



Çapkın, Ç. (2022). Trends in genre analysis articles on scientific abstract structures: A Quantitative content analysis. *Journal of Librarianship and Information Science*, [in press]. <https://doi.org/10.1177/09610006221132274>

Under SAGE's Green Open Access policy, this is the **Accepted Version** of the article.

Reuse is restricted to non-commercial and no derivative uses.

You may visit SAGE for published version of the article.

<https://doi.org/10.1177/09610006221132274>



Journal of Librarianship and Information Science

Trends in Genre Analysis Articles on Scientific Abstract Structures: A Quantitative Content Analysis

Çağdaş Çapkın*

Abstract

Genre analysis is a methodologically prominent approach to segmenting a scientific abstract into discourse units. Genre analysis studies on scientific abstract structures have valuable outputs not only for secondary information services, including bibliographic databases and online services but also for scientific communication and library and information science (LIS) education. However, trends of research on this topic have not been investigated yet. This study identifies research trends and reveals knowledge gaps and research opportunities in genre analysis articles on scientific abstracts. For this purpose, Web of Science and Scopus databases were searched to identify the articles. According to the study selection criteria, 75 articles were included in the quantitative content analysis. It was found that the most frequently studied genres were research articles (73.3%), proceedings (12%), and thesis/dissertations (8%). The sample size of the corpus ranged from 5 to 4214 abstracts ($M = 223.8$, $MD = 94$, $SD = 523.8$). The most frequently used abstract genre models were (Hyland, 2000) ($n = 24$), (Swales, 1990) ($n = 18$) and (Santos, 1996) ($n = 13$). In 18.7% of articles, at least one of the abstract standards was cited. Approximately, two-thirds of the articles were comparative. Languages (44.7%), disciplines (25.5%), genres, and native/non-native authors (8.5%) were compared most frequently. English was the most frequently studied language, both individual (72.4%) and comparatively (25.9%). The results of this study suggest that the LIS community, as well as applied linguistics, can seize the opportunity to address gaps in academic genres, disciplines, and languages. In addition, future studies are expected to have generalizable results to assist the scientific communication and LIS communities.

Keywords: Scientific abstracts, genre analysis, move analysis, step analysis, discourse analysis

* **Corresponding author:** Çağdaş Çapkın, Department of Information and Records Management, Cankiri Karatekin University, Uluyazi Campus, Cankiri 18100, Turkey. Email: cagdascapkin@karatekin.edu.tr

Introduction

Soon after World War II, scientific literature increased at an “explosive” rate, responding to the stimulus of sharp government expenditures on research and development for decades (Garfield, 1979: 6). The annual compound growth of journals was 3.3% between 1900 and 1996 when this period was divided into sub-periods of various socio-political developments (between 1900 and 1944, between 1944 and 1978, between 1978 and 1996) it was 3.3%, 4.68%, and 3.31%, respectively (Mabe and Amin, 2001: 156). A recent study found that the compound growth of scientific journals rose to 4.7% between 1986 and 2013, indicating the second journal growth explosion period in history (Gu and Blackmore, 2016: 714).

Due to the growing scientific communication, researchers have needed to follow the literature with selective readings. To meet this need, researchers have used scientific abstract which is “an abbreviated, accurate representation of a document” consisting of purpose, methodology, results, and conclusions (Weil, 1970). In the words of linguists, the abstract has been used to “sell” a full-text paper (Samar et al., 2014: 770). In addition, abstracts are valuable in the creation of information retrieval systems and secondary information services, as they do not have copyright restrictions (ISO, 1976: 2). So, editors of reputable journals read abstracts for preliminary consideration (Alspach, 2017: 12), and often check abstracts for compliance with the abstract writing guidelines before accepting articles for publication (Khansari et al., 2016: 39).

Since the paragraph abstract does not have predefined labels found in the structured abstract, its discourse units have been the focus of considerable research in recent decades. On the other hand, genre analysis is a prominent methodological approach to segmenting the abstract into discourse units. According to Swales and Feak (2009a: 1), simply, “genre is a name for a type of text or discourse designed to achieve a set of communicative purposes”. For example, scientific communication tools such as research papers or conference papers are genres, and their various parts, such as abstract, introduction, and conclusion are sub-genres (p.1). A part-genre consists of moves which are text extensions that do specific jobs (p.5). So, move and step (sub-moves that carries more fine-grained functional meanings) analysis as an analytic approach within genre analysis, is used for academic and professional genres to

develop a new model by defining moves and steps or to validate an existing model (Cotos, 2018). Genre analysis is also a key to a theory of English for Special Purposes for teaching students from different disciplines how moves and steps in genres work in academic writing (Dudley-Evans, 2000). Thus, genre analysis studies on scientific abstract structures have pedagogical implications for scientific communication and Library and Information Science (LIS) communities.

Although genre analysis is a prominent methodological approach to segmenting a scientific abstract into discourse units, research trends in this topic have not been investigated yet. Therefore, this study investigates trends in genre analysis articles on scientific abstract structures, including key corpus attributes (genres, languages, disciplines, and periods of abstracts), and methodologic components (genre models, sample sizes, and inter-coder agreement approaches). The results of this study are expected to be beneficial for researchers as they highlight future research opportunities on scientific abstract structures that have the potential to contribute to genre theory.

A brief background

The criteria for acceptable abstracts have long been a research topic in the LIS literature. To determine the types, contents, and formats of abstracts, Borko and Chatman (1963) analyzed 130 instructions published by various indexing and abstraction services to guide abstracters. They found that (1) abstract types were informative (discusses the research in active voice and past tense), indicative (discusses the article which describes the research in passive voice and presents tense,) and "informative and indicative", (2) it was generally agreed that abstracts should contain purpose, methods, results, and conclusions, (3) abstract lengths varied between 100 and 500 words by communication types.

