

A Corpus-Based Model for Successful Computer Science Research Rhetoric

R R Simanjuntak, D M Bourne, I I Dewi, Muhartoyo, and A A Sasongko

English Department, Faculty of Humanities, Bina Nusantara University, Jakarta, Indonesia, 11480

*Corresponding author email: risarsimanjuntak@binus.edu

Abstract. Many good research publications were not getting the attention of prospective readers due to a lack of structure and persuasive arguments. This paper investigated the rhetorical strategies to identify the generic structure and language features in successful research rhetoric. Research abstracts as representations of the research papers were used as corpus. A corpus-based study was conducted to identify the recurring patterns of lexical use and lexical structures in research rhetoric. The results showed information used in abstracts to contain an introduction to the issue followed by research purpose, methods, results, conclusion consecutively. Interestingly, sometimes abstracts directly started with the purpose of research and are followed by a discussion on the research topic. Some abstracts did not include a conclusion of the research. Certain lexical expressions were also used in presenting the thoughts and arguments of the authors. These recurring lexical items showed the common practice of research authors in Computer Science. Following the findings, a model of successful abstract was presented. Implications for further study were also discussed.

Keywords: *A Corpus-Based Model, Computer Science*

1. INTRODUCTION

In the breadth of knowledge, hundreds or even thousands of articles are published every year. Amongst these articles, only some have been able to read and appreciated by their audience. This is because the amount of available publications has become such information-overload to the readers, Along with the limited time available, most often readers would only browse the abstracts and not the entire articles [1, 2]. Avid readers would also become selective in what they consider to be the most suitable articles to read [1]. An abstract nonetheless becomes the "window-of-display" for a research article [1, 3] and only convincing abstracts would successfully get the attention of the readers [1, 2]. However, not many authors have the knowledge and skills to write good abstracts.

An abstract is a concise, condensed version of the whole article/paper. Abstract in general has been considered as a specific genre of its own [4, 5, 6]. It has increasingly become an important facet in academic endeavors [5, 6, 7, 8]. A good abstract needs to consist of the key communicative purposes, or rhetorical moves [5, 9], as well as the realizations of these purposes through the use of convincing and engaging language [10, 11]. A good abstract therefore need to be well-structured to help readers follow the ideas easily. A good abstract also needs to be argumentative as it represents the main idea proposed in the full article/paper [12]. Only a good and convincing Abstract will attract the attention of readers in further reading the full article/paper [1].

Previous research has identified some problems with research writing. For non-native speakers of English, there are problems in grammar as well as in structuring the ideas [13]. Besides, different disciplines have been identified to have certain rhetoric in presenting epistemic qualities. A model of successful abstract would be very beneficial to help authors in publishing their research.

This present research aims at identifying the structure and language used in successful abstracts to offer a model of good and convincing abstract.

2. CORPUS-BASED RESEARCH IN ACADEMIC TEXTS

2.1. Corpus-based research

Corpus as a systemized large collection of English usage has been widely analyzed to identify the structure and elements of texts [14, 15, 16, 17]. Corpus takes its collections from authentic sources, such as newspapers, magazines, or even people's social conversations. The real use of language reveals the true nature of words, be it to inform or to convince the readers [18].

Corpus becomes the penultimate source of information for any learning purposes [19, 20]. For novice authors and language learners, the corpus may also provide references on how language is used in a specific context [14, 16, 21, 22, 23]. However, not many kinds of research have dealt with different rhetoric and types of texts [16, 22, 23]. A specific corpus, such as a corpus of abstract, would provide

the generic/formal structure of abstract [24, 25, 26, 27, 28].

2.2. Structure of research article abstracts

An Abstract consists of at least four important moves: Purpose, Method, Results, and Discussion [29]. These are respectively the sections found in research articles/reports. Current studies also mentioned Introduction move as always appear in Abstract [10, 30]. The five sections model (Introduction-Purposes-Method-Product-Conclusion) has become the common structure used in research article abstracts and many studies have confirmed these across disciplines [27, 31, 32].

