descriptions of patterns of tense, aspect, and voice validated and refined by subsequent corpus analysis?

RQ3: How can these findings inform ESP instruction for writing engineering RA abstracts?

## 2. Methods

### 2.1 AI prompting procedure

The following prompt was submitted in both English and Japanese to ChatGPT, a widely used generative-AI system:

> Please pay attention to the tenses, aspects, and voices that are often used in chemical engineering paper abstracts, civil engineering paper abstracts, and computer science paper abstracts, and tell us the characteristics of each.

Three independent prompting sessions were conducted between August 7 and 9, 2026, each initiated as a new chat session. Requests were made in Japanese on the first day and in English on the second and third days. The version of ChatGPT used was GPT-5.6 Sol.

### 2.2 Engineering Research Abstract Corpus (ERAC)

First, using the SCImago Journal Rank website, we compiled a list of high-impact international journals for each of the three major engineering subfields: chemical engineering, civil engineering, and computer science. Next, in collaboration with faculty members who specialize in each field, we selected journals from these lists that were widely recognized and highly rated within their fields from those journal lists. We selected multiple journals to reduce the influence of any single journal's writing style on the data. Finally, for each discipline, we randomly collected a total of 1,000 abstracts published in 2025 or 2026 from the selected journals.

The selected journals are as follows:

Chemical: *Journal of the American Chemical Society*, *Nature Biotechnology*, *Nature Catalysis*, *Nature Chemistry*, *Nature Nanotechnology*

Civil: *Automation in Construction*, *Computer-Aided Civil and Infrastructure Engineering*, *Engineering Structures*, *Structural Safety*, *Water Research*

Computer: *IEEE Transactions on Knowledge and Data Engineering*, *IEEE Transactions on Neural Networks and Learning Systems Publication Information*, *IEEE Transactions on Pattern Analysis and Machine Intelligence*, *Nature Computational Science*, *Nature Machine Intelligence*

The outline of the data analyzed in this study is shown in Table 1.

| | Chem | Civil | Comp |
|---|---|---|---|
| Abstracts | 1,000 | 1,000 | 1,000 |
| Tokens | 168,239 | 214,081 | 203,012 |
| Types | 12,316 | 10,724 | 11,004 |

Table 1: Overview of the ERAC

### 2.3 Data tagging and linguistic features

Data were tagged using the Multidimensional Analysis Tagger (MAT), which was developed based on the framework proposed by Biber (1988). According to Nini (2019), MAT uses the Stanford Tagger (Toutanova et al., 2003) to perform the part-of-speech tagging and then identifies the 67 lexico-grammatical patterns described in Biber (1988). Some POS tags are replaced with new tags that represent more complex linguistic features.

In this study, we examine the following tags related to tense, aspect, and voice. Regarding tense, the prevailing view recognizes two types (Huddleston & Pullum, 2002; Quirk et al., 1985). Here, we focus on the past and present tenses, specifically examining the MAT tags VBD and VPRT.

Regarding aspect, because the progressive aspect rarely appears in abstracts, we limit our investigation to the perfect aspect. We use the MAT tag PEAS to analyze instances of "have/has + past participle" and "had + past participle."

Regarding voice, we conduct the analysis using the MAT tags BYPA and PASS. We also examine the use of active voice with the first-person pronoun "we." Table 2 summarizes the linguistic patterns examined.

| | |
|---|---|
| Tense | - VBD: Past tense<br>- VPRT: Present tense |
| Aspect | - PEAS: Perfect aspect<br>VBD/VPRT + PEAS |
| Voice | - BYPA: By-passives<br>- PASS: Agentless passives<br>VBD/VPRT + BYPA/PASS<br>- FPP1: First person pronouns<br>WE + VBD/VPRT |

Table 2: Linguistic patterns examined

AntConc was used to analyze both plain and tagged texts. The UCREL Significance Test System is used for statistical testing.