According to Weil (1970) as a Chairman of the *American National Standards Institute* (ANSI, now *National Information Standards Organization* [NISO]) Committee Z.39, a proposed ANSI standard for writing abstracts was drafted by the *International Union for Fundamental and Applied Physics*, the *American Institute of Physics* and UNESCO primarily to guide authors and editors in 1969. Next, the draft was sent for evaluation to all members of Committee Z.39, as well as to the many individuals and

groups interested in writing abstracts, including representatives of the social sciences and humanities. Finally, ANSI/NISO Z39.14, "American National Standard for Writing Abstracts" was first issued in 1971, followed by *International Organization for Standardization's* ISO 214, "Documentation - Abstracts for Publications and Documentation" which was largely based on ANSI Z39.14-1971 in 1976 (ANSI/NISO, 2015: V).

Although ANSI/NISO Z39.14 (1971) and ISO 214 (1976) differed in some details, they had similarities with the findings of Borko and Chatman (1963) in terms of important definitions and determinations. According to both standards, (1) abstract types were informative, indicative, and "informative-indicative", (2) abstracts should be used in journals, reports and theses, monographs and proceedings, and patents, (3) all three abstract types, especially informative, should consist of purpose, methodology, results and conclusions sequences as much as possible; however, to give information more quickly, the most important results and conclusions could place first, followed by other details, findings, and methodology, (4) a short abstract should be structured in a single, combined paragraph, but for long abstract multiple paragraphs can be used, (5) verbs should be in the active voice whenever possible, and (6) while the maximum abstract length was between 100 and 500 words by communication types, less than 250 words could be adequate for most papers.

Abstract writing has changed over time with the dynamics of scientific communication by responding to the needs of the scientific communities. For example, due to the lack of systematic structure and standard content in the abstracts, the *Ad Hoc Working Group on the Critical Evaluation of Medical Literature* (1987) proposed a structured abstract guideline consisting of seven distinct labels such as (1) objective, (2) design, (3) setting, (4) patients or other participants, (5) intervention(s), (6) measurements and main results, and (7) conclusion(s) for specific types of communication in the medical literature. This proposal is followed by a new proposal for structured abstract of review articles consisting of six labels such as (1) purpose, (2) data identification, (3) data extraction, (4) results of data synthesis, and (5) conclusions (Mulrow, Thacker and Pugh, 1988). The proposals were first accepted by the *Annals of Internal Medicine*, followed by various reputable medical journals

(Haynes et al., 1996). In response to these developments, structured abstract has also been added to the revised ANSI/NISO Z39.14:1997 (1997).

Although the objectives, types, and structures of scientific abstracts were defined by standards (ANSI/NISO Z39.14, 1971; ISO 214, 1976) before Swales (1981) proposed the move analysis, the structures of abstracts have also been extensively studied by applied linguistics, and various models have been proposed to explain the rhetorical structure of scientific abstracts. For example, two similar five-move abstract structuring models in Table 1 were proposed by analyzing abstracts from applied linguistics and multiple disciplines. However, unlike LIS, which focuses on “what should be written” in a scientific abstract with predefined rules or standards, applied linguistics focuses on “what is written” (Montesi and Urdiciain, 2005: 65). For example, while the standards recommend writing purpose, methodology, results, and conclusions for the research article abstracts (ANSI/NISO Z39.14, 2015; ISO 214, 1976), some genre analysis studies on particular domains recommend obligatory and optional moves after determining the frequency of moves in the genre (e.g., Kanoksilapatham, 2005: 272; Santos, 1996). On the other hand, it has long been known that the types and length of the abstracts affect information retrieval performance (Liddy, 1991). So, from a LIS perspective, recommending optional to write some of the moves for certain domains may not be accepted as it shortens the abstract and reduces information retrieval performance. It is also claimed that the IMRD type (a miniature version of the full article) abstract preferred by abstracting services may not appeal to all readers and that the reader may be attracted by telling something from the author's intellectual and research journey rather than announcing important findings first (Alharbi and Swales, 2011: 84). Although LIS and applied linguistics approach the objectives of scientific abstracts differently, the results of genre analysis studies are valuable in terms of revealing how the scientific abstract structure differs by disciplines, languages, and genres.

Table 1. Example for five-move abstract structuring models

Applied Linguistics (Santos, 1996: 485)	Multiple disciplines (Hyland, 2000: 67)
Move 1 - Situating the research	Introduction: Establishes the context of the paper and motivates the research or discussion
Sub-move 1A - Stating current knowledge and/or	
Sub-move 1B - Citing previous research and/or	
Sub-move 1C - Extended previous research and/or	
Sub-move 2 - Stating a problem	
Move 2 - Presenting the research	Purpose: Indicates purpose, thesis, or hypothesis, outlines the intention of the paper
Sub-move 1A - Indicating main features and/or	
Sub-move 1B - Indicating the main purpose and/or	
Sub-move 2 - Hypothesis raising	
Move 3 - Describing the methodology	Methods: Provides information on design, procedures, assumptions, approach, data, etc.
	Results: States main findings or results, the argument, or what was accomplished.
Move 4 - Summarizing the results	Conclusion: Interprets or extends results beyond scope of the paper, draws inferences, points to applications, or wider implications.
Move 5 - Discussing the research	
Sub-move 1 - Concluding and/or	
Sub-move 2 - Giving recommendations	

Research questions

The purpose of this study is to identify research trends and to reveal knowledge gaps and research opportunities in genre analysis articles on scientific abstract structures.

For this purpose, the study addresses the following research question:

RQ1: Which journals published genre analysis studies on abstracts?

RQ2: What were the characteristics of the corpus (disciplines, genres, languages, number of abstracts and period)?

RQ3: Which inter-coder agreement approaches reported?

RQ4: Which genre models were reported?

RQ5: What comparisons were made?

RQ6: How many articles reported cross-language differences as "cultural"?

RQ7: How many articles reported percentages, sequences, lengths, and sentence voices (active and passive) of moves?

