SEBASTIAN JAN DUNAT

ACADEMIC LANGUAGE FUNCTIONS IN CONFERENCE ABSTRACTS

CORPUS ANALYSIS OF ACADEMIC LANGUAGE FUNCTIONS

PH.D. THESIS

Supervisor:
prof. dr hab. Andrzej Łyda

Sosnowiec, 2025

Słowa kluczowe: Academic Language, Academic Vocabulary, Conference Abstracts, Corpus analysis

(max 5)

Oświadczenie autora pracy

Ja, niżej podpisany: Sebastian Jan Dunat,

autor pracy dyplomowej pt. Academic Language Functions in Conference Abstracts: Corpus Analysis of Academic Language Functions

Numer albumu: 7701
Student,
(Nazwa wydziału/jednostki dydaktycznej)

kierunku studiów: Językoznawstwo,

specjalności*

*(wypełnić, jeśli dotyczy)

Oświadczam, że ww. praca dyplomowa:

- została przygotowana przeze mnie samodzielnie[1],
- nie narusza praw autorskich w rozumieniu ustawy z dnia 4 lutego 1994 r. o prawie autorskim i prawach pokrewnych (tekst jednolity Dz. U. z 2006 r. Nr 90, poz. 631, z późn. zm.) oraz dóbr osobistych chronionych prawem cywilnym,
- nie zawiera danych i informacji, które uzyskałem/łam w sposób niedozwolony,
- nie była podstawą nadania dyplomu uczelni wyższej lub tytułu zawodowego ani mnie, ani innej osobie.

Oświadczam również, że treść pracy dyplomowej zamieszczonej przeze mnie w Archiwum Prac Dyplomowych jest identyczna z treścią zawartą w wydrukowanej wersji pracy.

Jestem świadoma/y odpowiedzialności karnej za złożenie fałszywego oświadczenia.

…… …….
Data Podpis autora pracy

[1] uwzględniając merytoryczny wkład promotora (w ramach prowadzonego seminarium dyplomowego)

TABLE OF CONTENTS

LIST OF ABBREVIATIONS

AL - Applied Linguistics
ALLA - Applied Linguistics Association of Australia
ANC - American National Corpus
AWL - Academic Word Lists
BH - Biology and Health
BNC - British National Corpus
BOE - Bank of English
CARS - Create a Research Space
CHILDES - Child Language Data Exchange System
CS - Computational Statistics
EAL - English as Additional Language
EFL - English as a Foreign Language
ELF - English as Lingua Franca
ESP - English for Specific Purposes
F - Female
F1 - Formulating
F10 - Reinforcing research process
F2 - Investigating
F3 - Analyzing
F3.1 - Presenting methodology
F3.2 - Design the experiment
F3.3 - Simulate the situation/develop a model
F4 - Reporting results
F5 - Drawing conclusions
F6 - Reviewing
F7 - Evaluating
F8 - Criticizing
F9 - Organizing order of information
G – Multi-authored
GSL - General Service List
H - Humanities
IPMPrC - Introduction, Purpose, Method, Product, Conclusion
ISO - International Organization for Standardization
JOP - Journal of Pragmatics
JRR - Research in Reading
JSLW - Journal of Second Language Writing
KWIC - Key Word in Context
L - Linguistics
M - Male
MD - Multi-dimensional
MS - Microsoft
N - Native
NF - Native Female
NLP - Natural Language Processing
NM - Native Male
NN - Non-Native
NNF - Non-Native Female
NNM - Non-Native Male
NNS - Non-Native Speaker
NS - Native Speaker
P-M-Pr-C - Purpose, Method, Product, Conclusion
R&W - Reading and Writing
RA - Research Article
VOA - Vocabulary of Analysis
VOE - Vocabulary of Evaluation
VORP - Vocabulary of Research Process

LIST OF TABLES

LIST OF FIGURES

Introduction

As Hyland (2014) claims, the first encounter of the interested reader with an academic text during the research process is an abstract. Several aspects of abstracts have been pointed to in the literature. First, the importance of abstract summarizing functions was underlined by such scholars as Graetz (1985), Salager-Mayer (1990), Bhatia (1993), Kaplan et al. (1994), Ventola (1994a, 1997) and Hyland (2000). Secondly, the structure of abstracts is described in the works of Bhatia (1993), Dahl (2000), Stotesbury (2003), and even defined by the American Standards Institute (ANSI) in 1979. Last but not least, the rhetorical functions used in abstracts are considered in the works of, e.g., Swales (1990), Santos (1996), Hutz (1997), Hyland (2000) and Lorés (2004). All of the rhetorical functions presented in these studies enhance the abstract's task, i.e., they encourage further examination of the article itself.

On the other hand, the significance of evaluating discourses, or considering how a few different genres are related, as well as searching for a variation between registers, are also crucial in the area of academic text analysis; such issues are addressed by, e.g., Stotesbury (2003) and Samraj (2005), Biber (2006), Ehlers-Zavala (2008); Scott, Nagy, and Flinspach (2008); Pilgreen (2007), Watinee and Siriluck (2013), Pho (2008), Samraj (2002a). Among other aspects, one cannot deny the importance of corpus linguistics, which is discussed by, e.g., Biber (2006), Teubert (2004), Orasan (2001), and Busch-Lauer (2014). Hasko (2012) emphasizes that the complexity of analysis requires proper compilation of the corpus, Biber (2006) presents quantitative measures for vocabulary analysis, Sinclair (1992) studies the purposes of corpus analysis, O'Keffe and McCarthy (2010) present views on computational corpus analysis, Flowerdew (2012) discusses historical aspects of corpus analysis, and McEnery and Wilson (2001) underline the importance of English as lingua franca in corpus linguistic works. Further, Teubert and Čermáková (2004) define types of corpora in use, Grishman (1986), Hausser (2014), Sinclair (1992) present computational linguistic approaches of analysis and Barlow (1996) and Scott (1996) present computer tools for linguistic analysis.

Moreover, one cannot forget about a variety of sources analyzing linguistic and statistical aspects in various settings, e.g., Birkeland, Belić, Firth, De Groot, Hocket, Reichling, Voegelin (1956) on structural linguistics; de Saussure (1983), Furdal (1977), Milewski (2004) on general linguistics; Steponavičius (2006) on diachronic linguistics; Evans (2007) on cognitive linguistics; Heinz (1983) on historical aspects of linguistics; Nowak (1985) on sociolinguistics methodology; MacPherson (2004) on English for Academic

Purposes; Łyda (2007) on Academic Spoken Discourse; Walesiak and Gatnar (2009), Stąpor (2015), or Wickham and Grolemund (2017) on a statistical analysis of data.

Apart from the structure and functions of abstracts, it is necessary to consider, in the present study, academic vocabulary discussed by, e.g., Martin (1976) and Baumann and Graves (2010). While corpus vocabulary analysis of academic registers has been carried out by, e.g., Nation (2001), Coxhead (2000), and Fløttum et al. (2006, 2007), one cannot neglect the importance of academic language presentation and vocabulary typologies found in: e.g., Coxhead (2000), Beck, McKeown and Kucan (2002, 2008), Jetton and Alexander (2004), Pilgreen (2007), Fisher and Frey (2008), Harmon, Wood and Hendrick (2008), Hiebert and Lubliner (2008), Townsend (2008) or Baumann and Graves (2010). Furthermore, broad academic literacy issues, which influence the academic discourse and abstracts as its part, are presented in Gutiérrez (2008), Janzen (2008), Lewis and Reader's (2009); reading and writing use, as well as reading proficiency and meaning construction is discussed in, e.g., Torgesen's et al. (2007), Moore's (2008).

The issue of abstracts as an academic genre, in various scientific dimensions, has been addressed by a number of scholars: diachronic rhetorical features in abstracts by Gillaerts (2014), diachronic comparison of authorial voice and self-mention in fields of economics, history, and linguistics in Bondi (2014), evaluative *that* in abstracts in Hyland and Tse (2005), prominence of abstract as a scientific genre and its functions in Busch-Lauer (2014), analysis and categorization of evidentials and epistemic markers, the genre and register connotations in abstracts in Alonso-Almeida (2014), abstracts rhetorical distinctions in Ortega-Barrera & Torres-Ramirez (2010), length of abstracts in Tibbo (1992), Melander et al. (1997), and Orasan (2001), differences between English, German and Russian abstracts in sociological journals in Baβler (2003), comparison of elements of self-advertisement in English and Russian abstracts in Müller (2008), and ELF abstract writing in Swales and Feak (1994).

There are various ways to present, define and analyze the structure and rhetorical functions of abstracts, as the above-mentioned academic works illustrate. Although previous studies have addressed many purposes of research on abstracts, not only in terms of corpus linguistics, none have focused on academic language functions in abstracts, especially using computational linguistics methods for research. The present dissertation is organized in the following order: the material is assessed and described with the use of a special corpus and computational linguistic methods to compare the academic vocabulary used in conference abstracts by both sexes in all of the characterized academic fields. The methodology of Averil Coxhead (2000) is used to establish the most frequent vocabulary data in question. The

analysis aims at detecting possible interdisciplinary variation in the frequency of occurrence of academic language functions. Gender is established as the leading parameter of the study. The presented corpus data, used for lexical analysis, is labeled, firstly into parts of speech, and secondly into three categories: Vocabulary of the research process, Vocabulary of analysis, and Vocabulary of evaluation, based on Martin (1976). The corpus data is chosen from five books of abstracts from five conferences, each on a different continent. It is selected by hand to equalize the number of examples, according to the gender factor.

The corpus data is divided into three academic fields: Computational Statistics, Humanities, and Biological & Health Sciences. The corpus has been designed in the present study to determine the quantitative and, to some extent, qualitative differences in the use of academic language functions in various fields of written academic language. The data of the corpus provide the following elements: Source, Reference number and page number, Author, Gender, Speaker type, Institution, Speaker origin, Discipline, Title, Abstract text, Abstract tagged for the parts of speech, and Number of nouns, verbs, adjectives, and adverbs used in the corpus texts. It is expected that there will be differences in the realization of language functions and academic vocabulary in the groups under analysis.

The thesis is divided into five chapters. The first presents a theoretical background of communication in academic discourse, the genre of abstracts, corpus, and principles of computational linguistics, gender aspect in linguistic research, data evaluation in statistics, and academic vocabulary and functions of academic language. This discussion, comprised in five sections of the first chapter, forms the basis of the introduction of important notions used for the implementation of this dissertation. The second chapter consists of descriptions of practical approaches to abstract analysis, not only in corpus studies. The discussion focuses on seven studies reviewing presented journal articles, which underline cross-linguistic, disciplinary, intercultural, and diachronic perspectives as well as rhetorical structure, and corpus study elements. The third chapter presents the analysis design and introduces computational means of the analysis, i.e., software programs used in the dissertation (R, AntConc, ProtAnt, TagAnt). The central part of the thesis, the fourth chapter, provides an analysis of the corpus data, researched in terms of gender and speakers' origin to visualize better the differences and similarities occurring in the researched material. The quantitative computational approach is the primary method of analyzing the chosen data samples, i.e., the interpretation of the presented tables, considering the female and male usages of specified lexis, and the quantitative use of specified academic language functions, is verified by Pearson's Chi-square test conducted in the R program, and the Ant software tools. The last

part of the dissertation contains conclusions corresponding to the data interpretation presented in each section of the fourth chapter and suggestions for further investigation. It offers a survey of the similarities and differences of the research data for gender and writers' native/non-native factors. The outcomes of the study are discussed in reference to selected sources. Conclusions are backed up by Pearson's Chi-Squared Test (where appropriate), extracted from the R software, visualizing the numerical differences between the specified data.

Chapter One
Theoretical background

1.1 Communication within academic discourse

The continuing development of any kind of research worldwide is conducted predominantly in English, the language that serves as a medium for sharing findings on an international scale (Swales & Feak, 2009). This trend, reinforced by the policies of higher education ministries across many countries, has led to the situation where scholars are determined to publish their texts in high-impact factor literature to receive primary credit for this kind of work. Since 2017, similar trends have also been observed in Poland. The already known labels of English as lingua franca (ELF) and English as an additional language (EAL) might replace native/non-native distinctions, thus changing the view on the scholars' nativity factor in research labeling. As stressed by Swales and Feak (2009, introduction), "today, the more valid and valuable distinctions are between senior researchers and junior researchers, on the one hand, and between those who have a broad proficiency in research English across the four skills of reading, writing, listening, and speaking and those with a narrow proficiency largely restricted to the written mode, on the other" (p .11).

The academic community, sharing views on an international scale, bases its communication on the skills of a particular individual, acquired during the long years of work, on the one hand, and the well-known, but not always thought of, basis of rhetorical constructs, e.g., academic discourse and genres. Bucholtz (2003) proposes a twofold definition of discourse, where, from a formal point of view, it is "the linguistic level in which sentences are combined into larger units" (p. 44). Secondly, the alternative view describes discourse as a "language in context: that is, language as it is put to use in social situations, not the more idealized and abstracted linguistics forms that are the central concern of much linguistic theory" (p. 44). For academic discourse the latter definition is more applicable. The reason for this type of thinking was perfectly clarified by Blanton (1994), who argues that "if, however, we treat academic discourse itself as static form rather than the fluid expression of individuals their thoughts, experiences, and discoveries we guarantee that many of our students will not stay long in our institutions, nor even feel included enough to want to stay; we preclude their participating in the process of academic discourse in the making" (p. 4).

For Giannoni (2010), academic discourse "is evolving not only in response to changes in scientific communication but also in the student population" (p. 27). The awareness of linguistic disadvantages felt by international students is widespread. One of the solutions to

end the dilemmas is to teach how to understand the relations built into languages, the power relations, or underline helpful multilingualism implications (Giannoni, 2010). Thus, a deep understanding of the scientific community discourse is needed to accomplish that goal. Therefore, considering a position where a text is "a site of engagement with peers rather than a mere product of linguistic craftsmanship" (Giannoni, 2010, p. 29) is an essential assumption to end the dilemmas. The second most crucial aspect is understanding the disciplinary cultures, which, according to Flottum (2006, as in Giannoni, 2010), "may in fact be more important than languages when accounting for textual variation" (p. 37). This supposition makes disciplinary cultures influence languages, work across and within them, and might change proclivities based on nationality and language, for example, in communication between scholars.

Academic communication is directly linked with academic texts and spoken genres, which serve as a medium to transfer information within the academic community. In general, genre links the individual dimension of language with its social dimensions (Giannoni, 2010). Also Swales and Feak (2009) define genre as "a type of text or discourse designed to achieve a set of communicative purposes" (p. 1). On the other hand, Biber (1988) argues that "genres characterize texts on the basis of external criteria, while text types represent groupings of texts that are similar in their linguistic form, irrespective of genre" (p. 170). Moreover, Biber (1988) adds that "genre categories are determined on the basis of external criteria relating to the speaker's purpose and topic; they are assigned on the basis of use rather than on the basis of form" (p. 170). However, what should be underlined here is the genre use in the academic community. Thus, Bhatia (1997) suggests that "genres are essentially defined in terms of the use of language in conventionalized communicative settings. They are meant to serve the goals of specific discourse communities" (p. 181). Understanding the interrelation between language and its social dimensions is of great value; these relations make genres "an expression of shared cultural norms" (Giannoni, 2010, p. 38). There is a crucial point in perceiving what genre types are used in the academic setting, for example, the ones presented in Swales' taxonomy, which clarifies the ways of communication within the academic community, as presented below in Table 1.

Table 1 Types of academic genres

Primary genres		Secondary genres		Occluded genres	
Written	**Spoken**	**Written**	**Spoken**	**Written**	**Spoken**
Research article	Conference presentation	Textbook	Lecture	External evaluation	Discussion between examiners
Journal abstract	Plenary lecture	Introductory text	Tutorial	Evaluation letter for tenure and promotion	Initiating or responsive phone call
Monograph	Thesis defense	Post-introductory text	Seminar	Book or grant proposal review	
Conference abstract		Course description		Review of articles submitted to refereed journals	
Dissertation		Lecture notes		Research grant proposal	
Thesis		Handout		Initiating or responsive e-mail	
Chapter in an edited book				Application, invitation, request, submission, and editorial-response letters	
Case report					
Book review					
Review article					
Editorial					

Note. Adapted from Giannoni (2010, p. 39)

Many of the genres are treated separately although they are interconnected, because "abstracts are always abstracts of some larger text. A conference talk may be based on a dissertation chapter and may end up as an article. Grant proposals lead to technical reports, to dissertations, and to further grant proposals" (Swales & Feak, 2009, introduction, p. 11). It is obvious that some genres influence other genres, or develop a link between them. Swales and Feak (2009) presented the following academic genre network scheme to signify the importance of these inter-connected networks, as shown in Table 2.

Table 2 Academic Genre Network (Adapted from Swales & Feak, 2009, Introduction, p. 10)

Open genres			
Book chapters	Literature reviews	Conference and other talks	Research articles
Conference posters	Theses and dissertations	Technical reports	Books and monographs
↑	↑	↑	↑
Job interviews	Curricula vitae	Job applications, Fellowship applications	Grant proposals
Practice talks (also known as "dry-runs")	Research paper reviews and responses to reviewers		Submission letters
Supporting Genres			

1.2 The genre of abstract

Abstracts are an essential constituent part of research articles. Hyland and Jiang (2017) present the following description of research articles:

> "research articles are the sites where academics negotiate and make sense of the issues which preoccupy them as members of particular disciplines. These sites, moreover, are not merely storehouses of arcane, abstract practices, monolithic and forever frozen in time, but responsive to changing contexts and the demands of new conditions" (p. 49).