## 3. Results and Discussion

### 3.1 AI-generated responses

The three responses generated by ChatGPT were highly similar in content, although they varied somewhat in format, number of examples, and length of explanations. The responses were compared manually, and recurrent claims appearing in at least two of the three responses were identified and summarized in Table 3.

| | Chem | Civil | Comp |
|---|---|---|---|
| Tense | mainly past + present | past + present | present + past |
| Aspect | simple + some present perfect | simple + occasional perfect | strongly simple |
| Voice | frequent passive | passive + active | active + we |

Table 3: Summary of ChatGPT responses

According to ChatGPT, chemical engineering abstracts mainly use the past tense to report that an experiment or process was performed. They also use the present tense, often in the final sentence, to describe general truths, establish knowledge, and state the conclusion. Simple past and simple present are overwhelmingly preferred, and present perfect is occasionally used in the background section to indicate previous research. Passive voice is extremely common because the focus is placed on the process, material, or experiment rather than the researcher. The overall characteristic is therefore "past tense + passive voice".

Civil engineering abstracts tend to use the past tense to report analyses, tests, field measurements, or case studies and the present tense for conclusions and statements of significance. Simple aspect is again dominant with the present perfect occasionally used in literature-gap statements. Civil engineering abstracts use both passive and active voice. A common pattern is to begin with the present in the background or objective section, move to the past in the methods and results, and then return to the present in the conclusion or implications. Compared with chemical engineering, civil engineering often shows a slightly higher proportion of active constructions, especially in journals that encourage concise writing.

In computer science abstracts, the present tense is predominant and is more frequently used than in the other two fields to describe algorithms, systems, models, and contributions. Simple aspect overwhelmingly dominates, with the present perfect used in literature reviews. Active voice is much more frequent than in the traditional engineering fields, so the combination of active voice and present tense appears more frequently than in chemical or civil engineering abstracts. Computer science journals generally favor clear, author-centered expressions, and their abstracts are therefore contribution-oriented.

## 3.2 Corpus-based analysis

This section analyzes the MAT-tagged ERAC and examines the results regarding tense, aspect, and voice.

### 3.2.1 Tense

Nearly 90% of the verbs used in Chemical Engineering and nearly 95% of those in Computer Science are in the present tense, whereas the rate is less than 80% in Civil Engineering. The combined frequency of present-tense and past-tense verbs ranges from 5.6 to 6.1 per 100 words.

These figures are presented in Table 4, along with the separate frequencies of present-tense and past-tense verbs per 100 words for each field.

| | Chem | Civil | Comp |
|---|---|---|---|
| Proportion of Past-Tense Verb | 12.3% | 22.1% | 5.4% |
| Proportion of Present-Tense Verb | 87.7% | 77.9% | 94.6% |
| Freq. of Past-Tense Verb (per 100) | 0.69 | 1.27 | 0.33 |
| Freq. of Present-Tense Verb (per 100) | 4.93 | 4.47 | 5.79 |

Table 4: Proportion of past and present tenses used in each discipline

In Computer Science, which has the highest proportion of present-tense verbs, 348 of the 1,000 abstracts contain at least one past-tense verb, whereas in Civil Engineering, which has the lowest proportion of present-tense verbs, 610 abstracts contain at least one past-tense verb. This indicates that a substantial number of abstracts contain no past-tense verbs, suggesting that present-tense forms may also be used to describe methods and results.

We investigated which verbs are frequently used in the present tense across the three fields. Lemmatization was not performed. Table 5 presents ten most frequent present-tense verbs in descending order of frequency.

| | Present | | |
|---|---|---|---|
| | Chem | Civil | Comp |
| 1 | is | is | is |
| 2 | are | are | are |
| 3 | has | proposes | propose |
| 4 | report | demonstrate | demonstrate |
| 5 | show | provides | has |
| 6 | remains | show | have |
| 7 | have | has | introduce |
| 8 | demonstrate | have | show |
| 9 | enables | presents | achieves |
| 10 | reveal | achieves | outperforms |

Table 5: Ten most frequent present-tense verbs

While Civil Engineering and Computer Science differ significantly in their proportions of present-tense verbs ($\chi^2 = 295.13$, $p < 0.001$), the same verbs appear in the top four rankings for both. However, the verb "propose" appears as "proposes" in Civil and as "propose" in Computer. The most frequent collocations are "this paper proposes" and "this study proposes" in Civil Engineering, and "we propose" in Computer Science. These findings suggest that Computer Science tends to favor human grammatical subjects, whereas Civil Engineering tends to favor nonhuman grammatical subjects.