Methods

Data collection

Web of Science (SCI-EXPANDED, SSCI, A&HCI, and ESCI) and Scopus citation indexes were searched to identify articles on August 31, 2021. The searches were limited to journal articles in English published until 2021. Two searches on the title and topic (containing title, abstract, and keywords fields) were combined with the OR operator to lower the recall and increase the precision. The keywords "genre", "step", "move", "rhetoric", "discourse", and "abstract" were used for the title search. "Genre analysis", "step analysis", "move analysis", "rhetorical analysis", "discourse analysis", and "abstract" keywords were used for topic search. The exact search strategies were as follows:

Web of Science search strategy: TS= (("genre analysis" OR "step analysis" OR "move analysis" OR "rhetorical analysis" OR "discourse analysis") AND abstract*) OR TI= ((genre* OR step* OR move* OR rhetoric* OR discourse*) AND abstract*)

Scopus search strategy: (TITLE-ABS-KEY (((("genre analysis") OR ("step analysis") OR ("move analysis") OR ("rhetorical analysis") OR ("discourse analysis")) AND (abstract*))) OR (TITLE ((genre* OR step* OR move* OR rhetoric* OR discourse*) AND (abstract*))))

In the study selection phase, Web of Science (n = 451) and Scopus (n = 498) records merged with Mendeley Desktop software. After merging duplicates (n = 271), the remaining 678 records were screened on the title and abstract, resulting in 586 irrelevant records exclusion. Next, a total of 92 records were screened in full text, and 16 records were excluded that did not report a genre model. Two duplicate (redundant) articles were detected (same data published in two journals). One of the duplicate articles was randomly excluded. Finally, 75 records (1) on paragraph abstract, and (2) reported at least one genre model were included in the data analysis.

Data analysis

Quantitative content analysis, which is "a research method in which features of a text are systematically categorized and recorded so that they can be analyzed" (Coe and Scacco, 2017: 1), was used to answer the research questions. To analyze full-text articles, the following coding scheme was developed: journal name, article year, author

count, cited standard (available or not available - N/A), corpus discipline(s), genre (article, conference, thesis, etc.), corpus period, number of abstracts in the corpus, abstract language(s) in the corpus, reported corpus population size (available or N/A), reported corpus sample size calculation (available or N/A), reported inter-coder agreement (Cohen's kappa, percentage agreement, and N/A), reported genre model(s), reported comparison(s) (languages, disciplines, genres, etc.), reported abstract lengths (available, not exact or N/A), reported move percentage (available or N/A), reported move sequences (available or N/A), in language comparative article conclusion was "cultural" without any data (yes or no). The coded data were analyzed using descriptive statistics.

Study selection and data coding were done by the author. To minimize bias in these phases, assistance was obtained from a Ph.D. candidate experienced in genre analysis on scientific abstracts. A 100% agreement was achieved between the two coders in the study selection phase, by screening the title and abstract in randomly selected search results (n = 50). In the phase of data coding from randomly selected full-text articles (n = 15), 96% agreement was achieved.

Results

Journals and articles

According to the study selection criteria, 50 journals published a total of 75 articles between 1990 and 2020. As seen in Figure 1, 21.3% (n = 16) of the articles were published in 22 years between 1990 and 2011, and 78.7% (n = 59) in nine years between 2012 and 2020. Approximately 75% (n = 56) of articles were published in linguistics, language, and related discipline journals, 9.3% (n = 7) in LIS, and the remaining in other disciplines such as interdisciplinary, medical, social sciences & humanities, applied sciences, communication and media, and education.

While 74% (n = 37) of these journals published only one article, 26% (n = 13) between two and five. Journals that published more than one article were *Text and Talk* (n = 5), *3L: The Southeast Asian Journal of English Language Studies* (n = 5), *English for Specific Purposes* (n = 4), *Asian ESP Journal* (n = 3), *English Language Teaching* (n = 3), *International Journal of Applied Linguistics and English Literature* (n = 3), *Journal of English for Academic Purposes* (n = 3), *Discourse and Interaction* (n

= 2), *Discourse Studies* (n = 2), *GEMA Online Journal of Language Studies* (n = 2), *Journal of Language and Linguistic Studies* (n = 2), *Languages in Contrast* (n = 2), and *Information Processing & Management* (n = 2).