They underline that some academic fields change in rhetorical options used by the researchers, i.e., options to establish a more personal connection with readers.

The importance of abstracts as the second essential part of a paper after the title is noted by Hyland (2014):

> "the abstract is generally the readers' first encounter with a text, and is often the point at which they decide whether to continue reading and give the accompanying article further attention, or to ignore it. (…) Abstracts are worthy of study because they are significant carriers of a discipline's epistemological and social assumptions, and therefore a rich source of interactional features that allow us to see how individuals work to position themselves within their communities" (p. 63).

According to Hyland, persuasion used in abstracts needs specific words, i.e., academic vocabulary. The organizational structure has to be replicated accordingly to the field the abstracts represent. It has to be clear for the readers of specific discourse and within its boundaries as well as capture beliefs and authorized practices (Hyland, 2014).

There are many alternative definitions of abstracts, each presenting a different perspective. One of the definitions which is worth mentioning maintains that an abstract "is basically a concise summary of a much longer report" (Lorés, 2004, p. 281). This definition does not seem to cover the whole spectrum of this genre. The following definition seems to capture the complexity: abstract "is an abbreviated, accurate representation of the contents of a document, preferably prepared by its author(s) for publication with it" (ANSI, 1979, p. 1). At this place, as a complement to the definitions presented above, I would like to propose the following definition: abstracts serve the purpose of presenting elements of the whole text as well as shortly underlining the purpose, defining method, and presenting the outcome of the given study to hook the readers' interest into further reading.

As concerns functions of abstracts, Hyland (2014) argues that many scholars emphasize the summarizing function of abstracts as the most frequent (see, e.g., Bhatia, 1993; Ventola, 1997; Graetz, 1985; Kaplan et al., 1994; Salager-Mayer, 1990). The summarizing function is also mentioned by Donesch-Jeżo (2013, p. 122), who argues that the function of abstracts is to briefly summarize the topic, methodology, and the main findings. Abstracts also help readers to make a decision whether to read an article or not; they constitute a structured overview of the article for those who want to read it (Donesch-Jeżo, 2013).

Other scholars present abstracts as independent discourse structures with *representation* functions (Bazerman, 1984, as in Hyland, 2014, p. 64) or *distillation* functions (Swales, 1990, as in Hyland, 2014, p. 64). As Hyland (2014) argues, "the research article is, in essence, a codification of disciplinary knowledge, where writers seek to persuade their communities to accept their claims and certify them as recognized and legitimate knowledge" (p. 64). Further, he emphasizes that abstracts serve a more urgent and modest purpose "to persuade readers that the article is worth reading" (p. 64). That is why Hyland (2014) describes abstracts as a form of *selective representation* more than the exact *knowledge on an article's content*. This selectiveness influences abstracts' function to encourage further examination of the article itself.

As has already been mentioned, abstracts serve as sources to hook readers; on the other hand, they are tools to select contributions in journals, organize the conference schedule or select the topics for research (Lorés, 2004). Abstracts have also "become a tool of

mastering and managing the ever-increasing information flow in the scientific community" (Ventola, 1994b, after Lorés, 2004, p. 281). To pinpoint the functionality of abstracts within the academic community, it is vital to specify other scholars' views on functions of abstracts. According to Ventola (1994b), one of the functions of abstracts is that English-language abstracts are a guarantee of scientific information circulation worldwide. Furthermore, Lorés (2004), and other scholars (i.e., Day, 1988; Graetz, 1985; Jordan, 1991; Ventola, 1994b) argue that two major types of functionality of abstracts can be distinguished: (1) indicative, and (2) informative.

> "the function of indicative abstracts is to help readers understand the general nature and scope of the research article; it indicates the subject and the main findings of the paper, but it does not go into a detailed step-by-step account of the process involved. On the other hand, informative abstracts encapsulate the whole paper" (Lorés, 2004, pp. 281-282).

The first function bears resemblance to representation functions, and the other, i.e., *informative*, to a summarizing function mentioned above.

It could be argued that abstracts are multifunctional texts. One of the reasons that various linguists show interest in research of this genre is to understand "the mechanisms which underlie these multifunctional texts" (Lorés, 2004, p. 281). Thus, it is also essential to look at the abstracts' structural features. Among various scholars' attempts to identify the features of abstracts, as Hyland (2014, p. 65) argues, the work of Graetz (1985) is worth mentioning; Graetz's (1985) attempts might also be labeled as the distinction of abstracts as a genre:

> "the abstract is characterized by the use of past tense, third person, passive, and the non-use of negatives. It avoids subordinate clauses, uses phrases instead of clauses, words instead of phrases. It avoids abbreviation, jargon, symbols and other language shortcuts which might lead to confusion. (...) In short it eliminates the redundancy which the skilled reader counts on finding in written language and which usually facilitates comprehension" (p. 125).

Among other findings, it is important to note Kaplan et al. (1994, after Hyland, 2014, p. 66), who found that the use of subordinate clauses, abbreviations, and variations of tense and voice is common. Salager-Mayer (1992, as in Hyland, 2014) found that the present tense use signals emphasis on generalisability of specific findings, and noted the frequent use of past tense in medical abstracts. Finally, Rounds (1982, as in Hyland, 2014) found that the extensive use of mitigation and hedging had its source in writers' rhetorical thought to avoid audience rejection by toning their claims down.

It can be assumed that all authors want to present themselves as competent community members; they do it by thorough and careful underlining the central claims in their abstracts

(Hyland, 2014). They also do it "to gain readers' attention and persuade them to read on, writers need to demonstrate that they not only have something new and worthwhile to say, but that they also have the professional credibility to address their topic as an insider" (Hyland, 2014, p. 63). To conclude, the use of summarizing, representation, and persuasive elements, i.e., the selection of the functions used in the text, as well as the structural features, is generally believed to be crucial to writing a good abstract.

It is important to obey the rules of specific academic genres, including the choices in grammar, lexis, discourse organization, and evaluation in academic texts (Stotesbury, 2003). Undoubtedly, functions and features used by the writers in their abstracts help the reader to evaluate the information presented in the text. Stotesbury (2003) states that the presence of evaluation in discourse has been researched by Russian and Finnish scholars, i.e., Volosinov (1986) and Heikkinen et al. (1999). It could be claimed that evaluation is "anything which indicates the writer's attitude to the value of an entity in the text" (Hunston, 1993, after Stotesbury, 2003, p. 328).

Stotesbury (2003) distinguishes four categories of evaluation in research reports, based on Hunston's (1993) research on evaluative-that clauses. The first category, status, "reflects the writer's degree of certainty and commitment towards the proposition and is always present and constantly evaluated in research writing" (Stotesbury, 2003, p. 328). Issues similar to the material mentioned above can also be found in Hyland (2005), particularly, hedging constructions in text. For Stotesbury (2003) the evaluation of status might refer to the different commitment of the author, e.g., in terms of probable, unlikely, certain or uncertain. The second category, value, "denotes quality on the good-bad scale (...) [it] takes place through lexis expressing accuracy, consistency, verity, simplicity, usefulness, reliability or importance, which renders the other language items traditionally regarded as evaluative redundant." (Stotesbury, 2003, p. 328). The third is relevance, "which refers to evaluation of the degree of significance or relevance of the argument in a research article" (p.329). These categories of evaluation may be useful when focusing on the analysis of abstracts. The first three categories are implicit evaluation. The fourth category, the explicit type of evaluation, as Stotesbury (2003) argues, is the attitudinal language, similar to Hunston's (1993) terms *value* and *modality*. To conclude, the genre of abstracts is an excellent example of an evaluative category of texts. Thus, the interests of scholars should be turned more into the genre of abstracts in terms of their functionality.

In order to stress the importance of abstracts as a genre, the following part will present the theory of genre analysis. Genre analysis may be defined as "the study of situated linguistic

behavior in institutionalized academic or professional settings" (Bhatia, 1997a, pp. 134-135). Bhatia (1997a) argues that there are four characteristics of genre analysis:

1. "Rather than providing a detailed extension, validation, or otherwise of one linguistic framework or the other, genre analysis shows a genuine interest in the use of language to achieve communicative goals. In this sense, it is not an extension of linguistic formalism.
2. However, genre analysis does not represent a static description of language use but gives a dynamic explanation of the way expert users of language manipulate generic conventions to achieve a variety of complex goals. In this sense, it combines the advantages of a sociolinguistics perspective, especially the use of ethnographic information, with those of a cognitive perspective, especially regarding the tactical use of language.
3. It is primarily motivated by applied linguistic concerns, especially language teaching at various levels.
4. It is narrow in focus but wide in vision, focusing on specific differentiation in language use at various levels of generality" (pp. 134-135).

Samraj (2005) may serve as an example of genre analysis research. The work focuses on determining whether abstracts are different from introductions and explores how the two genres interact in this setting. As Samraj (2005) emphasizes, the macro-organization of the genre, in terms of English for Specific Purposes (ESP), was studied by many scholars alongside the lexico-grammatical features of constituent moves. Some scholars found that genres vary across linguistic and cultural communities (like Connor, 1996) or disciplines and linked particular discourse features to disciplinary communities (i.e., Melander, Swales & Fredrickson 1997, Samraj 2002b). However, Samraj (2005) argues that there is almost no research on the structure related to genres within academic communities. The works enlisted by Samraj, i.e., the comparison of academic research articles made by Fahnestock 1986, Smith 1990, and Myers 1990, though centered around one area, are the only examples of the articles comparing relations within academic discourse. Further, Samraj (2005) underlines that it is vital to consider and investigate the relations between genres in different disciplines to understand the writing across disciplines better, as in, e.g., Miller's (1994) work, which refers to disciplinary norms in academic writing shown in relationships among related genres.

Samraj (2005) comments on traditional move structures and typically found moves used in the genera of abstracts and introductions. A further argument of Samraj's study refers to abstracts from the Wildlife Behavior and Conservation Biology fields, which are similar in rhetorical structure; on the other hand, they vary in other aspects, i.e., they do not use traditional moves (Samraj, 2005), where moves can be defined as "a functional term that refers to a defined and bounded communicative act that is designed to contribute to one main communicative objective, that of the whole text" (Lorés, 2004, p. 282).

Apart from the functionality and features of abstracts, as well as the aspect of their evaluation and genre analysis, it is crucial to present the rhetorical structure of abstracts briefly. Stotesbury (2003) presents the following structures of abstracts proposed in the literature:

Table 3 Structures of abstracts (Based on Stotesbury, 2003)

Bhatia (1993)	Dahl (2000)	Stotesbury (2003)	ANSI (1979)
Purpose	Background	Topic	Scope and Purpose
Method	Purpose	Argument	Methodology
Findings	Methodology	Conclusion	Results
Conclusion	Results		Significant
	Comments on results		Conclusions

It is obvious that some of the moves of the rhetorical structure of abstracts are described by many scholars, whereas other moves differ substantially. Nonetheless, their specific structure serves communicative purposes, i.e., they are semantic units which convey the required content (Donesch-Jeżo, 2013). Rhetorical moves are indicated by the use of lexis, tenses, verb forms, or introductory phrases (Donesch-Jeżo, 2013). Some works on the rhetorical structure of abstracts suggest the general absence of purpose move (e.g., Ahmad, 1997), some suggest overall differences in organization (Melander et al. 1997), and others suggest the differences between closely related disciplines in terms of various aspects, like modality and tense use (Samraj, 2002a, 2002b). But what is crucial to underline is that rhetorical moves used by the academic community in any field can also be predetermined by the publication guidelines fixed by publishers of the given academic field. There is a twofold division of abstracts based on their macrostructure, defined by Swales and Feak (2012, as in Donesch-Jeżo, 2013): *Traditional* type without the predetermined sections and *Structured* type with specified sections. Interestingly, the second type appeared in 1987 in medical research articles (Hartley, 2004, as in Donesch-Jeżo, 2013).

The most commonly known structure, which can be found in use in the literature, is Swales' (1990) "Create a Research Space" (CARS Model). It served as a basis for the development of other models of rhetorical functions in abstracts:

Move 1: Establishing a territory

Step 1: Claiming Centrality

Step 2: Making Topic Generalizations

Step 3: Reviewing Previous Items of Research

Move 2: Establishing a Niche

Counter-claiming

Indicating a Gap

Question-raising

Continuing a Tradition

Move 3: Occupying a Niche

Step 1A: Outlining process

Step 1B: Announcing Present Research

Step 2: Announcing Principal Findings

Step 3: Indicating the Structure of the Research Article

The first move provides the context and sets the background of the study or article. The steps presented within it are optional, one can use whichever is suitable. The first step is commonly used in academic disciplines where the discourse community plays the role of the audience. It sets the paper in a significant, already established area of research. The second step clarifies the knowledge or common practices within the field of study. The third step relates to previous findings on the given topic.

The second move can be executed in one of four ways. Firstly, as a counter-claim to the earlier research. Secondly, by demonstration the still existing questions that have not been answered. Thirdly, by raising or suggesting an additional question about previous research. And, lastly, by the presentation of existing research which can be extended by some additional data or information.

The third move consists of several steps which somehow answer the questions presented in the second move. It can be done in several steps, dependent on the method used in the previous, i.e., the second move. Firstly, the most common way is to outline the purposes of the paper in progress, and then announce or describe the current research. Secondly, descriptions of the findings, although it is less common in the introductions or written works, as Swales (1990) concluded. Lastly, the third element, the preview of the article's structure closes the rhetorical structure of the third move.

Table 4 is presented to more fully describe the rhetorical functions of abstracts based on their structure. The rhetorical functions presented below are probably the most common in various research disciplines.

Table 4 Models of rhetorical functions in abstracts (Based on Swales (1990, p. 142), Hyland (2000, p. 67), and Lorés (2004, p. 283)).

Three move structure	Description	Four-move structure	Description	Five-move structure	Description
Establish territory	Research aim, goal, "current capacity".	**Introduction**	Outlines the purpose or objective, the goals of the research, or the problems that the author wishes to undertake.	**Introduction**	It should establish the context of the paper and the motives of the research or discussion.
Establish a niche	A space for research gap and possible research questions.	**Methods**	Should indicate how the problem has been studied, or the goal set out. Specifies methodology followed, and data used.	**Purpose**	It indicates the purpose, thesis, or hypothesis, the intention behind the paper should be outlined.
Occupy the niche	Description of evaluation criteria.	**Results**	A concise summary of the general findings.	**Method**	It provides information on, e.g., design, procedures, assumptions, approach, data.
		Discussion	It might include an interpretation of the results and some implications for further research or application of the findings.	**Product**	States the main discoveries or results, the argument, or accomplished outcomes.
				Conclusion	Results should be extended beyond the scope of the paper, interpretation given, inferences drawn, and broader implications pointed.

As can be seen, although Swales's CARS model originally refers to the introductions sections of a research article, it was also used as a model of a structure for the genre of abstract. Other models of rhetorical functions in abstracts, presented in Table 4, are loosely or more strictly based on the well-known Swales's CARS model. Despite the same foundation, visible in various rhetorical function models of abstracts, as Lorés (2004) concluded, there is a lack of analysis standards in the research genre of abstracts.

Members of any discipline, occupied with issues which interest them, are aware of changing contexts and demands of any new conditions their written works have to stand against and abstract seems to be their calling card, of any written research paper, which has to stand the pace of the academic community. The issue of interest presented in the following section turns to corpus and computational linguistics.

1.3 Corpus and computational linguistics

One may argue that the beginnings of corpus linguistics stem from the 13th-century Bible Scholars like Anthony of Padua (1195-1231) or Cardinal Hugo of St. Caro (~1230) and their works: Concordiante Morales (the first concordance of the Bible), Word index for "Vulgate". Others might take Cruden's Complete Concordance of the Holy Scriptures (1737), or Strong's Exhaustive Concordance of the Bible (1890) as a root in this case. These are significant works that influenced future generations' research, but modern-era scholars look for the notions and terms in the specified literature of their field. Thus, their perspective is somehow shaped by the society of scholars around the globe. A good part to start with here is the O'Keffe and McCarthy (2010) claim that modern corpora were shaped by three elements:

1. Pre-Chomskyan structural linguists - who worked as lexicographers,
2. 1755 Johnson's First Comprehensive Dictionary of English,
3. Oxford English Dictionary - starting from the 1880's slips of paper version

However, as they conclude, the term "corpus-linguistics" was coined by Aarts and Meijs in 1984.

Of course, time changes perspective. The American Structuralists of the 1950's probably knew the works of the Bible Scholars but had the resources to develop new ways of automatizing the work with the machines. The first computer-generated concordances appeared in the 50's. With the use of punch-card technology, they were able to process 60.000 words in 24 hours. At the same time the notion of collecting real data for analysis appeared.

Other influential scholars who have shaped modern corpus linguistics are Busa and Juilland. Between 1949 and 1967, they worked with IBM and developed a concordancer to

analyze a corpus of medieval philosophy texts. Juilland in his works "addressed issues of balance, covering a wide range of genres, representativeness, examining a range of writers within various genres of 500,000-word corpora and dispersion, developing Harris' concept of distributionalism" (Flowerdew, 2012, p. 38). Since they worked mainly on the Spanish language, the contribution of the two scholars is not widely known due to the language they studied, because corpus linguistics is strongly associated with the works on the English language (McEnery and Wilson, 2001). The work that stands out in many sources as the precursor of corpus linguistics is The Brown Corpus by Francis and Kucera (1967), the milestone which commenced the era of machine-readable corpora. It is referred to as the first-generation corpora, alongside LOB (Lancaster-Oslo/Bergen). While second-generation corpora examples are: British National Corpus (BNC), American National Corpus (ANC), and Bank of English (BOE), the third-generation corpus is the internet (Flowerdew, 2012).