Research abstracts for Biology journal articles, for example, used the construction of “X is Y” to present the finding [33]. On the contrary, abstracts for educational journal articles often contain expressions, such as “it is suggested...” in the

conclusion part [27]. In Computer Science, however, the authors use “we” to refer to themselves and to include their readers in the discussion [31]. The language used helped authors to not only present their arguments but also to convince their readers by engaging themselves with their readers.

2.3. Research Methods

2.3.1 Corpus

The corpus for this study as the sources for this qualitative study were 90 research Abstracts from Computer Science of repuTABLE journals. There are 6 journals from 4 publishers used as the reference corpus, which the details provided in the following TABLE.

The distribution of each journal can be seen from the following TABLE:

TABLE 1. List of Journals for Computer Science Articles

No	Name of Journals	Name of Publishers	Year of Publication	Number of Articles
1	Transactions on Cloud Computing	IEEE	2016-2020	135
2	Transactions on Services Computing	IEEE	2016-2020	131
3	Network Science	CUP	2016-2020	134
4	The Computer Journal	OUP	2016-2020	130
5	Information and Management	Elsevier	2016-2020	129
6	Information Systems	Elsevier	2016-2020	131
	Total number of articles			790

The first publisher selected as the data source is IEEE, which stands for the Institute of Electrical and Electronics Engineers. IEEE previously is known as AIEEE (American Institute of Electrical and Electronics Engineers), which was founded in New York in 1884. Up to this present day, IEEE becomes the world's largest technical professional association, as stated in its website: “IEEE is the world’s largest technical professional organization dedicated to advancing technology for the benefit of humanity” (<https://www.ieee.org/about/vision-mission.html>).

Elsevier is selected for the reason of its standing to academic and research publications. The infamous SCOPUS indexing has made Elsevier a recognized publisher. Every journal under SCOPUS indexing undergoes at least a double-blind review process. With this indexing quality and rigorous process, articles published under Elsevier are considered as worth-reading. Journals selected in Elsevier for this present study were selected based on their H-index,

or based on the influence they are making to the Computer Science discipline. The H-index is calculated by the number of citations made for each of the articles in the journals.

The Information and Management (IM) and The Information Systems (IS) journals are selected based on this reason The Computer Journal (TCJ) from Oxford University Press (OUP) is highly recommended since the journal is the oldest and of high reputation and therefore included in CSJ, which positions this journal at the top of its respective field with a long history.

2.3.2 Instrument Used for Generating Data

AntConc is a freeware tool kit for corpus analysis. This tool can be obtained freely from <http://www.laurenceanthony.net/software.html>. For this present study the latest (2019) version of AntConc is used as the instrument for data collection, which is AntConc 3.5.8.

2.4. Research Procedure

2.4.1 Research procedure follows the steps as seen in Fig.1:

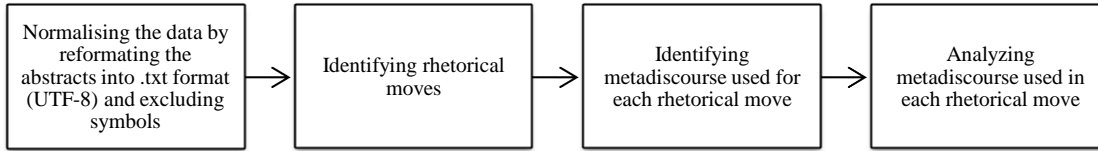


FIGURE 1. Research Procedure

First, all abstracts were reformatted, from pdf into txt files. The texts were normalised, excluding all symbols. When normalised, AntConc. would generate the data, which would be used for further analysis.