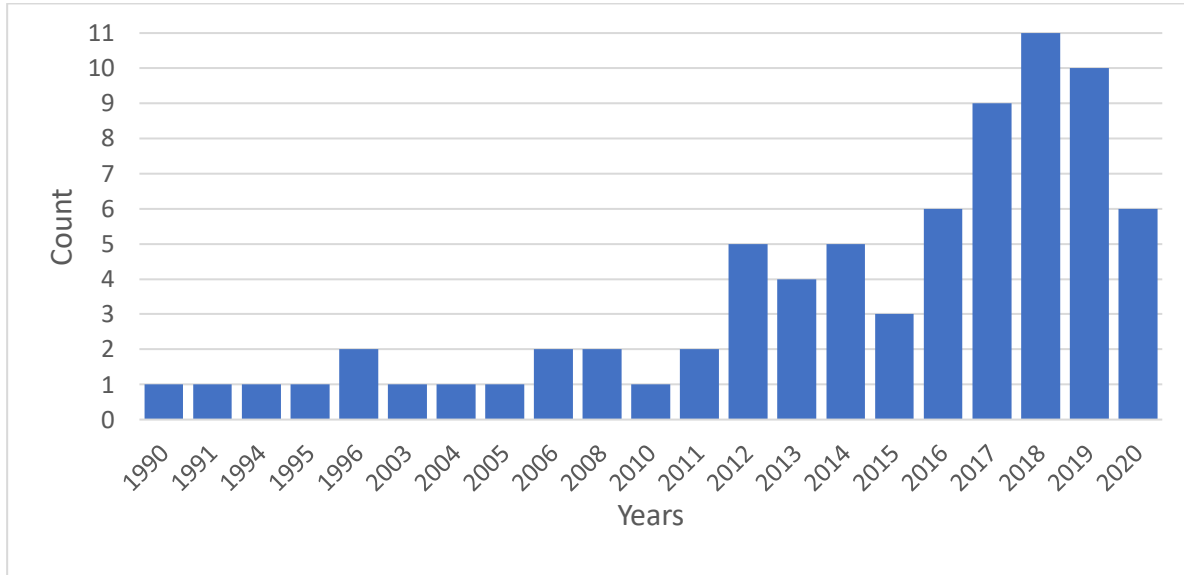


Figure 1. The distribution of articles over time

Corpora disciplines

One article did not report the discipline(s) of the analyzed abstracts. Approximately 71% (n = 53) of the articles analyzed abstracts from a single discipline (see Table 2). The single disciplines studied more than once were applied linguistics (n = 13, 17.3%), linguistics (n = 7, 9.3%), medical (n = 4, 5.3%), business (n = 2, 2.7%), computer and communication systems engineering (n = 2, 2.7%), education (n = 2, 2.7%), and English language teaching (n = 2, 2.7%).

Table 2. Corpus disciplines

Discipline(s)	n	%
Applied Linguistics	13	17.3
Linguistics	7	9.3
Medical	4	5.3
Applied Linguistics - Educational Technology	2	2.7
Business	2	2.7
Computer and Communication Systems Engineering	2	2.7
Education	2	2.7
English Language Teaching	2	2.7
Applied Linguistics - English Literature	1	1.3
English Studies	1	1.3

Environmental science - Applied Linguistics	1	1.3
Languages in Contrast	1	1.3
Language Sciences – Education	1	1.3
Linguistics - Applied Linguistics	1	1.3
Linguistics - Language - Literature - World English’ – Education	1	1.3
Literature	1	1.3
Mathematics - Applied Linguistics	1	1.3
Translation Studies	1	1.3
Library and Information Science	1	1.3
Scient metrics	1	1.3
Accounting	1	1.3
Applied Mathematics	1	1.3
Arts Education - Sports and Exercise - Management	1	1.3
Biology	1	1.3
Chemistry - Biology - Natural Sciences - Geology – Agriculture	1	1.3
Civil Engineering	1	1.3
Composition – Communication	1	1.3
Conservation Biology - Wildlife Behavior	1	1.3
Criminology	1	1.3
Earth - Formal - Life - Physical - Social Sciences	1	1.3
Education - Economy and Management – Humanity	1	1.3
Education – Psychology	1	1.3
Structural - Environmental - Electrical - Chemical - Computer Engineering	1	1.3
Experimental Social Sciences	1	1.3
Experiment Based Dental Research	1	1.3
Food Technology	1	1.3
General Public Health - Teaching English as a Foreign Language - Foreign	1	1.3
Literature (English) - Mathematics - Biology - Psychology - Curriculum		
Studies - Educational Leadership - Educational Planning and Management		
Geoscience	1	1.3
Higher Education	1	1.3
Human Necessity	1	1.3
Humanities - Social Sciences	1	1.3
Law	1	1.3
Management	1	1.3
Marine Engineering	1	1.3
Nanoscience – Nanotechnology	1	1.3
Physics - Medicine - Electronic Engineering - Philosophy - Marketing -	1	1.3
Applied Linguistics		
Protozoology	1	1.3
Sociology - Education - Electronics – Agronomy	1	1.3
N/A	1	1.3

Periods

Thirteen articles (17.3%) did not report data on the time of the abstracts included in the corpus (see Table 3). In the remaining articles, the periods ranged from 1 to 40 years ($M = 6.5$, $SD = 6.6$). The median time of the corpus was five years. The most frequent time periods of the corpora were one ($n = 12$, 16%), two ($n = 7$, 9.3%) and four ($n = 6$, 8%) years, respectively.

Table 3. Periods of analyzed abstracts in the corpus

Years	<i>n</i>	%
1	12	16
2	7	9.3
3	5	6.7
4	6	8
5	2	2.7
6	5	6.7
7	5	6.7
8	3	4
9	4	5.3
10	3	4
11	4	5.3
13	1	1.3
16	2	2.7
21	1	1.3
26	1	1.3
40	1	1.3
N/A	13	17.3

Genres

The research article was the genre its abstracts most frequently subject to analyses (73.3%), followed by conference (12%). In addition, abstracts of theses in degrees (undergraduate, MS, MA, and Ph.D.), patents, and students' term papers were the subject of limited research (see Table 4).

Table 4. Genres of abstracts

Genre	<i>n</i>	%
Research article	55	73.3
Conference	9	12.0
Conference and research article	2	2.7
Ph.D. theses	2	2.7
Undergraduate theses	2	2.7

MA theses	1	1.3
MS-MA theses	1	1.3
Students' term papers and research article	1	1.3
Undergraduates' final term papers	1	1.3
Patent	1	1.3

Languages

English was the most frequently studied language, both individually ($n = 53$, 70.7%) and comparatively ($n = 21$, 28%), followed by Arabic (see Table 5). The languages compared to English were Arabic, Persian, Russian, Spanish, Thai, Turkish, Chinese, French, German, Iraqi, Slavic, and Vietnamese.

Table 5. Languages of analyzed abstracts

Language(s)	<i>n</i>	%
English	53	70.7
English – Arabic	4	5.3
English – Persian	3	4.0
English – Russian	2	2.7
English – Spanish	2	2.7
English – Thai	2	2.7
English – Turkish	2	2.7
English – Chinese	1	1.3
English – French	1	1.3
English – German	1	1.3
English – Iraqi	1	1.3
English – Slavic	1	1.3
English – Vietnamese	1	1.3
Arabic	1	1.3

Sampling and corpus size

In 8% ($n = 6$) of the articles, the corpus consisted of the full population. Except for these articles, population size was reported in 2.9% ($n = 2$) of the remaining articles. In 89% ($n = 67$) of the articles, the population size was not reported. In addition, the sample size was not calculated in 98.5% ($n = 68$) of the articles. Only one article (1.5%) reported the sample size calculation. As seen in Table 6, the sample size of the corpus ranged from 5 to 4214 abstracts ($M = 223.84$, $SD = 523.8$). The median sample size of the corpus was 94 abstracts. In other words, half of the corpus sample sizes were 94 abstracts or less.