What is interesting, the term *corpus* was first used as a notion in linguistic literature by Allen (1956) in his work Transactions of the Philological Society, with a meaning of "the body of written or spoken material upon which a linguistic analysis is based" (OED, 2009 as in O'Keffe & McCarthy, 2010, p. 5).

As can be seen, there is a gap of 28 years between the notions of *corpus* and *corpus linguistics*, during which some interesting works were made, of which a few elements are worth looking at from the perspective of establishing the foundation of modern corpus linguistics as it is today. Primarily, the study of first language acquisition is based on transcribed data, i.e., CHILDES Language Database from the 1960's. Secondly, in the 1970's the Key Word in Context (KWIC) concordances started replacing catalog indexing cards, which helped with the automatization of the subject analysis (Hines et al., 1970 as in O'Keffe & McCarthy, 2010).

Before computers started to play a significant role in linguistics in 1980's and 1990's, there were three, already secured traditions within the research society (O'Keffe & McCarthy, 2010):

a. "trawling through texts to find all examples of a particular piece of language

b. writing dictionaries based on attested usage

c. analyzing language based on actual informant data" (p. 5).

Despite the limitations of the early machines (punch-card data) that influenced the studies until 1980's, considering today's perspective, alongside the availability of today's software, the 1987 DOS-based Oxford Concordance Program still lacked the efficiency of the GUI-based software. The development of Monoconc (Barlow, 1996) or Wordsmith (Scott, 1996)

ceased the software limitations and GUI-based programs became widely accepted and easy-to-use tools of linguistic study (O'Keffe & McCarthy, 2010). The widespread of linguistic software was broadened even more by the work of Laurence Anthony since 2014, i.e., his open-source software like *AntConc* and other "Ant-like" programs.

Apart from the historical background of corpus linguistics, it is essential to cover some of the definitions and approaches established by the research community. Thus, Teubert and Čermáková (2004) maintain that corpus linguistics is interested in:

- "the change of frequency of words or other units of meaning (compounds, multi-word units, collocations, set phrases), which is often indicative of a change in meaning or a change in the domains in which words are used;
- the occurrence of new words;
- the occurrence of new larger units of meaning;
- changing context profiles, i.e., changes in the frequencies of words occurring in the contexts of words or other units of meaning" (p. 121).

Research interests are closely connected with their purposes, as the purpose of any research results from the interests of a given field. Sinclair (1992, pp. 379-380) distinguishes three purposes of corpus analysis, presented in Table 5.

Table 5 Sinclair's (1992) Corpus analysis purposes

Purpose	Definition
Specific	purposes where the end result is the only important one
General	purposes where the task is so complex that we have to rely on the application of linguistic principles
Language as an object of study	purposes where the application of pre-existing descriptions to corpora is made to check and improve

Additionally, purposes should reinforce a background for analysis. Then, corpus-based analysis can be characterized, after Biber, Conrad and Reppen (1998), as consisting of several steps. Firstly, constructing corpora from a large collection of natural texts which serve as a basis for the analysis. Secondly, conducting empirical analysis of natural texts, a search for patterns occurring within natural texts. Thirdly, using computers to enhance the quantitative and qualitative techniques of the analysis.

As Hasko (2012) argues, the qualitative and quantitative corpus analysis attempts to

deeply interpret and apply empirical data and insights gathered with the help of computer-aided retrieval of data, i.e., authentic examples of investigated fragments of language, to a vast range of intellectual explorations in language studies.

It is evident that the data needs to be carefully selected for research and that it should present elements in question. Based on the type of elements used for the corpus preparation. Teubert and Čermáková (2004, pp.118-123) present the types of corpora given below. Table 6 has been constructed to display a more precise visualization of the descriptions:

Table 6 Types of corpora (Based on Teubert & Čermáková, 2004)

Corpus type	Description
Reference corpus	Contain the standard vocabulary of a language. Reference corpus is the linguist's primary resource to learn about meaning. If they are large enough, they reveal the contexts into which words are typically embedded and with which other words form collocations.
Special corpus	A Corpus that fits a specific research focus. It has no standardized composition. It is usually based on a set of hypotheses that guide its definition. All in all, it is any reasonable creation of a researcher.
Opportunistic (Cannabalistic) corpus	This type of corpus should include information about the text such as genre/domain/text type, author's name, date of publication, title, and other bibliographically relevant information. This kind of corpus does not claim to represent a language or mirror a discourse; it assumes that each corpus is imbalanced. Some corpora of this type can fall into the category of reference corpora. However, almost all of them will be labeled 'special'.
Monitor corpus	This corpus type monitors language change, open-ended, and is updated regularly. To be relevant, it should retain its initial composition of text types, e.g., journal abstracts.
Parallel corpora (Translation corpus)	Parallel corpora are repositories of the practice of translators. It is a corpus of original texts in one language and their translations into another (or several other languages). Sometimes parallel corpora contain only translations of the same texts in different languages, but not the text in the original language. Thus, we can extract a wider variety of translation

	equivalents embedded in their contexts from parallel corpora, making them unambiguous. All of the above makes parallel corpora exceptionally attractive.

In corpus-based research, various possibilities exist, e.g., to select the most suitable type of a corpus, or pick an already existing one; construct a new corpus, or choose to use a virtual corpus. A type of virtual corpus may be additionally specified as follows:

> "recently it has become quite common among corpus linguists to consult the Internet as a virtual corpus. This is particularly useful when we want to find out if a word or a phrase we have heard really exists and in which kinds of texts it occurs. Whenever we cannot find evidence of words or units of meaning in our classic corpora, we can turn to the Internet. (…) The Internet is a virtual corpus, and, like the discourse of any language community, we cannot expect to access it as a whole. Normally, if someone wants to use the Internet as a source, they should, therefore, download all the texts they are working with, and compile them in a special corpus; and they should document them with their web addresses and other bibliographic information and the date of the download" Teubert and Čermáková (2004, pp. 124-125).

Teubert (2004, p. 97) argues that there is no secret formula for corpus creation, nor is there a formula in natural language. All in all, language is a social phenomenon in the eyes of corpus linguistics (Teubert, 2004).

One of the methods used to document the descriptive metadata is a corpus markup. As Hasko (2012) highlights, the corpus markup applies the use of standardized, coded information inserted into corpus files' raw data. It can be divided into three schemes, namely, document-wide information (e.g., distributor, date), structural elements (e.g. chapters, headings) and subparagraph structure (e.g., quotations).

The second method of labeling the data might be corpus annotation; it is no less than the coding of linguistic information. As she states, it may vary in degree of specificity and might address such elements as, e.g., part-of-speech tagging, parsing, or annotation of phonetic/semantic/discourse/metaphorical elements (Hasko, 2012).

It can be observed that *parsing* and *tagging* are the simplest, most valuable and straightforward tools. The most basic form of parsing, used for the automatic analysis of language, may comprise:

1. "decomposition of a complex sign into its elementary components,
2. classification of the components via lexical lookup, and
3. composition of the classified components by means of rules in order to arrive at an overall grammatical analysis of the complex sign." (Hausser, 2014, p.29).

Not only *parsing* and tagging, but all of the above forms of automatic analysis of

language can be performed using various programs, either free (e.g. AntConc) or licensed (e.g. WordSmith). The methodological importance of parsers of natural language lies in the objectivity and automaticity of the process of testing the descriptive adequacy of formal rule systems on actual data (Hausser, 2014). The conclusion is that the corpus-based analysis compiles more benefits, i.e., large data sets under analysis, automatic retrieval, other computer-aided methods, and ease in replication and sharing evidence, while qualitative types of analysis are limited in size and approach of accounting individual examples (Hasko, 2012).

It is worth mentioning that Hausser (2014) establishes three approaches to the analysis of natural language:

1) Traditional grammar
   a) "uses the method of informal classification and description based on tradition and experience,
   b) has the goal to collect and classify the regularities and irregularities of the natural language in question as completely as possible, and
   c) is applied mostly in the teaching of languages (originally Latin)" (p.23).

Theoretical linguistics discards traditional grammar, but computational linguistics is intensely interested in it because of the wealth of concrete data (Hausser, 2014).

2) Theoretical linguistics
   a) "uses the method of mathematical logic to describe natural languages by means of formal rule systems intended to derive all and only the well-formed expressions of a language (explicit hypothesis formation instead of the informal descriptions of traditional grammar),
   b) has pursued the goal of describing the 'innate human language ability' (competence), whereby aspects of language use in communication (performance) have been excluded, and
   c) has had rather limited applications because of its computational inefficiency and because of its fragmentation into different schools" (p.23).
3) Computational linguistics
   a) "combines the methods of traditional grammar and theoretical linguistics with the method of effectively verifying explicit hypotheses by implementing formal grammars as efficient computer programs and testing them automatically on realistic – i.e., large – amounts of real data,
   b) has the goal of modeling the mechanism of natural language communication – which requires a complete morphological, lexical, syntactic, semantic, and pragmatic

analysis of a given natural language within a functional framework, and

c) has applications in all instances of human computer communication far beyond letter-based 'natural language processing' (NLP)" (Hausser, 2014, p. 23).

As Hausser (2014) argues, formal language analysis and mathematical complexity theory are the two essential areas relevant for computational linguistics researchers. But, all of the approaches mentioned above use the same division of components (i.e., lexicon, morphology, phonology, syntax, semantics, pragmatics), though different methods, goals, and applications apply; the roles and methods of the components in which they are handled differ scientifically (Hausser, 2014).

One of the examples of a complex work which divides its components in the manner described by Hausser (2014) in computational linguistics/quantitative comparison studies is Biber's (2006) multi-dimensional approach. Biber (2006) states that quantitative comparisons determine whether the occurrence frequency under analysis commonly occurrs in target registers. Register treated as a continuous construct enables research concerning its core linguistic features full-spectrum, while the texts are situated in the continuous space of linguistic variation (Biber, 2006). Even though, as Biber (2006) argues, the distribution of individual linguistic features can be interpreted in functional terms, the reliable distinction of common patterns is not possible because of the number of different linguistic characteristics under consideration.

The multi-dimensional approach is primarily based on the tools that corpus-based research can provide. Biber (2006) argues that "the MD approach [multi-dimensional] was developed to analyze the linguistic co-occurrence patterns associated with register variation in empirical/quantitative terms" (p. 178). What Biber (2006) suggests is that to simplify matters, in any quantitative research which seeks to verify some of the complex linguistic phenomena, one must take into consideration many co-occurring linguistic features. Thus, he proposes the following distribution of linguistic features for corpus linguistic analysis based on a tagged corpus: vocabulary distributions, grammatical part-of-speech classes, semantic categories for the major word classes, grammatical characteristics, syntactic structures, lexico-grammatical associations and lexical bundles (Biber, 2006).

The data has to be considered in relation to selected linguistic features or research groups in a study, but it also has to compare the researched material in terms of selected variables, such as gender, nativity, or academic field settings. It might be examined for the markers of functions or other studied phenomena. For example, in this study, the most frequent AWL nouns, verbs, adjectives, and adverbs which occur frequently in the corpus are used to identify

the academic functions of abstracts in this research. Additionally, gender is one of the selected variables taken into consideration in this dissertation. This is why the next section presents gender aspects in linguistic research.

1.3.1 Gender in linguistic research

If one looks more broadly at the topic of researching gender-related schemes, the first thing that comes to mind is sociolinguistics. As Flowerdew (2012) argues much of the research in corpus sociolinguistics uses one of two approaches: interactional sociolinguistics or variational sociolinguistics. The first approach views language as a dynamic entity that is constructed in the process of interaction between people, most frequently in diverse speech acts; it analyses these conversations and situations from a pragmatic perspective. The other uses specified linguistic variables to research the language of the particular group selected for a study; the variables should be labeled with demographic data. This type of research is often diachronic and looks for changes in language (Flowerdew, 2012). Of course, there are some limitations to sociolinguistic corpus studies which a researcher has to face. McEnery et al. (2006, as in Flowerdew, 2012) present three of them:

- the operationalization of sociolinguistics theories into measurable categories suitable for corpus work,
- the lack of sociolinguistic metadata encoded in currently available software,
- the lack of sociolinguistically rigorous sampling in corpus construction.

The sociolinguistic variables used in corpus studies have to be simple enough so that they can be researched with the use of available software. Simple demographic information such as gender, nativity, or role in the academic community can be used as a significant variable in linguistic research. Even if, instead of encoding it into available software, the corpus data has to be stored in different folders or files to mark each of the variables separately. For this reason studies performed this way are feasible.

To the best of my knowledge, the importance of gender as a variable in linguistic research on abstracts was underlined by two papers. Firstly, Hyland and Tse (2006) claim that "we know very little about these gender-preferential features in academic writing and nothing about how such preferences interact with disciplinary preferences and conventions. The few studies that do exist, moreover, provide inconsistent findings" (p. 177). Secondly, Łyda and Warchał (2014) state that "we are not aware of any studies into the possible relationships between the gender of the author of the research article abstract and preferred – or dispreferred – structural, rhetorical or lexical choices" (p. 115). That is why gender-

preferential features within academic disciplines, as well as structural, rhetorical and lexical choices influenced by the gender of authors, lie within the scope of this dissertation.

There are, of course, works on this somewhat neglected aspect of gender-preferential features, but, largely it seems to be an influential characteristic of various research papers on social interactions, where the linguistic features used by men and women differ because of variables like interaction with others or expressing oneself (Hyland and Tse, 2006). Additionally, "while gender does not seem to be a major variable in writers' interactive choices overall, these choices are heavily influenced by disciplinary considerations which make gender an important source of disciplinary variation" (Hyland and Tse, 2006, p.178). Thus, it seems essential to consider gender alongside disciplinary cultures and to study it in an academic discourse setting.

What is interesting about the findings of Hyland and Tse's (2006) research is that it corresponds to the findings in Crismore et. al, 1993; Francis et. al. 2001; Johnson and Roen, 1992, and Herbert's (1990). The first two studies found that men used more boosters, while the remaining two sources claimed that females used more intensifiers while expressing compliments. A reason for these findings can be the male-dominant culture of academia, and higher career positions of men, which was also supposed by Kirsch (1993) in similar findings. Nevertheless, more important differences between the researched groups occur when disciplines are taken into consideration. For example, female philosophers use more transition markers and attitude markers, while male philosophers use more engagement markers and self-mentions. Secondly, in sociology, females use more code glosses, and males use more evidentials and hedges. What is more, in the field of biology, females use more attitude markers and self-mentions, while males use more transitions, code-glosses, and hedges (Hyland & Tse, 2006, p. 187, Table 3).

Łyda and Warchał (2014) emphasize that some previous works which analyzed academic texts, i.e., Lynch and Strauss-Noll (1987), Francis et al. (2001), Robson et al (2002), and Łyda (2006), reveal more similarities than differences between the sexes. But, there are also some which suggest trends towards the differences: in Montgomery (1995), males were found to care less about grammar form and politeness; Ellis et al. (2008) emphasize that females are better writers, write longer sentences, have better grades, but are less satisfied with their overall academic and intellectual abilities; Flynn (1988), and Rubin and Greene (1992) found that female writers tend to be less formal and more polite. Additionally, Łyda and Warchał (2014) underline the importance of several studies performed with a gender variable in question. Gender variable is used to assess different social expectations in Tannen

(1986, 1999, 2001), power relations between genders in Mills (2008), or communicative style affordances in Eckert and McConnel-Ginet (2003), McConnel-Ginet (2008).

Łyda and Warchał (2014) have found that male and female authors of research articles in linguistics do not rely on academic lexis to the same extent, although this trend is visible, and can be explained when other parameters in question are accounted for, i.e., native and non-native writer status. Secondly, disciplinary terminology used by the authors is dependent on the Native Speaker/Non-Native Speaker (NS/NNS) status as well. Łyda and Warchał (2014) analyzed texts from two journals *English for Specific Purposes* (ESP) and *Journal of Pragmatics* (JOP). The most prominent differences in academic lexis use between the females and males groups are present in ESP female writers, between NS and NNS female and male groups. In JOP, differences exist between native males and females as well as non-native groups in the same configuration. Moreover, diachronic changes in the data seem to confirm the differences, for JOP texts it is more visible than for texts presented in the ESP Journal.

As Łyda and Warchał (2014) conclude, "at the same time our study has brought not only answers but also further questions: our data and findings suggest even more complex interactions between gender and a whole array of variables, such as academic experience, academic recognition status and L1-L2 distance" (p. 128). Again, it seems substantial to consider gender alongside other variables or features while analyzing them in an academic discourse setting. The above-mentioned works confirm the importance of gender as a variable in linguistic research alongside various data, not only sociolinguistic but also specified by the field they refer to. As men and women have different preferences in the use of lexis and linguistic features, it is essential to investigate those matters properly with reference to all of the suggested elements of this complex linguistic phenomenon. Thus, the last section of this chapter presents the view on academic language function and academic vocabulary.

1.4 Academic Vocabulary and Academic Language Functions

The first section of this chapter presented communication within academic discourse. Academic vocabulary is inevitably a part of this discourse. It is utilized by anyone who is a part of, and communicates within the academic society. Regarding the definition of academic vocabulary, Baumann and Graves (2010) highlight that a plethora of terms and meanings suggest either similar definitions or different, commonly occurring terms. In addition, they discovered that "…terms such as: *general academic vocabulary, academic literacy, academic background, general academic words, domain knowledge, academic competence, linguistic knowledge, domain-specific vocabulary, content vocabulary, academic language, and*

academic language skills" (p. 4) are used interchangeably by the academic community. Thus, all of the above notions can constitute a part of academic vocabulary definition.