Second, manual identification was carried out to identify sections in the abstract. According to the classifications, an abstract could consist of Introduction, Purpose, Method, Product, and Conclusion [6,8,17]. Each function could be expressed through clauses or sentences. When each section was identified according to the function, a building block for that abstract was described. Ideally, an abstract would then have a building block containing five sections of I (Introduction), P (Purpose), M (Method), Pr (Product), and C (Conclusion), in which each section could be explained below:

- Introduction (I), which has the function to introduce the issue/problem in the research
- Purpose (P), which has the function to state the purpose of the research
- Method (M), which has the function to describe the steps in carrying out the research
- Product (Pr), which has the function to present the results and findings of the research, and
- Conclusion (C), which has the function to conclude the research

Third, language used for each section was identified. Using AntConc, words used in each of the rhetorical move is identified. In addition to the use of words and the frequency of use, AntConc would also identify whether certain markers, called metadiscourse, were used in each of the section. Metadiscourse is defined as the commentaries used by authors in their writings [7, 13] to arrange ideas and information, as well as to help readers in understanding the meaning of the writing. Metadiscourse is often used to highlight ideas or to support author’s opinion.

Finally, analysis on the use of metadiscourse was conducted to reveal the characteristic of Computer Science research abstract as a genre as well as presenting a model for a successful abstract.

3. RESULTS AND DISCUSSION

Results will show the structure of research abstract and the language used for each of the sections. The last part of this section is to show a model for successful research abstract.

3.1. Structure in Computer Science research abstracts

From the corpus, it can be seen that there is a pattern in presenting Introduction, Purpose, Method, Product, and Conclusion sections in Computer Science research abstracts. The pattern of sections used in Computer Science research abstracts could be seen in TABLE2:

TABLE 2. Structures found in Computer Science research abstracts

No.	Move Sequence	No.	Move Sequence
1	I>P>C	11	P>I>M
2	I>P>M	12	P>I>M>Pr
3	I>P>M>C	13	P>I>M>Pr>C
4	I>P>M>C>Pr	14	P>M>C
5	I>P>M>Pr	15	P>M>Pr
6	I>P>M>Pr>C	16	P>M>Pr>C
7	I>P>Pr>M	17	P>Pr>M
8	I>P>Pr>M>C	18	M>P>Pr
9	I>M>P		
10	I>M>P>Pr>C		

As can be seen in TABLE 2 there was more than one structure used in writing an abstract. The symbol ">" meant followed by. In an abstract where I>P>M>C>Pr structure was used, I (Introduction) was followed by P (Purpose), which was followed by M (Method), then by Pr (Product), and then by C (Conclusion) section. The data also showed that 10 out of 18 structures started with the I (Introduction) section, and 8 structures directly wrote P (Purpose) in the abstracts with or without the I (Introduction) section. It was also seen in the data that the M (Method) section would always be written in the abstract as well as the P (Purpose) section. The least used section in the Computer Science research abstract seemed to be the C (Conclusion) section. From this data, it could be said that a Computer Science research abstract would commonly consist of P (Purpose) and M (Method) sections.

The data also revealed the movement of the sections in the structure of Computer Science research abstracts. In a conventional structure, the I (Introduction) section would be followed by the P (Purpose) section. However, it would also be possible to write the P (Purpose) of the research first before explaining the issue or the context of the research. In this case, the movement of I>P was inverted into P>I. There were three structures in

which the abstracts were written with P>I movement. Another interesting construction, in which the movement of the section was inverted, could be seen in M>P, and Pr>M, which were used in consecutively 2, and 3 structures of abstracts. The only inversion which was rarely used was C>Pr, in which Conclusion was written before the P (Product), which was the results or findings of the research.

It could be seen from the data that the Purpose and Method always appear in an abstract. These sections always occur in every good and convincing research article abstract in Computer Science. The implication for this was that in a successful model for abstract, both sections have to be explicitly mentioned. The results support previous studies, which found a five-move structure for the abstracts [27, 31, 32, 33]. However, this present study does not go in-line with previous studies in that Introduction does not always appear in Computer Science research abstracts [10, 30].