### 3.2.2 Aspect

The simple present and simple past are frequently used in all three fields. Table 6 presents the rates of perfect-aspect use among verbs.

| | Chem | Civil | Comp |
|---|---|---|---|
| Proportion of Verbs in the Perfect Aspect | 5.1% | 3.2% | 5.0% |
| Pres. + PEAS / Pres.Verb | 4.9% | 3.0% | 5.0% |
| Past + PEAS / Past Verb | 0.2% | 0.2% | 0.0% |

Table 6: Rate of perfect-aspect use

The overall frequency of perfect-aspect use per 100 words is 0.24 in Chemical Engineering, 0.13 in Civil Engineering, and 0.29 in Computer Science. While the overall frequency is low and shows little difference between Chemical Engineering and Computer Science, the frequency of perfect-aspect use in Civil Engineering is lower than that in Computer Science ($\chi^2 = 117.14$, $p < 0.001$). According to Biber et al. (2021), the ratio of verbs in perfect aspects in in academic texts is approximately 5% (pp. 459-460); thus, the rates observed in Chemical Engineering and Computer Science are consistent with this norm.

In both Chemical Engineering and Computer Science, the most frequent verb phrase in the perfect aspect is "have/has + emerge," which appears mostly in the initial section of the abstracts, as shown below:

> As geospatial data from web platforms becomes increasingly accessible and regularly updated, urban representation learning has emerged as a critical research area for advancing urban planning. (Zhang et al., 2026)

> Recently, dianionic heavy group 14 cyclopentadienides have emerged as promising ligands due to their higher charge density. (Sun et al., 2026)

It is frequently used to describe the research background.

#### 3.2.3 Voice

According to Biber et al. (2021), the proportion of verbs in the passive voice in academic texts is approximately 25% (pp.474-478). Consistent with this benchmark, the proportion in Civil Engineering was 25.1%, whereas the corresponding proportions in Chemical Engineering and Computer Science were considerably lower, at 9.4% and 1.9%, respectively. The corresponding values are shown in Table 7.

| | Chem | Civil | Comp |
|---|---|---|---|
| Overall Passive Rate | 9.4% | 25.1% | 1.9% |
| Past Passive Rate | 16.4% | 42.3% | 6.9% |
| Pres. Passive Rate | 8.4% | 20.2% | 1.6% |

Table 7: Rate of passive voice

In all three fields, passive voice accounts for a larger proportion of past-tense verbs than of present-tense verbs. The frequency of passive voice per 100 words is 0.5 in Chemical Engineering, 1.4 in Civil Engineering, and 0.1 in Computer Science. The high-frequency verbs associated with passive voice in Civil Engineering are "developed," "proposed," and "conducted." Examples of passive-voice use in both the past and present tenses are shown below:

> A finite element model incorporating the Johnson-Cook constitutive model was developed and validated to simulate the ACEAS's dynamic response under impact loading. (Zhu et al., 2026)

> First, a new probabilistic transformation method named the uniformity-constrained equal area approach is proposed to transform the evidence variables into random variables, enabling the most probable focal element to be identified with the help of the first-order reliability method. (Duan et al., 2026)

As shown in Section 3.2.1, the present tense accounts for nearly 95% of verbs in Computer Science, and the passive voice is rarely used in that tense. The phrase "we + propose" is the most frequent collocation for the high-frequency present-tense verb "propose." Manual concordance inspection of all 574 instances of "propose" showed that the subject was consistently "we."

We next examined the distribution of the pronoun "we" across the three disciplines. Biber et al. (2021) demonstrate that the personal pronoun "we" is used approximately 3,000 times per million words in academic texts (p. 333), which corresponds to 0.3 occurrences per 100 words.

A comparison with the frequencies shown in Table 8 reveals that "we" occurs more frequently in Computer Science and Chemical Engineering than in the academic-text benchmark. The corresponding frequencies are 1.05 and 0.98 occurrences per one hundred, respectively.

| | Chem | Civil | Comp |
|---|---|---|---|
| We + Present Verb | 0.84 | 0.05 | 0.99 |
| We + Past Verb | 0.14 | 0.01 | 0.06 |

Table 8: Distribution of active-voice constructions with "we"

Civil Engineering abstracts contain far fewer instances of "we" than the other two fields, suggesting that researchers in this discipline tend to avoid active-voice constructions with "we" in their abstracts.