As Nation (2001) emphasizes, there are several reasons why academic vocabulary is regarded as important and valuable for the academic community.

1. It is common to a vast scope of academic texts, and not in other text types. Finding the common academic words shapes the research made in English for Specific Purposes. The possibility of creating a list of common vocabulary influenced, as Nation (2001) underlines, the "little research comparing the frequency of specific academic words in academic and non-academic texts" (p. 304).
2. It constitutes a significant number of words in academic texts. Nation (2001) enumerates two ways of measurement: firstly, the number of tokens (i.e., the coverage), and secondly, the number of types/lemmas/families.
3. It is usually less known than technical vocabulary. Some problems with understanding the academic vocabulary meaning among second language learners were underlined, as Nation suggests, by Cohen et al. (1988):
    a. "It was sometimes used with a technical meaning and sometimes not, and learners were not always aware of this.
    b. Learners were often not aware of related terms being used to refer to the same thing. That is, they did not pick up instances of lexical cohesion through paraphrase" (Nation, 2001, p. 305)
4. It is a vocabulary that can be used by English teachers to help learners to acquire specialized vocabulary in contrast to technical vocabulary which needs more knowledge based on the specific subject. Such perspective opens new possibilities for the academic lists, i.e., they would serve as an extension of the general service vocabulary, thus deserving attention of teachers and learners with no difference to/of the academic area they study. However, it is important to note that: "a difficulty with some academic vocabulary is that it takes on extended meanings in technical contexts, and in different technical contexts there may be quite different meanings." (Nation, 2001, p. 307).

Academic vocabulary is connected with the notions of word families, lemmas and types. Bauer and Nation (1993) argue that "from the point of view of reading, a word family consists of a base word and all its derived and inflected forms that can be understood by a learner without having to learn each form separately" (p. 253). They add that the main principle of the word family suggests that with the knowledge of the base word, the

identification of the other words of a family causes no endeavour (Bauer & Nation, 1993). That is why, e.g., Gardner (2013, as in Łyda, 2013) proposes the following structure of the English lexicon based on the notion of word families (see Figure 1).

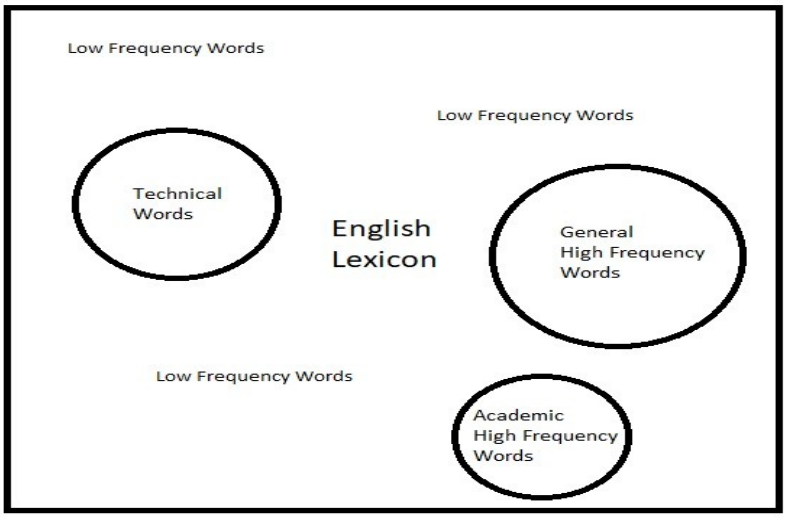

Figure 1 Structure of English lexicon (Gardner, 2013, p. 15, adapted from Łyda, 2013)

Word families can also be used to construct word lists. Coxhead's Academic Word List (2000, p 221) may serve as one of the examples. It was prepared according to the following three criteria:

1. *Specialized occurrence;*
2. *Range*;
3. *Frequency.*

Firstly, word families cannot be on the GSL list by West (1953) as the GSL is composed of 2,000 most frequent words, which are supposed to be used by English language learners and were sorted by the frequency of their occurrence. Secondly, a word in any family cannot occur less than 10 times in the main sections, and less than 15 in subject areas. Thirdly, a word in any family cannot occur less than 100 times in the Academic Corpus.

Unlike Nation (2001) and Coxhead (2000), who use word families in their approach to vocabulary analysis, Biber (2006) uses an alternative approach, i.e., lemmas. A lemma is a base form for any given word without inflectional morphemes. He proposes the following quantitative measures to diversify the vocabulary used for analysis in university registers:

A. "type/token ratio
B. mean word length
C. distribution of word types by part-of-speech

D. distribution of word types across registers
E. distribution of word types across academic disciplines
F. breakdown of word types by frequency level
G. interactions of part of speech/frequency level/register distribution" (Biber, 2006, p.242).

The above mentioned quantitative measures may be used to identify important linguistic patterns of variation between university registers, i.e., systematic difference between spoken and written discourses with respect to lexico-grammatical features and vocabulary characteristics; the difference between monologic and interactive registers; and differences between both spoken and written, academic and directive (behavior-management) registers (Biber, 2006).

Looking at the fourth point of the importance of academic vocabulary by Nation (2001), given above, one cannot forget about the basic aspects concerning the preparation of academic vocabulary lists or materials for English-language learners. As Martin (1976) states, there are four criteria which need to be considered in academic vocabulary preparation, e.g., for EFL students:

1. Unfamiliarity or incorrect use by the students.
2. Structural and contextual presentation of the material to help students with recognition of familiar items and extend knowledge to use unfamiliar items.
3. Usefulness in four categories: *speaking, writing, reading,* and *listening comprehension*.
4. Reinforcement of the essential academic skills like *paraphrasing, taking notes, outlining, writing papers, giving seminars,* and *taking exams.*

All of the above criteria, as she states, developed specific patterns appropriate for EFL teaching and learning.

Martin (1976) divides the academic vocabulary into the Research Process Vocabulary, Vocabulary of Analysis, and Vocabulary of Evaluation. In her view the EFL students have to integrate all three sets of academic vocabulary (VORP, VOA, VOE) into the activities performed in the university environment. In addition, the three-fold division mentioned here can be further divided into functions within each constituent point.

The following language functions used for the purposes of this dissertation, presented in Table 7, are constructed on the basis of Martin's (1976) research.

Table 7 Academic Language Functions (Based on Martin 1976)

Function	Academic Vocabulary
(F1) Formulating	Vocabulary of Research Process
(F2) Investigating	(Nouns and Verbs)
(F3) Analyzing	
1) Presenting methodology	
2) Design the experiment	
3) Simulate the situation/develop a model	
(F4) Reporting results	
(F5) Drawing conclusions	
(F6) Reviewing	Vocabulary of Evaluation
(F7) Evaluating	(Adjectives and Adverbs)
(F8) Criticizing	
(F9) Organizing order of information	Vocabulary of Analysis
(F10) Reinforcing research process	(Two-word Verbs, High frequency Verbs, and Nouns)

A detailed presentation of *academic language* in research on linguistic registers can be found in Ehlers-Zavala (2008) and Scott, Nagy, and Flinspach (2008). First, Ehlers-Zavala, defines academic language as "a specific register. (…) that students are expected to use in school subjects" (Baumann & Graves, 2010, p.5). The latter presents the following definition of academic language: "a register of English that has distinctive lexical, morphological, syntactic, and stylistic features" (Baumann & Graves, 2010, p.5). As Snow and Uccelli (2009, after Baumann & Graves, 2010) emphasize, the academic language is difficult for both native and non-native speakers, and what is more, there is less information on how to teach natives than English-language learners. A conclusion is that in most cases "academic language is represented as a rather extensive construct, somewhat akin to academic literacy" (Baumann & Graves, 2010, p.5).

Table 8 presents various typologies proposed between 2000 and 2010, given the similarities and differences in academic vocabulary definitions researched by Baumann and Graves (2010).

Table 8 Academic Vocabulary Typologies

Coxhead (2000)	Jetton and Alexander (2004)	Beck, McKeown, and Kucan (2002, 2008)	Pilgreen (2007)	Townsend (2008)	Fisher and Frey (2008)	Harmon, Wood, and Hendrick (2008)	Hiebert and Lubliner (2008)	Baumann and Graves (2010)
Academic words		Tier 2 words		General academic vocabulary	Specialized words		General academic vocabulary	General academic vocabulary
	"Language" of academic domains	Tier 3 words			Technical words	Academically technical terms	Content-specific vocabulary	Domain specific academic vocabulary
			Academic language				School-task vocabulary	Metalanguage
		Tier 2 words					Literary vocabulary	Literary vocabulary
					General words	Nontechnical words		
						Symbolic representations		Symbols
						Word clusters or phrases		

Baumann and Graves (2010), as well as Martin (1976) stress the importance of field-specific vocabulary and students' need to achieve fluency in it. Baumann and Graves (2010) conclude that "it is critical for learners to acquire the vocabularies of specific academic domains if they are to understand and learn the body of domain knowledge" (p. 6).

The last division that requires a comment is the difference between general and domain-specific academic vocabulary. As Baumann and Graves (2010) claim, there are two definitions of academic vocabulary used among the academic community: (1) *domain-specific academic vocabulary* and (2) *general academic vocabulary* (see Table 8). The former is linked with "domain-specific academic vocabulary [which] refers to the content-specific terms and expressions found in content area textbooks and other technical writing" (p. 6). The other, "general academic vocabulary is used to refer to words that appear in texts across several disciplines or academic domains" (Baumann & Graves, 2010, p. 6).

In this dissertation the following extended definitions of academic language and academic vocabulary have been proposed:

Academic language is a register used in an academic setting. It serves the purpose of communication within the academic society, in either spoken or written form. An

understanding of linguistic features of academic language can be acquired by being a member of academic society, i.e., using academic language on a daily basis, and/or mastered by learning the academic vocabulary of research process, analysis and evaluation.

Academic vocabulary is an essential part of academic language. It should be universal for all of the domains. The universal constituents of academic vocabulary can be found in any of the three parts: vocabulary of the research process, vocabulary of evaluation, and vocabulary of analysis.

Chapter Two
Approaches to abstract analysis

2.1 Introductory note

A vast majority of articles concerning journal article abstracts focus on their rhetorical structure. As stated in the previous chapter, the abstract structure seems to mirror the research study structure and its components. It has to mimic the rhetoric of a research article to catch the readers' attention successfully. The lack of research on the academic language functions of abstracts poses some difficulties concerning the material that can be presented and the practical approaches of such a study. However, some similarities seem to exist between the rhetorical structures and academic language functions of abstracts; the following summaries and reviews of the selected literature are presented to broaden the perspective on practical and theoretical approaches to abstract research. The summaries of the studies were presented one after another according to their research topics, whilst the concluding note tries to expose the similarities and differences of the presented studies.

2.2 Diachronic look at the abstract genre

The first study analyzed here, by Gillaerts (2014), focuses on applied linguistics's and diachronic rhetorical features in abstracts, particularly metadiscourse markers. Gillaerts (2014) selected sixty abstracts from The Applied Linguistics Journal of Oxford: 12 abstracts in each 5-year interval, between 1987 and 2007. The data was coded by two coders, including Gillaerts, who had the decisive say.

The use of Hyland's model of metadiscourse imposed two kinds of data division: interactive and interactional. The first one encompasses transitions, frame markers, endorphic markers, evidentials, and code glosses, whereas the other divides into boosters, hedges, attitude markers, self-mentions, and engagement markers. Gillaerts (2014) describes methodological issues concerning data coding under a selected model based on two approaches, thin and thick. The thin concentrates on quantitative occurrences and a list of metadiscourse categories. On the other hand, the thick determines how to label the specific instances, using context interpretation of metadiscourse instances occurring in a corpus. Gillaerts chooses to use the second approach. Additionally, this qualitative analysis is supported by quantification of the data for the research purpose of diachronic evaluation.

The analysis section in Gillaerts's (2014) diachronic study describes the density of metadiscourse markers used in abstracts per 100 keywords. There is a slight decrease in

metadiscourse markers' use between 1988 and 2002, and a visible increase in their use in 2007. In the material presented by Gillaerts (2014), there is a 13 percentage point increase in the use of interactive metadiscourse in total, and a decrease of the same value in interactional labeled instances. Based on the Gillaerts's previous studies, he claims that metadiscourse density tends to correlate with move density.

The comparison of specific markers in Gillaerts (2014) diachronic study reveals an interactive marker increase, but does not include endorphic markers. The frame marker section shows an increase from 15% (1987) to 21% (2007) of all studied examples, and the transitions change from 55% to 63% between 1987-2002. Furthermore, transitions decrease to 41% in 2007. The evidential section displays its peak in 1997 at 20% and a significant decrease in later years. In the interactional section one can see various changes, namely, an overall decrease in hedges, and an eight percentage point increase of boosters use across the studied years. Attitude markers peak in 1997 to 29% of all studied examples, and self mentions peak to 15% of all studied examples in the year 2002. What is interesting, engagement markers are diachronically stable (Gillaerts, 2014).

In Gillaerts' (2014) study metadiscourse markers' positions in the text are divided into two types, front and end. These are correlated with four moves: situate research, presenting research, findings, and conclusions. Thus, frame markers (34 examples) peak in front positions (i.e., at the beginning of sentences) along with evidentials (22 examples) and boosters (22 examples); all of these devices help to situate and present the research. There are almost no hedges (7) or self-mentions (1) in the front position. Interactionals are much more frequent in end positions: boosters (25), hedges (25), and attitude markers (19), and, additionally, transitions (24) are visible there.

In sum, the decrease in transitions is followed by an increase in frame markers and evidentials, as well as a decrease in hedges is somehow balanced by the gain of boosters and evidentials (Gillaerts, 2014). Thus, considering these two facts mentioned above, they tend to prove that abstracts' evolution shifts towards more complex text with informative and, what is more, persuasive intents in communication. Additionally, while transitions indicate reasoning and sign outlining, the lack of engagement markers suggests that the informative function is predominant. All of the above-mentioned demonstrates that the abstract genre can be more capable of evaluation than expected.

Gillaerts' diachronic study on metadiscourse markers provides a comprehensive analysis of the density and types of metadiscourse markers over a significant time frame divided by five-year periods between 1988 and 2007. This longitudinal approach allows for

observing trends and changes over time. The use of specific percentages and data points (e.g., 13 percentage points increase in interactive markers) adds precision and clarity to the findings. The division of metadiscourse markers into interactive and interactional categories, and further into specific types such as frame markers, transitions, evidentials, etc., helps in understanding the nuanced changes in the academic language used in abstracts. Linking the positions of metadiscourse markers to specific rhetorical moves in the text (i.e., situating research, presenting research, findings, and conclusions) provides insights into their functional roles in academic abstracts. The study highlights a shift towards more complex and persuasive texts, indicating an evolution in the abstract genre towards greater informativeness and persuasiveness.

While Gillaerts's study provides quantitative data, it lacks, in a few places, a deeper qualitative analysis of why these changes occurred. Understanding the reasons behind the increase or decrease in specific markers would add depth to the presented findings. The omission of endophoric markers in the analysis of interactive markers could be seen as a gap, as these markers also play a crucial role in guiding the reader through the text.

The claim that metadiscourse density correlates with move density is interesting but could be debated. Correlation does not imply causation, and further studies would be needed to establish a causal relationship. The functional roles of different types of metadiscourse markers (e.g., hedges vs. boosters) could be debated. Some might argue that the increase in boosters and decrease in hedges reflects a shift towards more assertive academic writing, while others might see it as a loss of nuance. The stability of engagement markers over time is noted as interesting. Still it could be debated whether this stability is due to their inherent nature or other external factors influencing academic writing.

The study mentions data from specific years (1987, 1992, 1997, 2002, 2007) but does not provide a continuous year-by-year analysis. This could overlook more subtle trends and fluctuations. An important aspect to explore further could be the research of changes in metadiscourse marker usage impact on the readability and comprehension of abstracts.

Diachronic evaluation used as a key factor is also presented, in the next section, by the study of Bondi's (2014), which focuses on the diachronic comparison of authorial voice and self-mention in economics, history, and linguistics.

2.3 Authorial voice in abstracts

Bondi (2014) set an interest in individual and social aspects of stance and its dialogic features, elements of disciplinary voice with a focus on the social dimension, in particular, forms of

authorial self-representation change over time. The research corpus of Bondi's (2014) work was based on three corpora, each comprising 300 abstracts from 1990, 2000, and 2010 in the fields of Economics and Linguistics. Economics abstracts were selected from 300 journals and 25 subjects. Linguistic abstracts were chosen from 137 journals and 20 subjects. The selected material was not balanced, namely, there were no 5 abstracts per subject for the year 1990 compared to other years. Additionally, historical abstracts were obtained from 60 journals, while its sampling concerned macro-categorization; the number of abstracts varied from 14 (1990) and 28 (2000) to 55 (2010).

Bondi (2014) combines discourse and corpus tools to concentrate on the prospective function of abstracts (i.e., influence on the reader and anticipating the structure of the argument). The study overviews the frequency data using Wordsmith 5 software, particularly markers of voice with change mapping over time and disciplines. Besides the lexical part, the search for collocation and phrasal patterns broadens her research. Syntactic patterns were researched in the view of semantic relations between elements. Thus, the study focused on variation and change in framing sequences, the locational and personal patterns, examining different verb types, as well as contrastive connectors, epistemic and attitudinal evaluative expressions, modal verbs, and evaluative adjectives. Next, wordlists were scanned for self-reference markers (i.e. illocution markers, argumentative connectors) and writer's stance markers (i.e., attitudinal adjectives).