Table 6. Corpus sample sizes by the range

Range	<i>n</i>	%
5-28	8	10.7
30-50	8	10.7
51-77	12	16.0
80-100	16	21.3
108-137	8	10.7
160-294	11	14.7
300-625	9	12.0
1180-4214	3	4.0

Authorship and inter-coder agreement

Authorship of the articles was 48% (n = 36) single and 52% (n = 39) multiple. In 56.4% (n = 22) of the multiple-author articles, neither inter-coder agreement nor the number of authors or coders involved in the data collection process was reported. The remaining multiple-author articles reported Cohen's kappa (n = 8, 20.5%), or percent agreement as "checked" (n = 9, 23.1%) for inter-coder agreement (see Table 7). Expert coders were involved in the data collection or control process in 16.7% (n = 6) of single-author articles. In these articles, 5.6% (n = 2) Cohen's kappa and 11.1% (n = 4) "percent agreement" were reported for inter-coder agreement.

Table 7. Inter-coder agreement by authorship

Inter-coder agreement	Multiple authorship		Single authorship	
	<i>n</i>	%	<i>n</i>	%
Cohen's kappa	8	20.5	2	5.6
Percent agreement	9	23.1	4	11.1
N/A	22	56.4	30	83.3

Genre models

No reference was given for the genre model used in six (8%) articles. The remaining 69 articles (92%) used a total of 35 genre models, 24 once and 12 multiple times (see Table 8). A single genre model was used in 75.4% (n = 52) of the articles. This was followed by two (n = 10, 14.5%), three (n = 3, 4.3%), five (n = 3, 4.3%), and six (n = 1, 1.4%) specific genre models. The most frequently used genre models were (Hyland, 2000) (n = 24), (Swales, 1990) (n = 18) and (Santos, 1996) (n = 13). In 75 articles, 14 (18.7%) cited at least one of the abstract standards.

Table 8. Cited genre models

Genre model	<i>n</i>
(Hyland, 2000)	24
(Swales, 1990)	18
(Santos, 1996)	13
(Swales, 2004)	4
(Bhatia, 1993)	3
(Pho, 2008)	3
(Swales and Feak, 2009b)	3
(Swales, 1981)	3
(Hyland, 2004)	2
(Lorés-Sanz, 2004)	2
(Salager-Meyer, 1992)	2
(Yakhontova, 2002)	2
(Aragonés, 2010)	1
(Crookes, 1986)	1
(Dahl, 2004)	1
(Dudley-Evans, 1986)	1
(Endres-Niggemeyer, 1998)	1
(Flower and Ackerman, 1994)	1
(Ge and Yang, 2005)	1
(Halliday and Hasan, 1989)	1
(Hyland, 2008)	1
(Kanoksilapatham, 2013)	1
(Ku, 2019)	1
(Lewin, 2010)	1
(Lim, 2006)	1
(Moreno and Swales, 2018)	1
(Nwogu, 1997)	1
(Pho, 2009)	1
(Salager-Meyer, 1990)	1
(Samar and Talebzadeh, 2006)	1
(Swales and Feak, 2010)	1
(Swales, Irwin and Feak, 2009)	1
(Talebzadeh et al., 2014)	1
(Weissberg and Buker, 1990)	1
(Yang and Allison, 2003)	1
Own model/no citation	6

Comparisons

Comparisons were made in 47 (63%) of the articles. In comparative studies, languages ($n = 21$, 44.7%), disciplines ($n = 12$, 25.5%), genre, and native/non-native writers ($n = 4$, 8.5%) were the most frequently compared subjects (see Table 9). In addition, comparisons were made on accepted-rejected conference abstracts, author –

abstractor, expert-novice writer, geographic, Scopus quartile, and Web of Science - non-Web of Science journal. In 76% (n = 16) of the language comparison articles, cross-language differences in abstracts were concluded as “cultural” without any data.

Table 9. Comparisons of analyzed abstracts

Comparison	n	%
Languages	21	44.7
Disciplines	12	25.5
Genres	4	8.5
Native vs. non-native writers	4	8.5
Accepted vs. rejected conference abstracts	1	2.1
Authors vs. abstractors	1	2.1
Expert vs. novice writers	1	2.1
Geographic	1	2.1
Scopus quartiles	1	2.1
Web of Science vs. Non-Web of Science journals	1	2.1

Move percentages and sequences

The percentages of moves such as introduction (I), purpose (P), methods (M), results (R), and conclusions (C) were reported in 67 articles (89.3%), incomplete in three articles (4%), and not reported in five articles (6.6%). Some of the articles using small-scale corpus preferred to report the frequency of moves rather than percentages. Move sequences (e.g., I-P-M-R-C, P-M-R-C, P-R-I, etc.) were not reported in 43 articles (57.3%), underreported in one article (1.3%), and fully reported in 31 articles (41.3%).

Abstract length and sentence voice

More than half of the articles (n = 42, 56%) did not report data on the length of abstracts. The average word length of the abstracts was reported in 28 (37.3%) articles. In five (6.6%) articles, abstract word lengths were reported in detail by their moves. Descriptive data on the sentence voice (active or passive) were reported for each move in 16 (21.3%) articles. The most frequently used voices for some moves were reported in 11 articles (14.6%).