### 3.3 Comparison of AI-generated descriptions and corpus-based analysis

The degree of agreement between the AI-generated descriptions and corpus findings varied across the three fields. The corpus-based analysis of Computer Science largely supports the AI findings by showing that researchers in the field almost always use the present tense in the active voice and make extensive use of the phrase "we propose." A notable difference from the AI descriptions is that 5.0% of verbs in Computer Science use the perfect aspect, which reflects the general usage patterns found in academic writing (Biber et al., 2021).

In contrast, the corpus findings for Chemical Engineering and Civil Engineering contradict the AI-generated descriptions. The corpus analysis reveals that approximately 88% of the verbs in Chemical Engineering are in the present tense, about 5% are in the perfect aspect, and less than 10% are in the passive voice. The personal pronoun "we" is frequently used as a subject. A more accurate conclusion would be that Chemical Engineering shares some important characteristics with Computer Science, particularly the strong preference for the present tense and the relatively frequent use of "we," although the two disciplines still differ notably in their use of passive constructions.

The corpus analysis indicates that Civil Engineering has the lowest proportion of present-tense verbs and the lowest proportion of verbs in the perfect aspect, as well as the highest proportion of verbs in the passive voice. The findings also suggest that the researchers may avoid constructions in which the first-person pronoun "we" is the grammatical subject. Furthermore, in active-voice construction, "this paper" or "this study" tends to serve as the grammatical subject.

## 4. Conclusion

This study analyzed research paper abstracts from three engineering fields—chemical engineering, civil engineering, and computer science—using AI and corpus-based methods, with a focus on tense, aspect, and voice.

Three research questions were addressed: (1) What usage differences exist among these three engineering fields according to AI analysis? (2) Does corpus-based analysis support the AI-derived findings? and (3) What implications do these findings have for ESP instruction in engineering?

Regarding RQ1, the AI-generated descriptions revealed that computer science abstracts frequently employ the simple present tense and active voice, notably featuring constructions with "we" as the subject. In contrast,

chemical engineering abstracts tend to utilize the past tense and passive voice. Civil engineering abstracts fall between these two, characterized by a mix of present and past tenses along with active and passive voices.

With respect to RQ2, the corpus analysis largely supported the AI findings for the computer science field, though it also revealed that 5% of the verbs appear in the perfect aspect. However, the corpus analysis contradicted the AI findings for chemical engineering and civil engineering. The analysis revealed a high frequency of present-tense verbs, a relatively low rate of passive constructions, and frequent active constructions with "we" in chemical engineering, while civil engineering exhibited the lowest rate of present tense verbs and the highest rate of passive voice usage. Unlike chemical engineering and computer science abstracts, civil engineering abstracts showed very limited use of active voice constructions employing "we."

Regarding RQ3, the findings of the study suggest that ESP instruction for chemical engineering students should not overemphasize the past tense and passive voice. Instead, active-voice constructions with the first-person pronoun "we" should also be introduced. Conversely, ESP instruction for civil engineering students should draw attention to the relatively frequent use of past-tense and passive constructions and to the comparatively infrequent use of first-person pronouns in the sampled abstracts.

This study has demonstrated that usage patterns of tense, aspect, and voice vary across engineering subfields. While AI is a valuable tool for ESP instruction, our findings suggest that it may fail to fully capture discipline-specific linguistic characteristics. Therefore, to deliver more effective, tailored instruction, combining AI analysis with corpus analysis is essential.

A limitation of this study is that the AI analysis relied on a single generative-AI system and three prompting sessions. In addition, the corpus included abstracts from only a limited number of journals and publication years and therefore did not represent each discipline in its entirety. Future studies could examine additional AI models, rhetorical moves within abstracts, and a wider range of engineering subdisciplines.

## 5. References

### 5.1 Bibliographical References


Apse, V., & Farneste, M. (2020). Fostering independent learning of English verb tense, aspect and voice forms at the tertiary level. *Baltic Journal of English Language, Literature and Culture, 10*. https://doi.org/10.22364 /BJELLC.10.2020.01

Biber, D. (1988). *Variation across speech and writing.* Cambridge University Press.

Biber, D. (1989). A typology of English texts. *Linguistics*, *27*, 3-43. https://doi.org/10.1515 /ling.1989.27.1.3

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

Caplan, N. A., & Johns, A. M. (2022). *Essential actions for academic writing: A genre-based approach*. University of Michigan Press.