Bondi (2014) presents the following outcomes. Self-mention "we" occurred 361 times (185 in economy, 139 in linguistics, 37 in history) throughout the corpus, with 183 total in 2010; 322 instances relate to authors, 30 to a generic researcher, and nine examples were excluded from research.

Personalization seems to gain ground in all researched corpora, emphasizing plural subjects for economics and linguistics (Bondi, 2014). Other self-referential occurrences (e.g., *study*) show a rise in the linguistics field: 18 (1990), then 23 (2000), and 60 (2010). Economics and linguistics have a higher average of self-referential nouns than history, respectively 2.90 and 2.65, to 1.77 (Bondi, 2014). The diachronic view reveals an increase in locational self-mentions and self-referential nouns for linguistics and history, and first-person self-mentions in economics.

Bondi (2014) claims that argumentative stance markers exhibit an increasing trend in using contrastive connectors (*but* and *however*). Though in the field of linguistics these constructions do not show an evident rise, the overall increase score is 113, 143, and 144, respectively, for 1990, 2000, and 2010. Furthermore, evaluative adjectives (e.g., *new* and

important) increased from 80 (1990) to 130 (2000) and 129 (2010) in the fields of economics and history. In linguistics *can* increases from 33 to 64, *may* from 16 to 22. While there is an overall increase for *could* from 16 to 30, and *may* from 44 to 59, in economics *should* increases from five to 17.

Overall, there is a rise in locational and personal self-mentions for all fields in Bondi's (2014) study. The historical sub-corpus relies on contrastive connectors, evaluative adjectives, and locational self-mention rather than personal references used in abstracts. Linguistics sub-corpus shows a preference for locational self-reference and modalization, whereas the economics sub-corpus visualizes average forms of all markers except for personal marker use, in the studied material (Bondi, 2014).

Bondi's study on self-mentions and argumentative stance markers provides detailed quantitative data on the frequency of self-mentions and other markers across different fields (economics, linguistics, history) and over time (1990, 2000, 2010). This allows for a thorough understanding of trends and changes. By comparing different academic fields, the study highlights how the use of self-mentions and argumentative markers varies across disciplines, providing valuable insights into disciplinary writing practices. The longitudinal approach offers a diachronic view, showing how the use of these markers has evolved over two decades. This helps in understanding the historical development of academic writing styles in that period. The study categorizes self-mentions into personal and locational, and further distinguishes between different types of argumentative stance markers (e.g., contrastive connectors, evaluative adjectives). This detailed categorization aids in a nuanced analysis.

While the study provides extensive quantitative data, it lacks, in some parts, an analysis of the reasons behind the observed trends. Understanding the contextual factors driving these changes would add depth to the findings. The heavy reliance on quantitative data might be seen as overlooking the subtleties and complexities of how these markers function in different contexts. One may argue that a balanced approach incorporating both quantitative and qualitative analyses would be more comprehensive. It could be noted that the study does not seem to address other potentially relevant markers, such as engagement markers or endophoric markers, which could provide additional insights into academic writing practices through further study.

An important aspect for further investigation could be an analysis of the contextual factors influencing the use of self-mentions and argumentative markers, such as changes in publication practices, or disciplinary norms. Additional discussion on how the changes in the use of these markers impact readability or persuasiveness, could be another aspect to explore.

The information on how the same changes influence the overall effectiveness of academic writing is scarce. These could be essential aspects to explore. Comparing the findings with those of other similar studies in different fields or languages can help in understanding whether the observed trends are universal or specific to particular disciplines.

The increasing trend of self-mentions, especially in economics and linguistics, could be debated. Some might argue that this reflects a shift towards more personalized and engaged academic writing, while others might see it as a move away from traditional objectivity. The rise in the use of contrastive connectors and evaluative adjectives could be debated in terms of their effectiveness. These markers, instead of enhancing the clarity and persuasiveness of academic writing, may introduce bias and subjectivity. The differences in the use of self-mentions and argumentative markers across disciplines could also be debated. These differences may be driven by the inherent nature of the disciplines or influenced by external factors such as publication norms or audience expectations.

All in all, the increasing use of argumentative stance markers, such as contrastive connectors and evaluative adjectives, suggests a trend towards more assertive and evaluative academic writing. This could reflect a broader shift in academic communication towards greater emphasis on argumentation and evaluation. The study reveals field-specific preferences in the use of self-mentions and argumentative markers. For example, the historical sub-corpus relies more on contrastive connectors and evaluative adjectives, while the linguistics sub-corpus shows a preference for locational self-reference and modalization. These preferences highlight the diverse rhetorical strategies employed in different academic fields.

Another study, presented below, partially taking into consideration the self-references, modalization, as well as human subject connotations, mainly in the form of pronouns, is the corpus study by Hyland and Tse (2004). It focuses on that-clauses in abstracts and researches two corpora comprising 465 abstracts in 6 fields.

2.4 A corpus study of evaluative "*that*" in abstracts

Hyland and Tse (2004) focus on "*that* retention" and omit "*that* sentences" used in other grammatical functions, such as relative and demonstrative pronouns. Twenty master's and doctoral abstracts were chosen from the following fields: applied linguistics, biology, business studies, computer science, electrical engineering, and public administration. Forty articles from leading journals from each discipline were used for comparison.

Hyland and Tse (2004) use WinMax Pro software to conduct a qualitative analysis.

They divide the researched clauses into four categories: evaluated entity, evaluative stance, source of evaluation, and evaluative expression, dependent on the role of that-clause usage in the text. The first category refers to connotations made in the studied abstracts with previous studies, statements of research goals, research methods, and models of theories evaluated. The second category is connected with attitudinal or epistemic stance and conveys, for example, affect or obligation. The third category attributes the evaluation to a general subject of a clause, i.e., a human source or an abstract and concealed entity. The fourth category corresponds to non-verbal and verbal predicates.

Hyland and Tse's (2004) analysis revealed 563 cases of that-clauses use, which gives the ratio of 5.5 per 1000 words. Differences between student and expert writers were evident. Experienced writers used that-clauses more frequently. Student-marked texts in social sciences used more that-clauses than in the fields of science and engineering. Further findings from the research presented by Hyland and Tse (2004) show that 80% of that-clause refer to writers' own findings, and 11% apply to the words of others, mainly in applied linguistics, business, and public administration. Nevertheless, most examples refer to generalized sources.

In half of the cases presented there, the sources are attributed to the study's results as a whole. Notably, only a small percentage of students' texts have a human subject attributed to them. Other groups, except electronic engineers, tend to preserve the presence of human-related evaluations. However, many fields where evaluation is considered without human reference to a subject, do it intentionally to strategy to strengthen the claim by maintaining personal interest. Some scholars account for removing the authorial self to demonstrate directness and intimidation, whereas students account for it for scholarly competence (Hyland & Tse, 2004).

Research results presented in Hyland and Tse's (2004) study, looking at the abstracts under evaluation, namely, the ones from dissertations' sub-corpus and articles' sub-corpus, show that the human subject in evaluation attribute plays the most important role in the computer science field in both corpora. The type of stance used in abstracts, in most cases, corresponds to epistemic judgment used in relation to the findings; this refers to doubt (35%) and certainty (55%). Students use certainty and doubt markers in a similar manner.

The use of certainty and doubt markers by both groups similarly occurs because of the tendency to use careful persuasion, to encourage readers to read the full text, which is desirable by all writers (Hyland & Tse, 2004). In consequence, electronic engineering has the most instances of certainty in dissertations sub-corpus and computer science in journal abstract sub-corpus (Hyland & Tse, 2004). Furthermore, "doubt" is most prevalent in the

fields of business studies and public administration. For both of the fields, it is present in dissertation and journal abstracts. In the dissertation sub-corpus, attitude and affect are most frequently used in computer science, and obligation in applied linguistics, whilst in the journal abstract sub-corpus there is no obligation visible. However, applied linguistics in journal sub-corpus has the highest number of examples of attitudinal and affect stances. According to Hyland and Tse (2004), a neutral stance occurred most in master's and doctoral abstracts in the field of public administration and in journal abstracts in the field of electrical engineering.

In Hyland and Tse's (2004) research, predicate forms of the evaluation show that verbal forms lead in both corpora, particularly in biology and public administration. Non-verbal predicates are almost twice as frequent in students' texts. Noun predicates lead in applied linguistics and computer sciences in thesis abstracts. While adjectives constitute 5.2% of examples of the total, verbal predicate forms occur in 86.9% (dissertation corpus), and 92.6% (journal corpus) of all texts (Hyland & Tse, 2004). The most frequent verbs in their research are: *show, demonstrate, find,* and *conclude,* but in soft knowledge fields, discourse verbs were more frequent. What is interesting, two-thirds of all discourse verbs occur in sentences with abstract subjects. As Hyland and Tse (2004) claim, it might suggest the preference for nouns over adjectives.

All in all, students use that-clauses appropriately to their discipline. This is due to the incorporation of evaluative that in technical or academic writing courses, which appear to be utilized to build confidence in making sophisticated judgments, thereby promoting students' texts in a manner suitable to their discipline.

Hyland and Tse's (2004) study of evaluative-*that* in abstracts covers a wide range of disciplines, providing a broad understanding of the use of that-clauses and stance markers in academic writing. The differentiation between student and expert writing offers valuable insights into the development of academic writing skills. The research provides specific data on the frequency and distribution of that-clauses, stance markers, and predicate forms, which can inform teaching practices. The identification of discipline-specific patterns helps tailor writing instruction to different academic fields.

The study focuses primarily on written academic texts, potentially overlooking the nuances of spoken academic discourse. The analysis is based on a specific set of disciplines, which may not fully represent the diversity of academic writing practices. While the study provides valuable insights, the findings may not be universally applicable across all academic contexts or cultural backgrounds. The reliance on quantitative data may not capture the full complexity of how that-clauses and stance markers function in academic writing.

The study suggests that that-clauses play a crucial role in constructing persuasive arguments, but there may be debate over the extent to which this is true across different disciplines. Some scholars might argue that other linguistic features are equally or more important in achieving persuasion. The recommendation to use concordance programs and focus on evaluative that-clauses in teaching may be debated among educators. There may be differing opinions on the best methods for teaching academic writing and whether these strategies are effective for all students. Overall, Hyland and Tse's (2004) study provides a valuable foundation for understanding the use of that-clauses and stance markers in academic writing, while highlighting areas for further research and potential debates in the field.

Establishing a relationship with readers by understanding how meaning is conveyed and how to accomplish persuasion is also a part of the following study by Alonso-Almeida (2014), who focuses on epistemic devices used in English and Spanish language abstracts.

2.5 Evidential and epistemic devices. A contrastive study of different language abstracts.

Alonso-Almeida's (2014) study aims to analyze and categorize evidentials and epistemic markers as well as check the relationship between genre and register in the corpus study. The data was obtained from the Corpus of Specialized Research Papers. It comprises randomly selected texts from several databases of scientific journals. Those with high impact factors were selected in the first place. The second selection factor concerned sociological features, e.g., writers' nativity. Alonso-Almeida (2014) stresses the role of the structural pattern of the abstract dependent on the move structure. He shows similarities between the four move strategies by Samraj (2005) developed on the basis of Santos (1996), and similar research by Martin-Martin (2003), to finally choose the fourfold distinction (Background, Purpose, Method, Results & Conclusions), by Ortega-Barrera and Torres-Ramirez (2010), for the research basis.

Alonso-Almeida (2014) compiled a corpus divided into two sub-corpora (English and Spanish) that consist of 60 abstracts. Data was tagged for moves and part of speech normalized to 10.000 words. Alonso-Almeida (2014) used the Pen Treebank tag set for the data optimization. Additionally, abstracts were tagged manually. Computational analysis was conducted in Onicom, a corpus tool of Alonso-Almeida's invention and design with the help of the Emerging Technology Applied to Language and Literature program from the University of Las Palmas de Gran Canaria.

Alonso-Almeida (2014) presents three perspectives to categorize evidentiality and epistemic modality, i.e., disjunction, inclusion, and intersection. The first is explained in Cornillie

(2009), the second in Chafe (1986) and Palmer (1991), and the third in Salkie (1996) and Dendale (1994). It is worth mentioning that Cornillie (2009) describes epistemic modality as an evaluative tool of likelihood and evidentiality as a reasoning process that results in a proposition. Inclusive approaches differentiate the two within the semantic nature of the device. On the other hand, the intersective approach asserts contextual categorization of epistemic markers as evidentials (Alonso-Almeida & Cruz-Garcia 2011). Alonso-Almeida (2014) chooses to treat both terms as separate concepts and underlines that evidentiality functions to show the authorial source of knowledge without excluding concepts of truth, reliability, and commitment; as he sees them, the effects, not functions of evidentials.

Results of Alonso-Almeida's (2014) study concerning lexical devices are divided into those with evidential or epistemic semantic load; modals were treated separately. Nouns, adjectives, adverbs, and verbs were chosen as the focus of the study. Alonso-Almeida (2014) reveals that in the researched material, rhetorical moves sometimes blend, particularly in the case of purpose and method moves. Furthermore, he states that some abstracts follow schemes outlined by the discipline, for example, in medical journals, which use labeling in the text before each paragraph. However, these text types are not the case in his study. He reported that the frequency of epistemic nouns is the highest in Spanish computing abstracts, and the occurrence of evidential nouns is the highest in medical English abstracts. Nouns have the highest frequency count in the Spanish sub-corpus. Additionally, in the English sub-corpus nouns are associated with judgements based on probability or assessment. The use of adjectives is not discipline or language-dependent; both sections seem to have an equal frequency. As Alonso-Almeida (2014, p.34) states, adverbs were not significant during his research. Verb tokens with evidential meaning outnumber epistemic verbs, in particular, in the Spanish sub-corpus.

Alonso-Almeida (2014) classifies evidential lexical verbs after Marin-Arrese (2009) into three matrices: experiential, quotative, and cognitive. The first category covers information gained through sense, e.g., *I see, it appears, that shows*. Alonso-Almeida (2014) interprets some examples concerning different verbs in use and proves them to be evidentials. For example, he argues that *show* and *monstrar* fall into the category of visual perception according to Dixon (2005), while the commitment and truth of these examples intertwine. They only inform how the authors obtained the knowledge, and, as a result, this fact reinforces these verbs as evidentials in experiential strategy. The second category, quotative, underlines third-party attribution (Alonso-Almeida, 2014). Cognitive evidentials represent the mode of knowing and include cognitive verbs. Both the second and the third category

examples occur only once in the researched corpus. Quotative examples are present in the Spanish sub-corpus, and cognitive examples are present in the English sub-corpus (Alonso-Almeida, 2014). Modal verbs with evidential attribution in the corpus are distributed chiefly in English medicine, Spanish law, and computing sub-corpora (Alonso-Almeida, 2014). There are nearly no modals with evidential or epistemic meaning in the computing sub-corpus for both languages.

In sum, authors of English abstracts use evidential and epistemic devices, along with modal verbs or lexical devices, as a strategy to avoid imposition. Authors of Spanish abstracts, on the other hand, use lexical evidential strategies and refrain from mitigating devices. Moreover, the English sub-corpus shows more politeness and collegiality. All in all, evidential and epistemic devices are both dependent disciplinarily and culturally. They are also dependent on the register used (Alonso-Almeida, 2014).

The study provides a detailed examination of lexical devices, focusing on nouns, adjectives, adverbs, and verbs, which offers a thorough understanding of their usage in different disciplines and languages. The classification of evidential lexical verbs into experiential, quotative, and cognitive categories facilitates an understanding of the the nuanced differences in how knowledge is conveyed. By comparing Spanish and English abstracts across various disciplines, the study highlights the cultural and disciplinary dependencies of evidential and epistemic devices.

The study's focus on specific sub-corpora (e.g., Spanish computing, English medical) may limit the generalizability of the findings to other disciplines or languages. The study mentions that adverbs are not significant, which might overlook potential subtle differences in their usage across disciplines and languages. The reliance on examples that occur only once in the corpus, i.e., quotative and cognitive evidentials, may weaken the robustness of the conclusions drawn about those examples.

Expanding the corpus to include more disciplines and languages could provide a more comprehensive understanding of the use of lexical devices. Further research into the role of adverbs in conveying evidential and epistemic meanings could fill the gap left by their minimal significance in this study. Investigating how contextual factors influence the use of evidential and epistemic devices could offer deeper insights into their variability.

The study suggests that the use of evidential and epistemic devices is both disciplinarily and culturally dependent. This raises questions about the relative influence of these factors and how they interact. The finding that English abstracts show more politeness and collegiality compared to Spanish abstracts could be debated, as it may reflect broader

cultural communication styles rather than specific disciplinary practices. The blending of rhetorical moves (e.g., purpose and method) in some abstracts suggests a need for further discussion on the methodological approaches used to categorize and analyze these moves.

Another study, which is similar to Alonso-Almeida (2014), takes the move structure into consideration as a part of the method used in researching the selected material. Suntara and Usaha's (2013) study, presented below, focuses on the rhetorical structure of article abstracts in two disciplines.

2.6 Research article abstracts in two related disciplines

The emphasis of Suntara and Usaha's (2013) study is put on the rhetorical structure of article abstracts in the disciplines of linguistics (L) and applied linguistics (AL). They claim, similar to Gillaerts (2014), that the study of abstracts could be an excellent basis for discourse analysis in various disciplines, as presented by, for example, Pho (2008) and Samraj (2002a).