Discussion

Journals and articles

It is inferred from the results that more than three-quarters of genre analysis studies on scientific abstract structures have been published since 2012. This result shows parallelism with the second period of scientific journal “explosion” in history (Gu and Blackmore, 2016: 714). In addition, the results of present study show that two-thirds of genre studies on scientific abstract structures were published by linguistics and related discipline journals. However, genre analysis is a recommended methodological approach to be in the toolbox of LIS researchers (e.g., Andersen, 2008; Bowker, 2018), it has received limited attention in LIS journals. Journals highlighted in present study results may be beneficial for researchers in selecting suitable journals, as they have the potential to publish future genre analysis studies on scientific abstract structures.

Corpora disciplines

Empirical results of corpora disciplines show that genre analysis studies on scientific abstract structures are conducted for two purposes. The first purpose is to reveal how abstracts in a given domain are segmented into discourse units. Studies conducted for this purpose have pedagogical implications. The majority of these articles analyze abstracts from a single discipline. Montesi and Urdiciain (2005) also reached a similar result. Their literature review of 23 studies concluded that linguistics studies on abstracts are restricted to certain scientific domains. The second purpose is to make comparisons and propose or test genre models on a multi-domain corpus. However, many disciplines not included in Table 2 have not been studied yet. Future studies are expected to fill this gap.

Periods

It is inferred from periods results that articles tend to analyze current abstracts. On the other hand, abstract writing changes over time to meet the needs of the scientific communication communities, as can be seen in revisions of ANSI/NISO Z39.14 (1971, 1997). For this reason, it is important to report data on the time of the abstracts included in the corpus. In addition, studies examining how scientific abstract writing changes over time in certain disciplines can be expected to increase in the future.

Genres

The result of present study show that three-quarters of genre analysis studies focused on “article” abstracts. Considering that we are in the second journal growth boom period in history (Gu and Blackmore, 2016: 714), this is an expected result for the research article genre. However, other genres have been studied limitedly. Moreover,

a genre analysis has not been conducted on scientific report abstracts. Future studies are expected to address this gap.

Languages

The social, technological, and scientific conditions after the World War II made English not only an important language but also the language of science and technology (Kaplan, 2011: 12). As a natural consequence of this, researchers with different native languages have also been required to write in English to contribute to the international scientific literature. In this regard, English may have been the most frequently studied language, both individually and comparatively. On the other hand, this study was limited to articles written in English. For this reason, articles written in other languages were not examined. However, there are limited studies in the English literature on scientific abstracts written in other languages. Future studies are expected to focus on languages besides English.

Sampling and corpus size

In previous studies, it was claimed that small-scale corpora were used both in general genre analysis (Cotos, 2018: 2) and in genre analysis on scientific abstracts (Montesi and Urdiciain, 2005). The empirical evidence presented in this study supports previous studies. On the other hand, considering that move analysis was proposed by Swales (1981) as a pedagogical approach to help authors improve their academic communication skills, the generalizability of the genre analysis studies gains importance in terms of pedagogical implications. In this context, conducting genre analysis on small-scale corpora makes the generalizability of research results a matter of debate (e.g., Montesi and Urdiciain, 2005). Moreover, the fact that population size and sample size calculation were not reported in almost any article can be considered a "methodological deficiency" or "risk of bias" in sampling. To avoid risky implications, in future studies, it will be significant to report bias-free sampling approaches that can ensure generalizability of results, albeit costly, as seen in a limited number of studies.

Authorship and inter-coder agreement

As identifying rhetorical functions can be limited to subjective human judgment and interpretation, there are usually multiple coders in the genre analysis literature to test inter-coder agreements (Biber, et al., 2007: 83; Cotos, 2018; Rau and Shih, 2021). Contrary to this, according to the results of present study, the intercoder agreement was not reported in %70 of the articles. For inter-coder agreement, percentage agreement (n = 14) was reported more frequently than Cohen's kappa (n = 10) in the articles. A recent study (Rau and Shih, 2021) also recommends using percentage agreement to test inter-coder agreement as the data is far from meeting Cohen's kappa requirements for move analysis.

Genre models

The results of present study show that the use of a single genre model was more common in genre analysis articles on abstracts. However, there are also articles in which more than one model is combined to conduct a more detailed analysis. For example, Swales' (1990) CaRS (Creating a Research Space) model, which describes the moves and steps in the introduction part of research articles, has been used as a model or applied to the introduction move of scientific abstracts. Among the most frequently used genre model, Hyland's (2000) genre model (Introduction - Purpose - Method - Product - Conclusion) was developed on a large-scale corpus in terms of both disciplinary diversity and the number of abstracts. Because of its development methodology, Hyland's (2000) genre model may have been most frequently used to analyze abstracts from many disciplines, genres, and languages. Another trend in the genre model was the use of study results on a genre, discipline, or language as a genre model in another study. So, many genre models have been used in articles. The abstract genre models in Table 8 can be used to select suitable models for future studies.

Comparisons

The results of present study show that in genre analysis articles on abstracts, comparisons are often used in many ways to identify the similarities and differences of abstracts. On the other hand, it seems debatable that in a majority of the language comparison articles (71%) cross-language differences in abstracts were concluded as "cultural" without any data. Although this conclusion may seem acceptable at first glance, there may be other factors that need further investigation. For example, cross-language differences in abstracts may be due to the lack of awareness of the authors about the abstract writing standards, as it can be seen from the results of present study that only 18.7% of the articles cited the standards.

Move percentages and sequences

The results highlight that the majority of the articles reported the move percentages. However, some of the articles using small-scale corpus preferred to report the frequency of moves rather than percentages. In addition, the move percentages presented in some articles were not suitable for descriptive statistics presentation. Reporting both frequency and percentages in descriptive statistics presentations can make the results easier to understand. On the other hand, more than half of the articles did not report the move sequences. Reporting the move sequences is important as it reveals how the authors structured the abstracts. According to ISO 214 (1976), the optimum sequence may depend on the target audience, for example, a results-oriented arrangement where the most important findings and results are placed first may be beneficial for some audiences.