3.2. Language used in Computer Science research abstracts

The language used for writing each section in successful Computer Science research abstracts can be seen in TABLE 3 below:

TABLE 3. Language used to identify different sections in Computer Science research abstract

Sections	Language used to frame the move
Introduction	in this paper
Method	First, second, finally
	instead
	unlike
	consequently
Conclusion	the paper ends
	in our study

As can be seen in the data, there were certain expressions commonly used for different sections in Computer Science abstracts. I (Introduction) sections specifically used the expression, "In this paper". This expression identified a section in which the author presented the issue being researched. Meanwhile, M (Method) section often used expressions of orders, such as: first, second,, finally as common expressions in describing the procedures of the research. There were also expressions such as instead, unlike, and

consequently, which were used when authors explain the method used in their research. Semantically, these words could be inferred as having a contrasting value, in which the author comparing the present research to the previous research.

Through corpus-based analysis, it could be seen that the language was not only used to help the author arrange the information but could also be used to address readers' expectations. This finding was in line with previous studies, in which

comments used by authors, or metadiscourse, were an important facet in a genre [5,7]. Corpus-based analysis conducted in this present study was able to reveal the common preference of a community of discourse [5,7,12].

Corpus analysis also showed a certain preference for linking each of the sections. These are called transition markers as part of interactive metadiscourse markers (IMM) [14,24]. The use of transition markers in Computer Science research abstracts could be seen in TABLE 4 below:

TABLE4. Transition markers used for linking each section in research abstracts

Transition markers	Frequency
And	488
Also	26
But	17
thus	12
in addition	7
as well as	6
as well	2
furthermore	2
besides	1
Total	524

As can be seen from TABLE 3, the most frequently used transition marker is “and” (used in 89 out of 90 abstracts). In specific, using corpus analysis metadiscourse markers for each section of the research abstracts. These markers are called frame

markers, in which each of the markers specifies a certain section in the abstract [6,8]. The frame markers used for each section of the research abstracts could be seen in the following FIGURE 2:

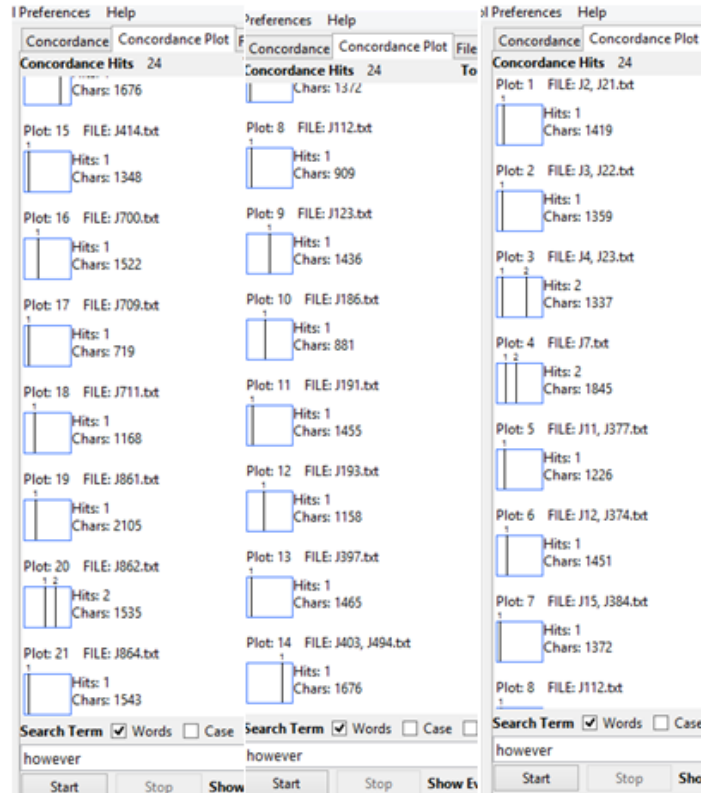


FIGURE 2. Concordance plot for transition marker “however”