Suntara and Usaha (2013) analyzed 200 abstracts from selected journals with high-impact factors published between 2009 and 2012. After several test approaches, they chose to use the methodology of 5-moves by Hyland (2000), which, in their view, covers the research fields best (Introduction, Purpose, Method, Product, Conclusion). Suntara and Usaha (2013) use a sentence as the unit of coding. The methodology for determining the conventionality of moves is covered by the Kanoksilapatham (2003) approach, which assumes that data (moves) can be accounted for as conventional when they exceed 60 percent of all examples. Some units were labeled double, according to Santos' (1996) "move embedding".

In Suntara and Usaha's (2013) study the median of moves used in abstracts varies between three and four. The most common pattern occurring in the researched text is P-M-Pr-C (Purpose, Method, Product, Conclusion) Moves of Purpose, Method, and Products were used the most. The Introduction move was counted as optional and occurred in only 45% of texts from linguistic journals and 44% of texts from applied linguistic journals. In the field of linguistics, the Conclusion move occurred only in 57% of the researched texts, whereas in applied linguistics it constituted 69%.

Next, Suntara and Usaha (2013) present data concerning 27 instances of gap description in the introduction move in abstracts and suggest that the field of linguistics (13 instances) tends to use the gap move in criticizing previous research. On the other hand, the field of applied linguistics (14 instances) describes a gap in previous research. The use of the tenses (present simple, present perfect and past tenses) for both disciplines in the introduction section was similar. Suntara and Usaha (2013) also explore the Purpose move. They examine

the deictic item along with the inquiry type and reporting verbs used. The quantitative use of inquiry type, e.g., *paper* (23 in L and 14 in AL), *study* (21 in L and 30 in AL), and reporting verbs, e.g., *investigates* (10 in L and 5 in AL), *examines* (6 in L and 13 in AL), as Suntara and Usaha (2013) claim, was similar for both researched fields, although one can see some differences looking at the examples given above. There is a minor move embedding of the Purpose and the Method move, similar to Santos's (1996) Presenting the research/Describing the methodology moves and Pho's (2008) Describing the methodology/Presenting the research moves occurrences in these studies.

The next rhetorical move, the Method move, was divided into three categories Purpose embedded, Product embedded, and Method only. Interestingly, no modal verbs occurred in the Method sections of the researched abstracts. There were a few instances of personal pronouns use ('*I*' and '*we*'), although there was no difference in their use, as well as, there were no differences in tense use and particular embedding methods between the two studied fields (Suntara and Usaha, 2013).

Furthermore, they describe Product move, marked as conventional in both fields of their study. This move shows the opening noun variation within the studied corpus, e.g., *result* (26 in L and 24 in AL), *finding* (7 in L and 19 in AL), *analyses* (6 in L and 9 in AL). The reporting verbs were similar for both fields. Suntara and Usaha (2013) suggest that self-reference words are similar to those from Hyland's (2000) research. The use of 'that' complement in their study was not generally did not refer to the subject of the studies themselves, but to the specific elements like data or results, in 90% of the cases. Additionally, the overall use of tenses in this section shows that abstracts in L journals tend to use present simple more often. On the other hand, the abstracts in AL journals use singnificantly more of the past simple tense.

The Conclusion move was optional for the field of linguistics and conventional in applied linguistics. There was a preference for the present tense use in the Conclusion move. The use of self-reference pronouns was present in their research material, similar to Hyland (2003), who found that it is most likely to occur in the introduction and at the end of abstracts with a goal of self-promotion. Pho's (2008) study also underlined this trend, where writers were using present tense as a strategy to explicitly present the author's presence in a text. Additionally, in Suntara and Usaha's (2013) study, the use of modal auxiliaries was found in the fourth move, specifically in sentences referring to possibility, ability and obligation. In sum, their analysis of the move pattern suggests an increasing trend of Introduction and Conclusion moves to occur. This trend was also indicated by Hyland (2000). Moreover,

Suntara and Usaha (2013) agree with Pho's (2008) research, which does not take the Introduction move as conventional, unlike Hyland's (2000) research. Although the Introduction move was not found to be conventional in both fields, what is interesting, more than half of the texts used it to acquaint readers with the background of the presented work.

The study of Suntara and Usaha (2013) on research article abstracts provides a thorough examination of the different rhetorical moves (Purpose, Method, Product, Conclusion) used in abstracts, offering valuable insights into their frequency and patterns. By comparing linguistic and applied linguistic abstracts, the study highlights differences and similarities in rhetorical strategies across these fields. The use of quantitative data to support findings, such as the frequency of specific moves and tenses, adds strength to the analysis.

The study focuses on a specific set of journals within linguistics and applied linguistics fields, which may limit the generalizability of the findings to other disciplines or broader contexts. The classification of certain moves as optional (e.g., Introduction and Conclusion in linguistics) might overlook their potential importance in certain contexts. The study notes minor embedding of Purpose and Method moves but does not explore this phenomenon in depth. This kind of further exploration could provide additional insights.

Expanding the corpus to include more disciplines and a wider range of journals could provide a more comprehensive understanding of rhetorical moves in abstracts. Further research into the embedding of moves, particularly Purpose and Method, could reveal more about the complexity of abstract structures. The study notes the absence of modal verbs in Method sections and the slight use of personal pronouns but does not explore the implications of these findings in detail.

The study finds differences in tense usage between linguistic and applied linguistic abstracts, which could be debated in terms of their impact on readability and clarity. The use of self-reference pronouns for self-promotion and author presence raises questions about the balance between objectivity and personal voice in academic writing. The increasing trend of Introduction and Conclusion moves suggests a shift in rhetorical strategies, which could be debated in terms of its significance and underlying causes.

Another study similarly focused on the rhetorical structure of article abstracts is the analysis of Tu and Wang (2013), who try to research patterns of rhetorical structure and find potential pedagogy-oriented instructions.

2.7 Tense analysis and rhetorical structure in JAA

Tu and Wang's (2013) study focuses on tense variation, mainly the reporting verbs within

structural transitions of the abstracts. They stress the importance of context, which significantly influences the patterns of rhetorical structure organization and describe the communicative role of each element used.

Tu and Wang (2013) present three theories used in the research. First, a CARS model by Swales (1990); second, the four-move theory by Ventola (1994a); and third, the IPMPrC structure proposed by Hyland (2000). The functions and moves of the theories mentioned here are described in the first chapter (see Table 4). A total of 1000 journal article abstracts, which constitute their research corpus, were examined. Tu and Wang's (2013) corpus consists of four Journal extracts; 250 examples each. Mono Conc Pro and WordSmith programs were used in the study to convey the qualitative analysis.

Tu and Wang's (2013) research results bring two-dimensional outcomes: illustrate verb tense and structure used in abstracts. They found that reporting verbs and be-verbs are prevailing. Moreover, frequent present tenses use is visible in Journal of Pragmatics (JOP), Journal of Second Language Writing (JSLW). The past tenses were found to be commonly used in Research in Reading (JRR), and Reading and Writing Journal (R&W). The rhetorical move analysis, on the other hand, has shown the usage of all three rhetorical structures presented by Tu and Wang (2013) with a highest percentage of four-move usage constituting 53.9%. Other rhetorical models in Tu and Wang's (2013) study constitute 28.8% (IPMPrC), and 17.3% (CARS) of the corpus.

Tu and Wang's (2013) study asserts that organization structure and verb-tense variation are robustly correspondent. They provide detailed quantitative data on the frequency of reporting verbs and be-verbs across different journals, offering a clear picture of verb usage patterns. The study highlights how verb tense usage varies across disciplines by comparing different academic journals (Journal of Pragmatics, Journal of Second Language Writing, Research in Reading, and Reading and Writing Journal). The inclusion of move analysis, showing the usage of different structures (four-move, IPMPrC, CARS), adds depth to the understanding of abstract organization. The use of tables to present data on reporting verbs and verb tenses helps in visualizing the findings clearly and concisely.

There are several aspects which may be debatable in Tu and Wang's (2013) study. Firstly, the observed trends presented in the study could benefit from a more thorough qualitative analysis rather than relying on quantitative data. A deeper examination of the contextual factors influencing these trends would enhance the clarity the presented findings. Furthermore, the study presents only a single example sentence for each examined verb and rhetorical pattern, which undermines the reliability of the presented tendencies. Additionally,

a comparison with similar studies conducted in different language or academic field could help determine whether the presented trends are universal or discipline-specific.

Moreover, the study notes that the pedagogical element was not conducted due to a lack of research design. This is a significant drawback as it limits the practical applicability of the findings. Another debatable point in Tu and Wang's study is the effectiveness of different move structures (four-move, IPMPrC, CARS) in enhancing the clarity and persuasiveness of abstracts. Tu and Wang do not provide any information whether these structures are equally effective across different disciplines or not, which leaves this question open for further investigation.

The study highlights distinct patterns in verb usage across different journals, with present tenses being more common in some journals and past tenses in others. This reflects the varying stylistic conventions in different academic fields. The preference for the four-move structure in over half of the abstracts analyzed suggests a dominant organizational pattern in academic writing. This information could be helpful for novice writers in structuring their abstracts.

The correspondence of tense variation and the organization structure is also present in the rhetorical characteristics of the genre of abstract studied by Busch-Lauer (2014), discussed in the following section.

2.8 Cross-linguistic, disciplinary and intercultural perspectives

The main focus of the Busch-Lauer (2014) research is a search for a connection between rhetorical structure and communicative characteristics of the abstract genre in the academic field. She emphasizes abstract prominence as a scientific genre and describes abstracts' primary functions as informative and evaluative, but the elements such as categories of abstracts, text composition, language features, as well as disciplinary, intercultural, and cross-linguistic aspects are also discussed.

Busch-Lauer (2014) uses a broadened definition of abstract from International Organization for Standardization (ISO) guidelines for writing abstracts. She defines abstracts as the representation of the original document without its interpretation or information about its author. Similar to Gillaerts (2014), Alonso-Almeida (2014), and Suntara and Usaha (2013) she claims that the abstract genre might be a basis for specialist vocabulary research, as well as preparing exercises for the foreign language classroom. A corpus used by Busch-Lauer (2014) is composed of her previously researched material, i.e., 30 English abstracts from 3 fields: linguistics, medicine, and nanotechnology. The features of abstracts are partitioned

according to two perspectives: (1) discipline and structure research; and (2) cross-linguistic and intercultural aspects. As Busch-Lauer (2014) remarks, the applied guidelines widely differ between disciplines, which is why the significant problems of researching abstracts in this view relate to, e.g., proper condensation of information, the difficulty of separating important information, and the composition of the text.

The first factor of disciplinary aspects discussed by Busch-Lauer (2014) is the length of abstracts. She refers to several studies, such as Tibbo (1992) and Melander et al. (1997), which established typical lengths of abstracts across different disciplines. Busch-Lauer (2014) also attempted to align with the previous findings of Orasan (2001). She has found an average of 175 words (7.4 sentences), with computer science text as the longest in the corpus, at 232 words, followed by chemistry at 215 words, biology at 196 words, and linguistics at 150 words. Additionally, the length is dependent on the type of document (RA, case study, or review). The descriptive or indicative forms determine short humanities abstracts; all other categories mentioned above tend to have an informative form (Busch-Lauer, 2014). Linguistic abstracts, as Busch-Lauer (2014) states, vary in length and structure. Abstracts from journals that use guidelines seemed to be better, as well as, double-blind peer reviews influenced the quality of the texts (Busch-Lauer, 2014). The genre of a journal differentiates medical abstracts. Characteristically, the direct author involvement does not occur in medical abstracts (Busch-Lauer, 2014). What is more, technology abstracts provide impersonal, informative texts.

Some deviations in abstract structure can occur according to discipline, subject, type of original text, or individual style of the authors (Busch-Lauer, 2014). The abstracts of non-native speakers often fail to meet the expectations of native speakers. Additionally, non-native authors' texts are often refused for being incoherent. Therefore, as Busch-Lauer (2014) emphasizes, the use of ISO standards should be more widespread, and international guidelines should be straightforward and readily available for novice writers to avoid mistakes.

Bush-Lauer's (2014) cross-linguistic perspective on German and English abstracts highlights four trends that appear to have changed since her previous research, conducted in 2002. First, the majority of journals provide guidelines to authors. Second, informative abstracts tend to overtake the field of linguistics. Third, the limitation of words forces authors to be more precise; they leave metadiscourse structures in favor of abstract structure, including topic presentation, purpose explanation, and results description. Fourth, the overall quality of texts has improved, probably due to strict reviews of scientific texts (Busch-Lauer, 2014).

Busch-Lauer (2014) provides a detailed examination of abstract lengths across various disciplines, supported by multiple studies (e.g., Tibbo, 1992; Melander et al., 1997). The study highlights differences in the lengths and structures across disciplines, offering valuable insights for academic writing. It also emphasizes the importance of journal guidelines and double-blind peer reviews in improving the quality of abstracts.

Some conclusions may overgeneralize the findings, not accounting for variations within disciplines or individual author styles. The undertaken theoretical approach seems to rely heavily on previous studies, which may not fully capture current trends or emerging disciplines. There is limited discussion on how digital tools and platforms might influence abstract writing and publication practices. While it mentions the need for guidelines for novice writers, it does not provide specific recommendations or strategies to support them.

There are a few points which seem to be debatable in the study of Busch-Lauer (2014). Firstly, the need for standardized guidelines does not align with the flexibility required to accommodate different disciplines and individual styles. Secondly, balancing the quality of abstracts with the need to make them accessible and understandable to a broader audience, including non-native speakers, presents a significant challenge. Thirdly, the role of peer review, namely maintaining quality, can turn into suppressing innovation and diverse perspectives in abstract writing.

2.9 Closing remarks

As can be seen in the studies presented above, the approaches used by scholars while researching the genre of abstract are various. Some studies focus on the interpretation of context and the quantification of data support for diachronic evaluation, as well as markers of voice with temporal mapping in Wordsmith, and the search for collocations and phrasal patterns. Others try to search for four categories of evaluative clauses and analyze them quantitatively, or present three perspectives on categorizing evidentiality and epistemic modality using corpus tools. There are also analyses of abstracts in high-impact factor journals in search for conventional rhetorical moves, using a sentence as a unit of coding, and the use of programs like MonoConc Pro and WordSmith to make a qualitative analysis in the search for the rhetorical structure of abstracts in 1000 journal articles. Finally, research features of abstracts within a discipline and their cross-linguistic and intercultural aspects, which seem very broad topics to cover in a journal article, were also presented. These diverse methodologies highlight the multifaceted nature of abstract genre research and underscore the importance of integrating various analytical approaches to gain a comprehensive

understanding of this field.

Gillaerts (2014) notes a shift towards more intricate and persuasive texts, signifying an evolution in genre of abstracts towards enhanced informativeness and persuasiveness. In Bondi's (2014) study the linguistics sub-corpus reveals a preference for locational self-reference and modalization, while the economics sub-corpus shows average forms of all markers, except for the use of personal markers. Notably, two-thirds of discourse verbs studied by Hyland and Tse (2004) are found in sentences with abstract subjects, suggesting a preference for nouns over adjectives. In Alonso-Almeida's (2014) study, authors of English abstracts tend to use evidential and epistemic devices, as well as modal verbs or lexical strategies, to avoid imposition. Conversely, Spanish abstract authors prefer lexical evidential strategies and steer clear of mitigating devices. Additionally, the English sub-corpus displays more politeness and collegiality (Alonso-Almeida, 2014). Suntara and Usaha's (2013) analysis suggests a growing trend of Introduction and Conclusion moves in abstracts. Tu and Wang (2013) found a high usage of all researched rhetorical structures, with a prominent 53.9% four-move usage. According to Busch-Lauer (2014), linguistic abstracts vary in length and structure, with those following journal guidelines and undergoing double-blind peer reviews showing improved quality.

Concerning similarities between the presented studies, one can find that all of them examine the rhetorical features or structures of academic abstracts, indicating a shared interest in understanding how these texts function within their respective disciplines. The use of qualitative and quantitative methods is common: many studies combine quantitative data analysis with qualitative approaches to enrich their findings. For instance, both Gillaerts (2014) and Alonso-Almeida (2014) employ a mix of computational analysis and qualitative coding. Several studies utilize established rhetorical models (e.g., Hyland's model, Swales' CARS model) to guide their analysis, reflecting a common methodological approach in the field. In reference to data sources used, most studies utilize abstracts from high-impact journals or specific academic corpora, ensuring that the data is relevant and reflective of current scholarly practices.

Regarding a wide range of issues analyzed in the presented studies, it might be stated that while Gillaerts (2014) concentrates on metadiscourse markers, Bondi (2014) emphasizes authorial voice and stance showcasing a broader range of rhetorical elements across studies. In contrast, Busch-Lauer (2014) focuses on the communicative characteristics of abstracts, highlighting genre-specific issues. Furthermore, each study employs different data coding methodologies and software tools. For example, Alonso-Almeida (2014) uses Pen Treebank

tagging, while Suntara and Usaha (2013) apply a specific move-based methodology, demonstrating the diversity in analytical approaches. Several studies focus on diachronic analysis, such as Gillaerts (2014) and Bondi (2014), which look at changes over time, whereas others, like Hyland and Tse (2004) and Tu and Wang (2013) do not emphasize a temporal dimension in their analysis.

There are some minor contrasting views on abstracts, namely, between Suntara and Usaha (2013) and Busch-Lauer (2014). The first see abstracts as neglected by discourse analysts in a previous decade and the other claims the opposite in reference to diachronic theory studies.