Abstract length and voice

The results of present study suggest that genre analysis articles on scientific abstract structures tend to report the word counts of the corpus or the average word length of abstracts as a whole, rather than abstract moves. On the other hand, it is valuable to reveal the lengths of moves in abstracts by genre, discipline, and language, as abstract types and lengths affect information retrieval performance (Liddy, 1991). In addition, a limited number of articles reported the sentence's voices (active or passive) by their moves. Reporting lengths and voices by moves could be beneficial for a better understanding of scientific abstract structures.

Conclusion

In the rapidly growing scientific literature, abstracts play essential roles both in deciding to read the entire document and in developing information retrieval systems and secondary information services. Hence it is important that scientific abstracts are well-structured and adequately provide the necessary information. On the other hand, genre analysis is a prominent methodological approach to revealing the structures of scientific abstracts. Therefore, this study represents a first attempt to conduct a quantitative content analysis of genre analysis articles on scientific abstract structures to reveal trends and possible gaps in the literature.

However, the present study has several limitations. First, the study analyzed articles indexed in Web of Science and Scopus databases. Specific field indexes or databases were not included in the study. Second, a manageable number of articles manually analyzed. Third, books, book chapters, theses, and proceedings were excluded from the scope of the study. Fourth, articles written in English were included the present study.

Despite the limitations listed above, present study has several practical implications. First, the journals reported in the present study results may benefit researchers as they have the potential to publish future research. Second, 35 models revealed by this study can be used in genre model selection for scientific abstract structures in future studies. Third, present results highlight several gaps in the literature, particularly in genres, disciplines, and languages. The LIS community, as well as applied linguistics, can seize the opportunity to address these gaps. However, the results of the present study provide empirical evidence that generalizability was questionable in previous articles, as most of them used small-scale corpus and did not

report inter-coder agreement and unbiased sampling approaches. Therefore, future studies are expected to be generalizable to avoid misleading pedagogical implications. In addition, it would be significant for future studies to report the sequence, length, voice (active/passive), and tense of moves to better understand how author abstracts are structured in practice.

In terms of future research, it would be significant to extend the current results by conducting more content analysis on books, book chapters, proceedings, theses, and articles indexed in more databases and languages. In addition, it would be useful to conduct a content analysis of studies exploring metadiscourse devices used in scientific abstracts by disciplines, genres, and languages.

References

- Ad Hoc Working Group on the Critical Evaluation of Medical Literature (1987) A proposal for more informative abstracts of clinical articles. *Annals of Internal Medicine* 106(4): 598-604.
- Alharbi LM, and Swales JM (2011) Arabic and English abstracts in bilingual language science journals: Same or different? *Languages in Contrast* 11(1): 70–86.
- Alspach JG (2017) Writing for publication 101: Why the abstract is so important. *Critical Care Nurse* 37(4): 12– 15.
- Andersen J (2008) The concept of genre in information studies. *Annual Review of Information Science and Technology* 42(1): 339-367.
- ANSI/NISO Z39.14-1971 (1971) Guidelines for abstracts.
- ANSI/NISO Z39.14-1997 (1997) Guidelines for abstracts.
- ANSI/NISO Z39.14-2015 (2015) Guidelines for abstracts.
- Aragonés M (2010) How to become a patent translator: Tricks and tips – notions of text genre and ceremony to the rescue. *Meta* 55(2): 212-236.
- Bhatia VK (1993) *Analysing genre: Language use in professional settings*. London: Longman.
- Biber D, Connor U and Upton TA (2007) *Discourse on the move: Using corpus analysis to describe discourse structure*. John Benjamins Publishing.
- Borko H and Chatman S (1963) Criteria for acceptable abstracts: A survey of abstracters' instructions. *American Documentation* 14(2): 149-160.

- Bowker L (2018) Corpus linguistics is not just for linguists: Considering the potential of computer-based corpus methods for library and information science research. *Library Hi Tech* 36(2): 358-371.
- Coe K and Scacco JM (2017) Content analysis, quantitative. In: Matthes J, Davis CS and Potter RF (Eds) *The international encyclopedia of communication research methods*. John Wiley & Sons, pp.1-11.
- Cotos E (2018) Move analysis. In: Chapelle CA (Ed) *The encyclopedia of applied linguistics*. John Wiley & Sons, pp.1-8.
- Crookes G (1986) Towards a validated analysis of scientific text structure. *Applied Linguistics* 7(1): 57-70.
- Dahl T (2004) Some characteristics of argumentative abstracts. *Akademisk Prosa* 2: 49-67.
- Dudley-Evans A (1986) Genre analysis: an investigation of the introductions and discourse sections of MSc dissertations. In: Coulthard M (Ed) *Talking about text*. Birmingham: English Language Research, Birmingham University, pp.128-145.
- Dudley-Evans T (2000) Genre analysis: A key to a theory of ESP? *Ibérica, Revista de la Asociación Europea de Lenguas para Fines Específicos* 2: 3-11.
- Endres-Niggemeyer B (1998) *Summarizing information: Including CD-Rom "SimSum", simulation of summarizing, for Macintosh and Windows*. Berlin: Springer.
- Flower L and Ackerman JM (1994) *Writers at work: Strategies for communicating in business & professional settings*. Fort Worth: Harcourt Brace College Publishers.
- Garfield E (1979) *Citation indexing, its theory and application in science, technology, and humanities*. New York: Wiley.
- Ge DM and Yang RY (2005) Genre analysis of academic articles. *Modern Foreign Languages (Quarterly)* 2: 138-146.
- Gu X and Blackmore KL (2016) Recent trends in academic journal growth. *Scientometrics* 108(2): 693-716.
- Halliday MAK and Hasan R (1989) *Language, context, and text: Aspects of language in a social-semiotic perspective* (2nd ed). Oxford: Oxford University Press.
- Haynes RB, Mulrow CD, Huth EJ, Altman DG and Gardner MJ (1996). More informative abstracts revisited. *The Cleft Palate-Craniofacial Journal* 33(1): 1-9.
- Hyland K (2000) *Disciplinary discourses: Social interactions in academic writing*. London, UK: Longman.
- Hyland K (2004). *Disciplinary discourses: Social interactions in academic writing*. University of Michigan Press.