As can be seen from FIGURE 2, the concordance plot showed in the corpus where the word “however” as a frame marker was used in the abstracts. The word “however” was used to help readers identify the transition of ideas from one section to another. This marker is used in differentiating these particular moves from another: I (Introduction) from P (Purpose) moves, P (Purpose) from M (Method) moves, Pr (Product) from C (Conclusion) moves. From the concordance plot, it can be seen that the occurrence of this marker is usually at the beginning of the abstract, such as the I (Introduction) and P (Purpose) sections. The finding suggests that this marker is used to help readers identify the importance of the issue and the purpose of the research. The word “however”, was never used to differentiate M (Method) move from Pr (Product) move.

3.3. The model for successful Computer Science research abstract

Based on the findings, the model for successful Computer Science is presented below:

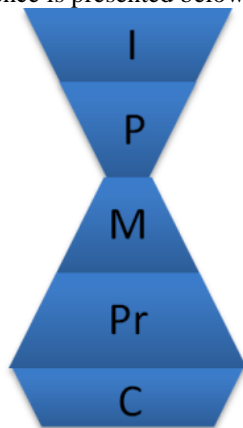


FIGURE 3. The model for successful research abstract in Computer Science

From FIGURE 3 above it can be seen that a successful Computer Science abstract consists of five sections, namely: Introduction, Purpose, Method, Product, and Conclusion. The model informs that the amount of information is relative to the hour-glass like structure, in which the length of sentences in Purpose and Method are relatively fewer compared to Product section. On the other hand, Introduction has more sentences when compared to Conclusion.

4. CONCLUSION AND IMPLICATION

From the analysis, it can be concluded that successful abstracts need to have structural and language elements. In the structure, a successful abstract requires the use of five sections, namely

Introduction, Purpose, Method, Product, and Conclusion. The sections which need to always appear in an abstract are Purpose and Method. In the model of a successful abstract, the length of sentences for each section can be illustrated to look like an hour glass, in which the Introduction usually is longer compared to Conclusion. The language model in a successful abstract consisted of the use of markers to identify the sections and to link one section to another.

ACKNOWLEDGMENT

The authors would like to express their gratitude to Research & Technology Transfer Office of Bina Nusantara University, which has provided the sufficient funding for our research entitled *Multidimensional Analysis in Learner Corpus: Embracing Identities in Learner Writing* with the grant number 026/VR.RTT/IV/2020.

REFERENCES

- [1] O'Connor, R. 2009. *Writing scientific research articles: Strategy and steps*. West Sussex: Wiley-Blackwell.
- [2] Cutting, J. & Fordycde, K. 2020. *Pragmatics: A Resource Book for Students (Routledge English Language Introductions) 4th Edition*. London: Routledge.
- [3] Sanz, L. S. 2014. Lost (and gained) in translation: A contrastive (English/Spanish) analysis of rhetorical and lexicogrammatical patterns in Sociology research article abstracts. In Bondi, M., and Sanz, L.S. (eds.). *Abstracts in academic discourse: Variation and change* (pp.85-110). Bern: Peter Lang AG.
- [4] Van Dijk, T. A. 1980. *Macrostructures*. Hillsdale, NJ: Lawrence Erlbaum.
- [5] Swales, J. M. 1990. *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.
- [6] Hyland, K. 2004. *Genre and second language writing*. Ann Arbor, MI: University of Michigan Press.
- [7] Swales, J. M., & Feak, C. B. 1994. *Academic writing for graduate students-essential tasks and skills: A course for nonnative speakers of English*. Michigan: The University of Michigan Press.
- [8] Hyland, K., & Tse, P. 2004. Metadiscourse in academic writing: a reappraisal. *Applied Linguistics*, 25, 156-177. doi:10.1016/j.esp.2004.02.002.