The studies cover a range of academic disciplines, from linguistics and economics to medicine and technology, each contributing unique insights related to their specific fields. Overall, the examined studies reveal a cohesive effort to analyze the rhetorical structures of academic abstracts, using a variety of methodologies and frameworks. Despite their commonalities, the studies differ in focus, methodology, and data characteristics, reflecting the diverse nature of research in academic writing and discourse analysis.

Chapter Three
Corpus Analysis Design

3.1 Introductory note

The following chapters focus on computational analysis and corpus analysis and provide examples and statistical data of selected conference abstracts from around the world. The corpus data, presented below in the form of tables, helps to give a better picture of academic language and its functions used in the conference abstracts.

3.2 Description of the data

The corpus consists of 480 conference abstracts from five conferences in 2016, each situated on a different continent. The means of categorization and calibration of the corpus data are presented according to three academic fields: Humanities (Applied Linguistics and Education), Biological and Health Sciences, and Computational Statistics. The corpus has 294,780 tokens and 91,097 types. The conference 'abstract books' used in this work are:

1. useR 2016 Conference. Book of Abstracts. Stanford University, California, USA. June 27-30, 2016.
2. EDEN 2016 Annual Conference. Re-Imagining Learning Environments. Book of Abstracts. Budapest, Hungary. June 14-17, 2016.
3. International Society of Behavioral Nutrition and Physical Activity 2016 Annual Meeting. Abstract Book. Cape Town, South Africa. June 8-11, 2016.
4. Applied Linguistics Association of Australia (ALAA) Annual Conference. List of Abstracts. Monash University. December 5-7, 2016.
5. ICHA The 17th International Conference of Harmful Algae. Abstract Book. Florianopolis, Santa Catarina, Brazil. October 9-14, 2016.

The special corpus has been designed with the aim of determining quantitative, and to some extent, qualitative differences in the use of academic language functions in various fields of written academic language. The data of the corpus provide:

- reference number and page number
- source
- author, gender, speaker type, institution, and speaker origin
- discipline, title, abstract text
- authorial stance (i.e., personal pronouns, if present)

- deictic item referring to type (if present)
- nouns, verbs, adjectives, adverbs
- abstract tagged for parts of speech

3.3 Aims of the research

The present study presents the research results, which comprise formal and functional analyses of academic language functions used across the disciplines in the examined material. The summarized aims of this study stand as follows:

- formal/structural analysis of the academic language used in abstracts by both sexes in all of the characterized fields and their comparison;
- functional analysis of specified academic language functions present in the corpus and their comparison with regard to the gender of the speakers, their origin, and field of study;
- analysis of the possible connection between formal and functional means of specified functions used in academic writing.

3.4 Hypotheses

On the basis of the preliminary analyses of academic language functions use, it can be hypothesized that:

1)
   a. Null hypothesis 1: There will be no differences in the frequency of occurrence of any academic language function according to the authors' gender, origin, and academic fields; functions will be distributed randomly in the academic fields.
   b. Alternative hypothesis 1: There will be differences in the frequency of occurrence of some academic language functions according to the authors' gender, origin, and academic fields; some functions will be found more frequently in the academic fields concerning Humanities.

2)
   a. Null hypothesis 2: It is expected that there will be no differences in academic vocabulary use and types of language functions (their formal realizations), in texts, for both gender and speaker origin.
   b. Alternative hypothesis 2: It is expected that there will be differences in academic vocabulary use and types of language functions (their formal realizations), in texts, for both gender and speaker origin. The occurrence of certain functions will be

higher and/or lower than those expected by chance. Vocabulary used by both sexes in different academic fields, and the specified origin of the speakers should be higher and/or lower than those expected by chance.

3)

a. Null hypothesis 3: Presumably, there is no connection between the functional and formal means of the researched academic language functions.
b. Alternative hypothesis 3: Presumably, there is a connection between the functional and formal means of the researched academic language functions.

Pearson's Chi-squared tests (if possible) or Correlation will be used to check whether the null hypotheses are false/true.

3.5 Method

After manually preparing the corpus by transferring data from the books of abstracts into Microsoft Excel and labeling it according to the scheme presented in section 3.2, the speaker origin labels were assigned based on the authors' last names and their affiliations. Prior to this, all names were verified through online searches, including Google Scholar and ResearchGate.

An effort was made to categorize speakers by type (scholar versus non-scholar) based on their affiliations. Authors of abstracts who represented a University (the information was found next to an abstract, enclosed in the book of abstracts) were given the appropriate labels of scholar. However, this category (scholar/non-scholar) was ultimately excluded due to a significant disparity in quantity: 392 texts were attributed to scholars compared to 87 texts from non-scholars (see Table 15).

Subsequently, data from Coxhead's (2000) Academic Word Lists were processed using TagAnt to separate the text into distinct parts of speech. This data was manually verified for accuracy and then transferred to Microsoft Excel, where it was saved into separate .txt files in UTF-8 encoding.

The corpus data within Excel was manipulated to generate quantitative data, tables, and figures to provide a numerical overview of the corpus. The abstract texts were copied into individual .txt files and organized into specific folders based on factors such as gender, speaker type, speaker origin, and discipline.

This .txt data was later analyzed in ProtAnt, using the appropriately coded AWL .txt files for parts of speech, with each part of speech being scanned separately. The same process was repeated for the subsequent categories:

1) Females and Males
2) Natives and Non-natives
3) Native Females and Native Males
4) Non-native Females and Non-native Males
5) Computational Statistics, Humanities, Biology and Health Disciplines
6) Females and Males within Disciplines
7) Native Females and Native Males within Disciplines
8) Non-native Females and Non-native Males within Disciplines

The acquired outcome data saved in MS Excel, after minor corrections, served as a basis for the preparation of tables presented in the fourth chapter, sections 4.5 and 4.6.

In the next step, a labeling system has been designed on the basis of Martin (1976)(see Table 7). The AntMover program data has been changed to suit the corpus data handling appropriately. These actions required the manual conversion of the corpus data to serve as training data for the program. Thus, 25% of the corpus data selected for the research have been labeled by hand and copied, firstly into .txt files, and then into the program data folders. A sentence has been used as the unit of coding. The AntMover helped to automate the work with the corpus data. The only disadvantage of the software is the lack of a double-labeling option. However, based on the preliminary labeling work, it has been found that the double-function sentences in the corpus data constitute less than 5% of the examples. Therefore, the decision for final labeling was to choose one function for each sentence in question. The AntMover program labeled the remaining 75% of the corpus data. Then the data, labeled with functions, were checked and saved manually into .txt files. Section 4.8 shows the excerpt from the labeled corpus, containing samples of the real language used in conference abstracts of the corpus. Samples were prepared manually, with each sentence representing one function. Next, the labeled .txt files from AntMover have been run through ProtAnt, using the prepared list of academic language functions. ProtAnt made a data summary, which was then copied into MS Excel and the tables for section 4.9 were prepared:

1) The number of the specified (1-10) academic language functions used within academic fields
2) The overall number of academic language functions used by Females and Males in academic fields
3) The number of the specified (1-10) academic language functions used by Females and Males in academic fields, overall

Section 4.9 shows the outcomes of the quantitative analysis of academic language functions.

Sections 4.9.1-4.9.3 present the outcomes of the statistical analysis of the academic language functions in use. The statistical tests of Pearson's Chi-squared were prepared for the three categories of academic language functions, i.e., VORP, VOA, and VOE, for the function numbers 1-10 in the academic field, gender, and speaker origin divisions. Next, the quantitative data for the use of academic language functions 1-10 by native and non-native writers, and their gender, within the academic fields was prepared. Further, statistical tests were conducted for the academic language functions 1-10 for all native and non-native writers' texts within academic fields. Lastly, the quantitative data was extracted for the average length of the abstracts in the corpus, and the tables for this section were prepared.

The following sections 3.5.1-3.5.4 present a brief description of how to use the computational programs selected to execute this research. Section 4.7 presents a similar description of the AntMover used to prepare the data in Sections 4.8 and 4.9.

3.5.1 **TagAnt**

It is a simple freeware software for tagging, which uses the TreeTagger engine (Anthony, 2014). The program was used to tag parts of speech in the corpus. It is simple to use and has only two functions.

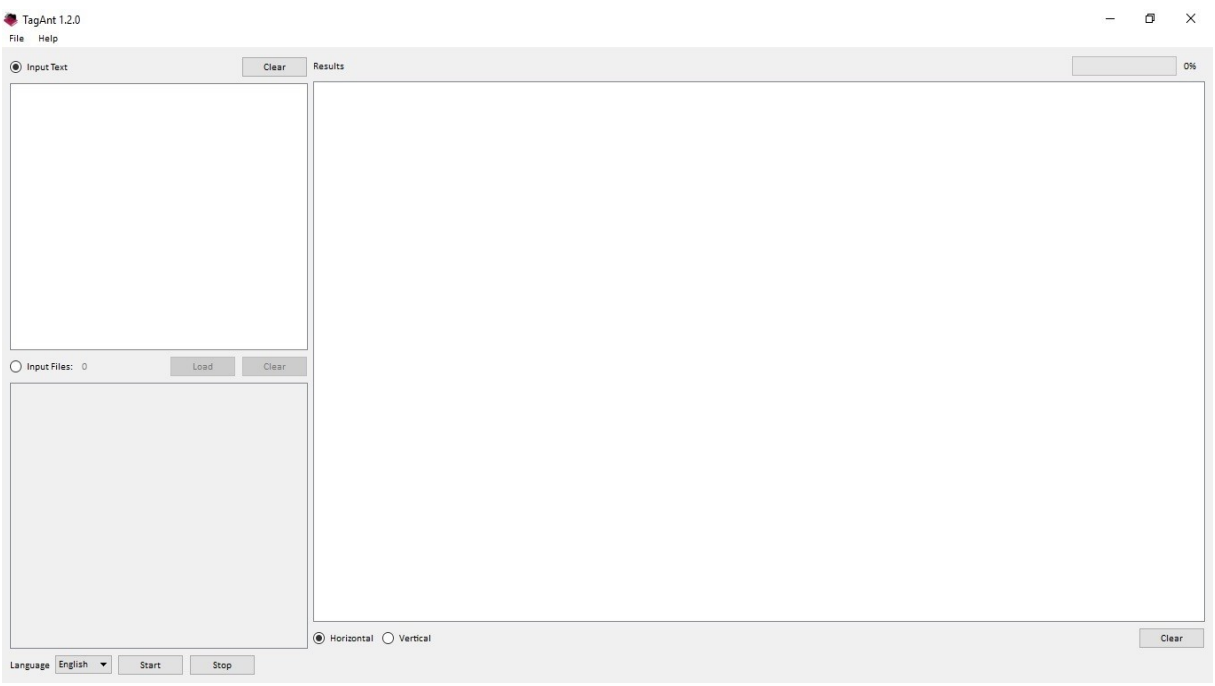

Figure 2 TagAnt clear window

Researchers can type in the text they desire to tag into the "Input text" window and execute the tagging by clicking the "Start" button in the bottom left-hand corner. In addition, the program allows choosing the *horizontal* and *vertical* tagging options.

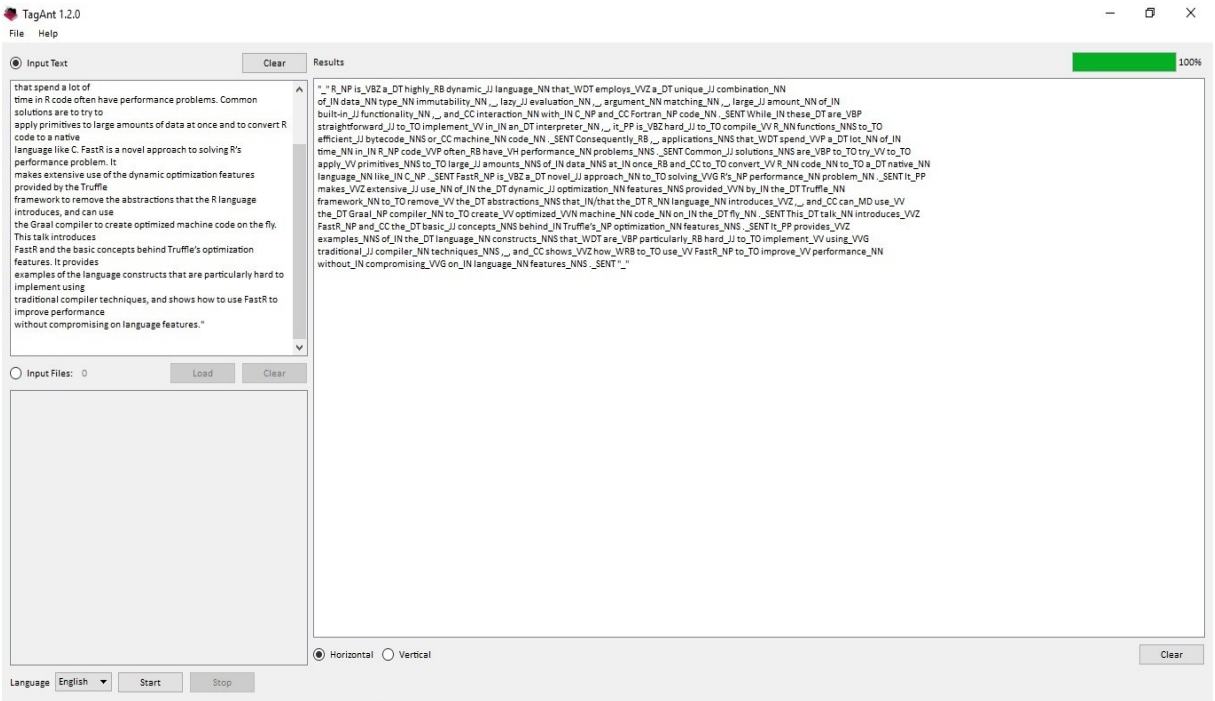

Figure 3 TagAnt Input text option & outcome

As an outcome, the program will display the tagged text in the “Results” window. It can be copied into any text processing program (see Figure 3).

3.5.2 **AntConc**

It is a freeware software multiplatform designed to simplify the analysis of corpus research. It comprises seven tools: *Concordance, Concordance plot, File view, Clusters/N-grams, Collocates, Word List, Keyword List.* Each of these serves a specific purpose: from a simple search of KWIC (Key words in context), collocates or clusters, to individual files check and access to word list made on the basis of data input to the program.

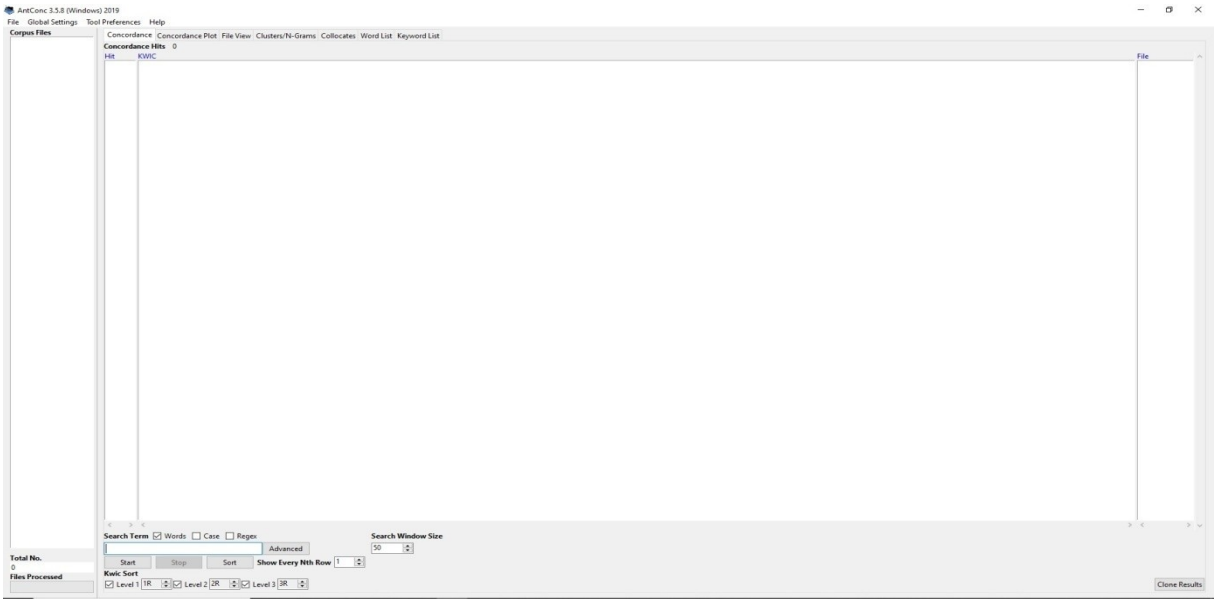

Figure 4 AntConc startup window

One can load the files in .txt (UTF-8 coded) file using *File>Open File(s)* option. All loaded files are visible in the *Corpus Files* window. The program allows adjusting the *Search window* size and highlighting the specific words in the desired left/right span from the searched term. The software allows searching for *Words, Case,* and *Regex.*