- Hyland K (2008) *English for academic purposes: An advanced resource book*. London and New York: Routledge, Taylor and Francis Group.
- ISO 214:1976 (1976) Documentation - Abstracts for publications and documentation.
- Kanoksilapatham B (2005) Rhetorical structure of biochemistry research articles. *English for Specific Purposes* 24(3): 269-292.
- Kanoksilapatham B (2013) Generic characterization of civil engineering research article abstracts. *JL: The Southeast Asian Journal of English Language Studies* 19(3): 1-10.
- Kaplan RB (2011) English - the accidental language of science. In: Ammon U (Ed) *The dominance of English as a language of science: Effects on other languages and language communities*. Berlin, New York: De Gruyter, pp.3-26.
- Khansari D, Chan SH, Chan MY and Tan H (2016) Regularities and Irregularities in Rhetorical Move Structure of Linguistics Abstracts in Research Articles. *JL, Language, Linguistics, Literature* 22(1): 39-54.
- Ku MC (2019) A comparative analysis of English abstracts and summaries of Chinese research articles in three library and information science journals indexed by the Taiwan Social Science Citation Index. *Journal of Library & Information Studies* 17(1): 37-81.
- Lewin BA (2010) *Writing readable research*. London: Equinox.
- Liddy ED (1991) The discourse-level structure of empirical abstracts: An exploratory study. *Information Processing & Management* 27(1): 55-81.
- Lim JMH (2006) Method sections of management research articles: A pedagogically motivated qualitative study. *English for Specific Purposes* 25(3): 282-309.
- Lor'es-Sanz R (2004) On RA abstracts: From rhetorical structure to thematic organization. *English for Specific Purposes* 23(3): 280-302.
- Mabe M and Amin M (2001) Growth dynamics of scholarly and scientific journals. *Scientometrics* 51(1): 147-162.
- Montesi M and Urdiciain BG (2005) Recent linguistic research into author abstracts: Its value for information science. *Knowledge Organization* 32(2): 64-78.
- Moreno AI and Swales JM (2018) Strengthening move analysis methodology towards bridging the function-form gap. *English for Specific Purposes* 50: 40-63.
- Mulrow CD, Thacker SB and Pugh JA (1988) A proposal for more informative abstracts of review articles. *Annals of internal medicine* 108(4): 613-615.
- Nwogu KN (1997) The medical research paper: Structure and functions. *English for Specific Purposes* 16(2): 119-138.

- Pho PD (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.
- Pho PD (2009) An evaluation of three different approaches to the analysis of research article abstracts. *Monash University Linguistics Papers* 6(2): 11-26.
- Rau G and Shih YS (2021) Evaluation of Cohen's kappa and other measures of inter-rater agreement for genre analysis and other nominal data. *Journal of English for Academic Purposes* 53: 101026.
- Salager-Meyer F (1990) Discoursal flaws in medical English abstracts: A genre analysis per research- and text-type. *Text* 10(4): 365-384.
- Salager-Meyer F (1992) A text-type and move analysis study of verb tense and modality distribution in medical English abstracts. *English for Specific Purposes* 11(2): 93-113.
- Samar RG and Talebzadeh H (2006) Professionals write like this: The case of ESP/EAP experimental research article abstracts. In: Post-graduate conference, University of Tehran, Iran.
- Samar RG, Talebzadeh H, Kiany G and Akbari R (2014) Moves and steps to sell a paper: A cross-cultural genre analysis of applied linguistics conference abstracts. *Text & Talk* 34(6): 759-785.
- Santos MB (1996) The textual organization of research paper abstracts in applied linguistics. *Text-Interdisciplinary Journal for the Study of Discourse* 16(4): 481-500.
- Swales JM (1981) *Aspects of article introductions*. Birmingham, England: University of Aston.
- Swales JM (1990) *Genre analysis: English in academic and research settings*. Cambridge: Cambridge University Press.
- Swales JM (2004) *Research genres: Explorations and applications*. Cambridge: Cambridge University Press.
- Swales JM and Feak CB (2009a). *Abstracts and the writing of abstracts*. University of Michigan Press.
- Swales JM and Feak CB (2009b) *Academic writing for graduate students: Essential tasks and skills* (2nd ed.). Michigan: University of Michigan Press.
- Swales JM and Feak CB (2010) From text to task: Putting research on abstracts to work. In: Ruiz-Garrido MF, Palmer-Silveira JC and Fortanet-Gómez I (Eds) *English for professional and academic purposes*. Leiden, Netherlands: Brill, pp.167-180.
- Swales JM, Irwin F and Feak CB (2009) *Online commentary for abstracts and the writing of abstracts*. Michigan: University of Michigan Press.

- Talebzadeh H, Samar RG, Kiany GR and Akbari R (2014) Steps to a successful abstract: A comparative genre analysis. *The International Journal of Humanities* 20(3): 1–25.
- Weil BH (1970) Standards for writing abstracts. *Journal of the American Society for Information Science* 21(5): 351-357.
- Weissberg R and Buker S (1990) *Writing up research: Experimental research report writing for students of English*. Englewood Cliffs, NJ: Prentice Hall Regents.
- Yakhontova T (2002) 'Selling' or 'telling'? The issue of cultural variation in research genres. In: Flowerdew J (Ed) *Academic discourse*. Harlow: Longman, pp.216–232.
- Yang R and Allison D (2003) Research articles in applied linguistics: Moving from results to conclusions. *English for Specific Purposes* 22(4): 365–285.