- [9] Samraj, B. 2005. An exploration of a genre set: Research article abstracts and introductions in two disciplines. *English for Specific Purposes*, 24, 141–156. doi:10.1016/j.esp.2002.10.001.
- [10] Arsyad, S., & Arono. 2018. *Memahami dan menulis abstrak artikel jurnal*. Bogor: Halaman Moeka Publishing.
- [11] Salazar, D. 2014. *Lexical bundles in native and non-native scientific writing: Applying a corpus-based study to language teaching*. Amsterdam: John Benjamins.
- [12] Conrad, S. & Biber, D. 2009. *Register, Genre, and Style*. Cambridge: Cambridge University Press.
- [13] Biber, D. & Barbieri, F. 2007. Lexical bundles in university spoken and written registers, *English for Specific Purposes*, 26, 263–286.
- [14] Hyland, K. 2008. Genre and academic writing in the disciplines. *Language Teaching*, 41(4), 543-562.
- [15] Firth, J. R. 1962. *Studies in linguistics analysis*. Oxford: Basil Blackwell.
- [16] Hanks, P. 2013. *Lexical analysis: Norms and exploitations*. Cambridge: MIT Press.
- [17] Flowerdew, L. 2012. Corpora in the classroom: An Applied Linguistic perspective. In K. Hyland, C. M. Huat, & M. Handford, *Corpus applications in Applied Linguistics*, pp. 208-224. London: Continuum.
- [18] Cheng, A. 2018. *Genre and Graduate-Level Research Writing*. Michigan: University of Michigan Press.
- [19] Adolphs, S. 2019. *Introducing Text Analysis*. Oxfordshire: Taylor Francis.
- [20] Leech, G, Rayson, P & Wilson, A. (2001). *Word frequencies in written and spoken English based on the British National Corpus*. London: Longman.
- [21] Belcher, D & Nelson, G. (Eds.). 2013. *Critical and Corpus-based Approaches to Intercultural Rhetoric*. Michigan: The University of Michigan Press, Ann Arbor
- [22] Tankó, G. 2017. Literary research article abstracts: An analysis of rhetorical moves and their linguistic realizations. *Journal of English for Academic Purposes*, 27(1), 42-55.
- [23] Kanoksilapatham, B. 2013. Generic characterisation of Civil Engineering research article abstracts. *3L: The Southeast Asian Journal of English Language Studies*, 19(3), 1 – 10.
- [24] Pho, P. D. 2008. Research article abstracts in applied linguistics and educational technology: a study of linguistic realizations of rhetorical structure and authorial stance, *Discourse Studies*, 10(2), 231–250. doi: 10.1177/1461445607087010.
- [25] Yaylı, D. & Canagarajah, A. S. 2018. The Processes Behind RA Introduction Writing Among Turkish Arts and Science Scholars. *Key Issues in English for Specific Purposes in Higher Education: English Language Education*, 11, 111-128.
- [26] Swales, J. M. & Santos, M. B.D. The futures of EAP genre studies: A personal viewpoint. *Journal of English for Academic Purposes*, 38, 75-82
- [27] John M. Swales Santos, M. B. D. 1996. The textual organization of research paper abstracts in Applied Linguistics. *Text*, 16(4), 481–99.
- [28] Cunningham, K. J. 2017. A phraseological exploration of recent mathematics research articles through key phrase frames. *Journal of English for Academic Purposes*, 25, 71-84.
- [29] Kelly J. Cunningham Ozmen, K. S. 2016. Rhetorical analysis of the doctoral abstracts on English Language Teaching in Turkey. *Journal on English Language Teaching*, 6 (1), 25-35.
- [30] Stotesbury, H. 2013. Evaluation in research article abstracts in the narrative and hard sciences. *Journal of English for Academic Purposes*, 2, 327-341. Doi: 10.1016/S1475-1585(03)00049-3.
- [31] Carrió-Pastor, M. L. 2020. *Corpus Analysis in Different Genres: Academic discourse and learner corpora*. New York: Routledge.