Fuchs, R., & Werner, V. (2020). Tense and aspect in second language acquisition and learner corpus research. In R. Fuchs & V. Werner (Eds.), *Tense and aspect in second language acquisition and learner corpus research* (pp. 1-22). John Benjamins.

Glasman-Deal, H. (2010). *Science research writing: For non-native speakers of English.* Imperial College Press

Hinkel, E. (2002). *Second language writer's text: Linguistic and rhetorical features*. Lawrence Erlbaum Associates.

Hinkel, E. (2004). Tense, aspect and the passive voice in L1 and L2 academic texts. *Language Teaching Research*, *8*(1), 5-29. https://doi.org/10.1191/1362168804lr132oa

Hyland, K. (2001). Humble servants of the discipline? Self-mention in research articles, English. English for Specific Purposes, *20*(3), 207–226.

Hyland, K. (2002). Authority and invisibility: Authorial identity in academic writing. *Journal of Pragmatics, 34*, 1091-1112.

Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing*. Bloomsbury.

Huckin, T. (2001). Abstracting from abstracts. In M. Hewings (Ed.), *Academic writing in context: implications and applications* (pp. 93-103). The University of Birmingham Press.

Huddleston, R., & Pullum, G. K. (2002). T*he Cambridge grammar of the English language*. Cambridge University Press.

Ishikawa, Y. (2024, October 25-27). *A multi-dimensional analysis of research abstracts: Disciplines and text types* [Conference presentation]. The 5th Asia Pacific Corpus Linguistics Conference (APCLC), Jiao Tong University, Shanghai, China.

Nini, A. (2019). The Multi-Dimensional Analysis Tagger. In T. B. Sardinha & M. P. Veirano (Eds.), *Multi-dimensional analysis: Research methods and current issues* (pp. 67-94). Bloomsbury Academic.

Quirk, R., Greenbaum, S., Leech, G., & Svartvik, J. (1985). *A comprehensive grammar of English*. Longman.

Swales, J. M. (2004). *Research genres: Exploration and applications*. Cambridge University Press.

Swales, J. M., & Feak, C. B. (2009). *Abstracts and the writing of abstracts*. The University of Michigan Press.

Swales, J. M., & Feak, C. B. (2012). *Academic writing for graduate students: Essential tasks and skills* (3rd ed.). The University of Michigan Press.

Toutanova, K., Klein, D., Manning, C. D., & Singer, Y. (2003). Feature-rich part-of-speech tagging with a cyclic dependency network. *Proceedings of the 2003 Human Language Technology Conference of the North American Chapter of the Association for Computational Linguistics,* 252–259. https://aclanthology.org/N03-1033.pdf


### 5.2 Language Resource References


Duan, J., Zhao, Q., Yu, D., Zhou, Y., & Hong, J. (2026). A unified evidence variable transformation framework for structural reliability analysis with aleatory and epistemic uncertainties. *Engineering Structures, 360*

(1), 122850. https://doi.org/10.1016/j.engstruct.2026.122850

Sun, X., Hinz, A., Maier, S., Gillhuber, S., Jin, D., Saito, M., Weigend, F., Schlittenhardt, S., Moreno-Pineda, E., Wernsdorfer, W., Ruben M., & Roesky, P. W. (2026). Contrasting single-molecule magnet behaviour in dysprosium and terbium bis(stannolediide) complexes. *Nature Chemistry*, *18*, 872–881. https://www.nature.com/articles/s41557-026-02114-9.pdf

Zhang, Z., Xie, M., Balsebre, P., Huang, W., Luo, S., & Cong, G. (2026), UrbanMFM: Spatial graph-based multiscale foundation models for learning generalized urban representation, *IEEE Transactions on Knowledge and Data Engineering*, *38* (3), 2064-2078, doi: 10.1109/TKDE.2026.3656202.

Zhu, W., Li, T., Liao, X., Wu, Y., Peng, Y., & Wang, K. (2026). Crashworthiness analysis and multi-objective optimization of an annular cutting-type energy-absorbing structure for high-speed trains. *Engineering Structures, 360* (1), 122771. https://doi.org/10.1016/j.engstruct.2026.122771

## 6. Acknowledgements

This work was supported by JSPS KAKENHI Grant Number JP 26K04064 and 23K00673.