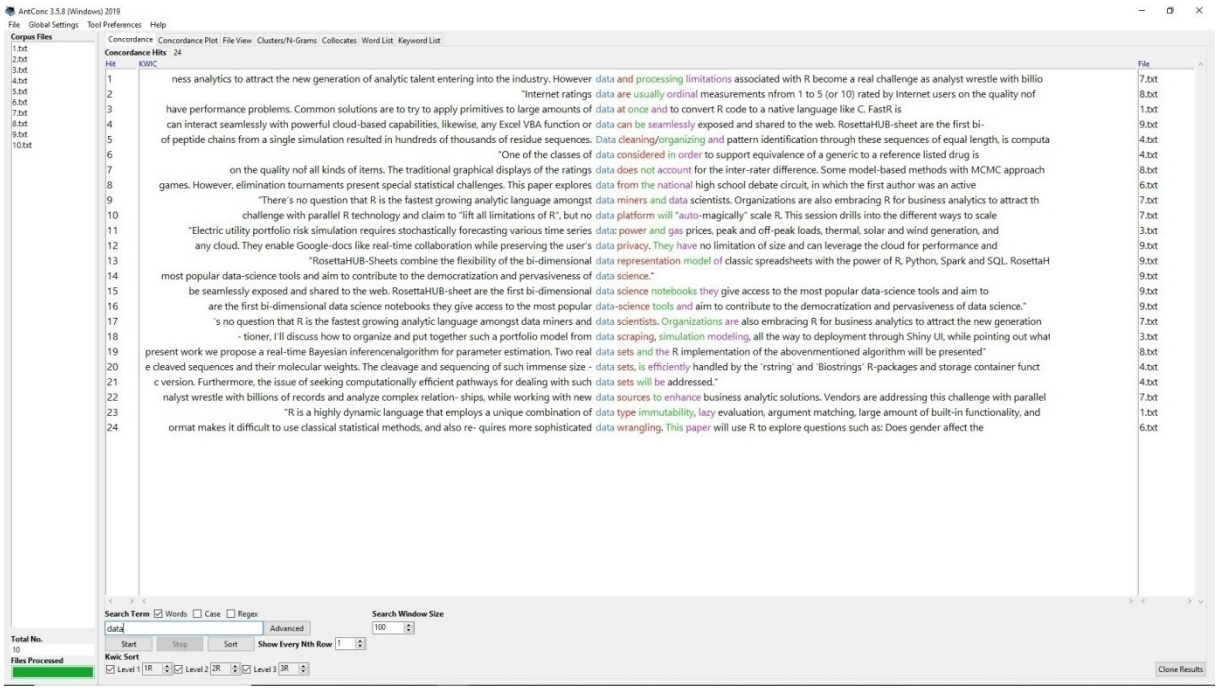

Figure 5 AntConc Concordance Tool

Figure 5 presents the outcome of the *Concordance Tool* search. The highlight colors can be adjusted in the *Global settings>Color.* Advanced search with the use of *wildcards* (might be adjusted in *Global settings>Wildcards*). By default, the search may be conducted with the following signs in front, between, or after the searched tokens:

Table 9 AntConc *wildcards* (Based on Anthony (2018), AntConc software)

Wildcard Sign	Definition
*	Zero or more characters
+	Zero or one character
?	Any one character
@	Zero or one word
#	Any one word
|	Search term 'OR' search term
&	Nonword

3.5.3 ProtAnt

It is a freeware text detection tool, and it can compare individual corpus to other reference corpus or a word list. ProtAnt counts characteristic features of target files and ranks the features on the basis of their prototypicality.

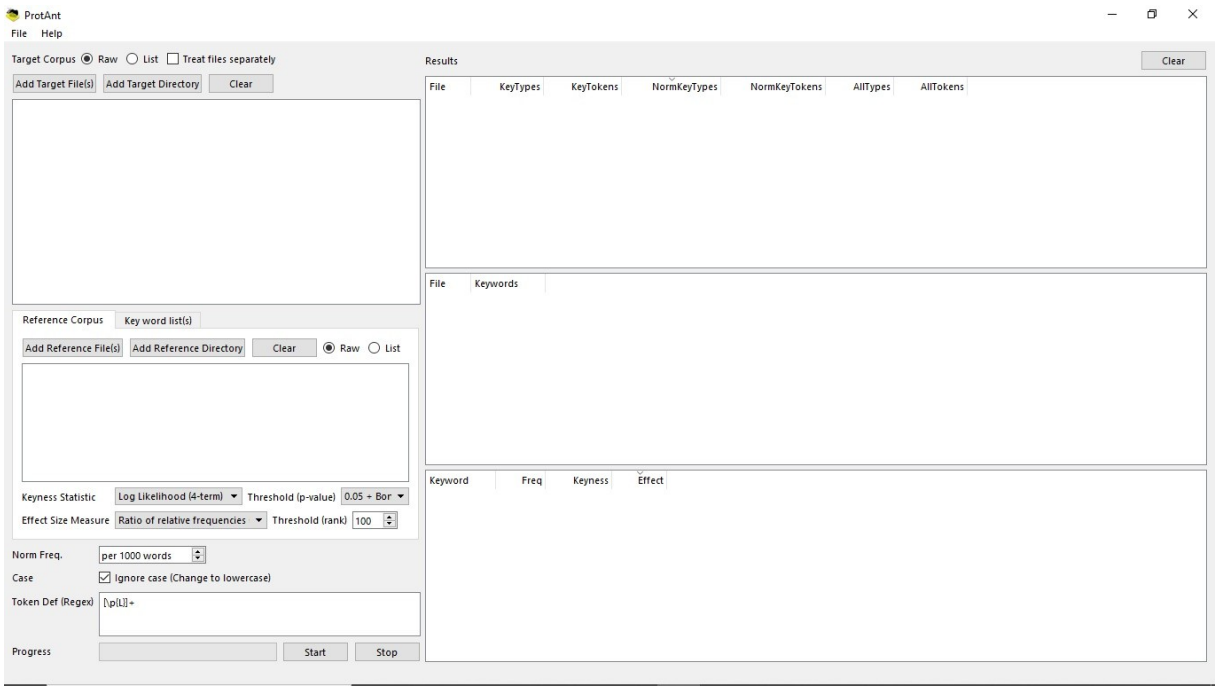

Figure 6 ProtAnt startup window

The program can load UTS-8 coded .txt files into *Target Corpus* and *Reference Corpus* or *Key word list(s),* dependent on the search type which is desired. To do that, one may use the *File* option in the top left-hand side corner or use/click the preset button(s) in the appropriate window sections. The Target Files can be treated separately or as a corpus. In addition, ProtAnt allows reading the loaded files in *Raw* format or *List* setting.

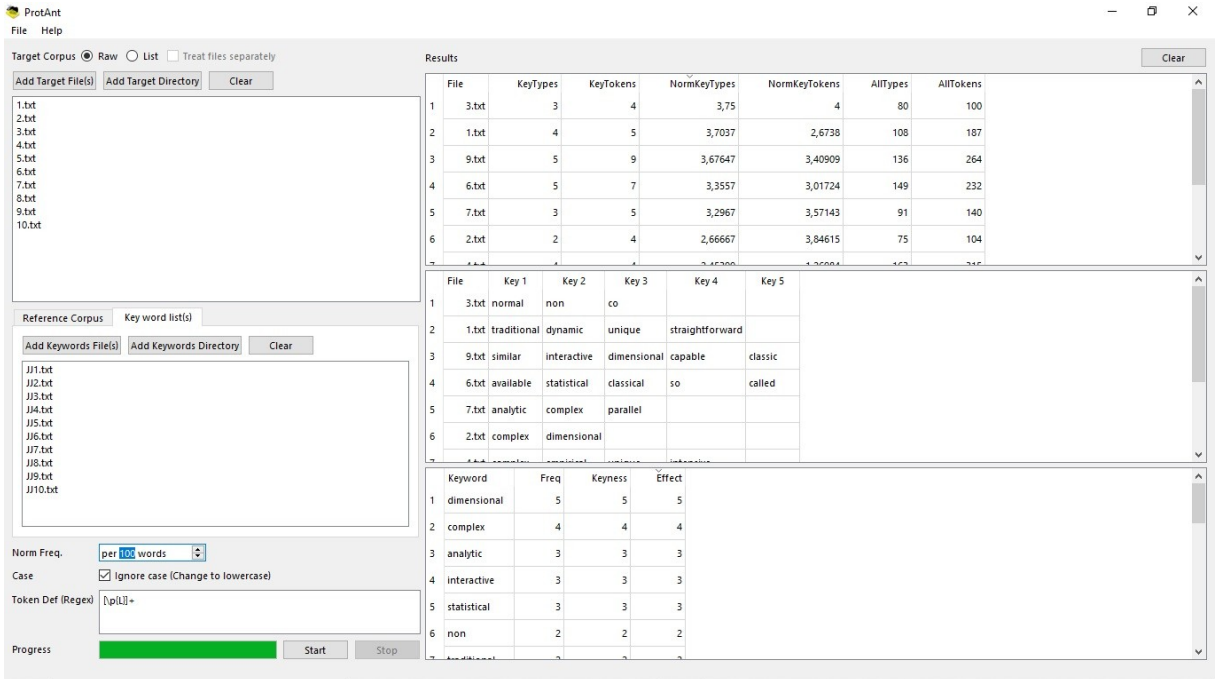

Figure 7 ProtAnt keyword list search

Normalized frequency might be adjusted to the desired level (e.g., per 1000 words) in the

bottom left-hand side corner. Finally, the *Token Definition* can be set and the *Case might* be ignored. *The start* button commences the corpus text analysis, displayed on the right-hand side (Results section).

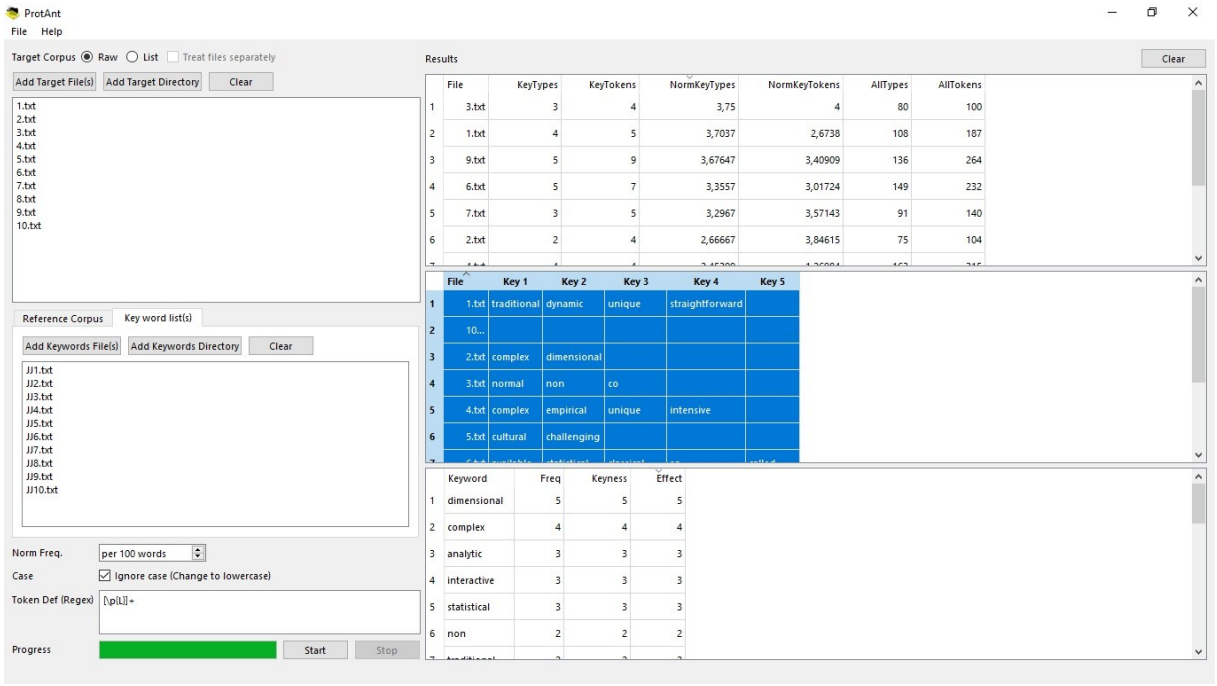

Figure 8 ProtAnt keyword list results in manipulation

The results of presented data can be adjusted in the three windows, on the right-hand side. The first one according to *File, KeyTypes, KeyTokens, NormKeyTypes, NormKeyTokens, AllTypes, AllTokens* The *s*pecified *Key number* can be selected in the second window. Moreover, *Keyword, Frequency, Keyness,* and *Effect* are displayed in the third. The data can be copied into any text editor, and it is possible to use Microsoft Excel to operate on the copied data in the table format at an instant.

3.5.4 R Gui environment

It is a freeware software for data manipulation, calculations, and graphical display. The basic features of the R program are:

1) "An effective data handling and storage facility,
2) A suite of operators for calculations on arrays, in particular matrices,
3) A large, coherent, integrated collection of intermediate tools for data analysis,
4) Graphical facilities for data analysis and display either directly at the computer or on hard-copy, and
5) A well-developed, simple, and effective programming language (called 'S'), which includes conditionals, loops, user-defined recursive functions, and input and output

facilities.” (Venables and Smith, 2015, p.2).

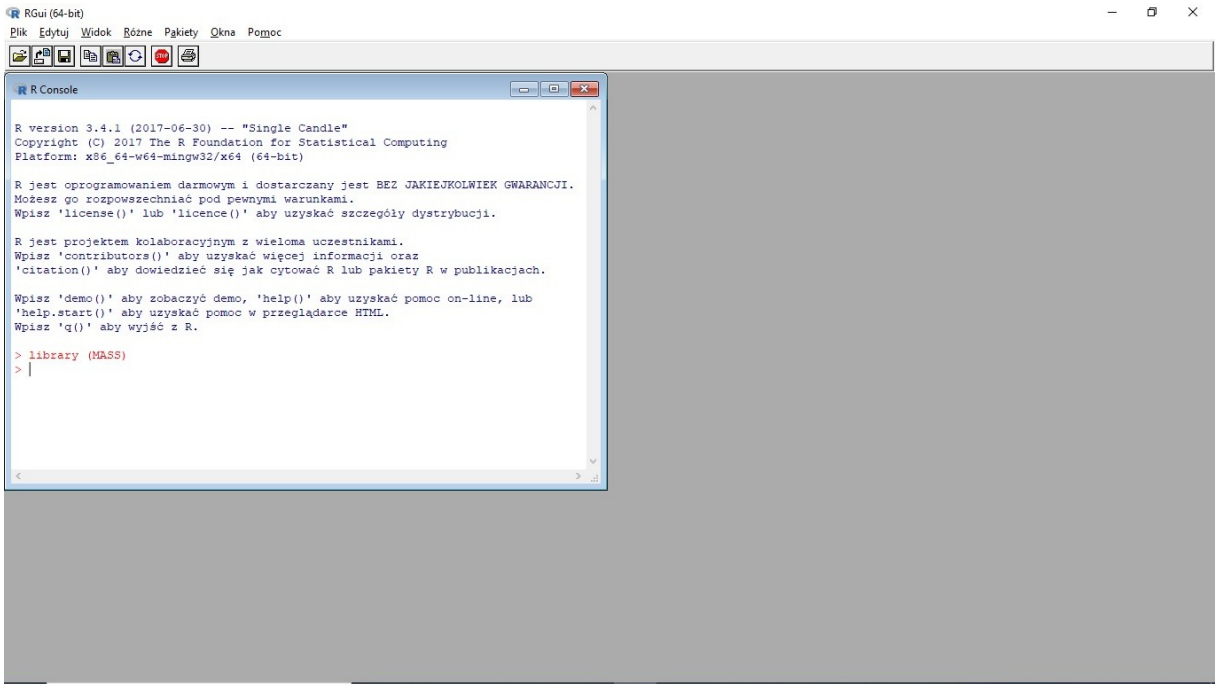

Figure 9 R Gui startup window

The program is operated with the use of text commands entered in the *R Console.* Before commencing analysis, the first and most important thing to do is to load the library data (e.g., MASS library). Most of the work done in R is the manipulation of data frames. Program introduction restricts the use of data frames specified below:

1) “The components must be vectors (numeric, character, or logical), factors, numeric matrices, lists or other data frames.
2) Matrices, lists and data frames provide as many variables to the new data frame as they have columns, elements, or variables, respectively.
3) Numeric vectors, logical, and factors are included as is, and by default, character vectors are coerced to be factors, whose levels are the unique values appearing in the vector.
4) Vector structures appearing as variables of the data frame must all have the same length, and matrix structures must all have the same row size”. (Venables and Smith, 2015, p.27)

The easiest way to use data frames is to load an external .txt file using the *read.table()* command. A list of simple R commands is presented in the table below.

Table 10 Simple R commands (Based on Dunat, 2015)

Action	Command in R
# Load data (frame) into R	data_name <- read.table(file.choose(), header=T, sep="\t")
# Look at the summary of data	summary(data_name)
# Look at detail of one column	table(data_name$NameOfColumn)
# Steps to do a Correlation Analysis	
# Do the correlation analysis	cor(data_name)
# Plot the correlation analysis graph (all data)	plot(data_name)
# Plot with main and axis titles	x <- data_name$NameOfColumn y <- data_name$NameOfColumn
# Change point shape (e.g. pch = 19) and remove frame.	plot(x, y, main = "MainTitleOfTheGraph", xlab = " NameOfXAxis", ylab = " NameOfYAxis", pch = 19, frame = FALSE)
# Add regression line	plot(x, y, main = " MainTitleOfTheGraph", xlab = " NameOfXAxis", ylab = " NameOfYAxis", pch = 19, frame = FALSE) abline(lm(y ~ x, data = mtcars), col = "blue")
# Two steps to do a Chi-square test for significance	
# Load an xtab form a text file	name <- read.table(file.choose(), header=T, row.names= 1)
# Do the Chi-Square	CHI = chisq.test(name) CHI
# Do the Chi-Square without automatic correction	chisq.test(name, correct = F)
# Three steps to obtain Observed vs. Expected outcomes	
# Count Observed and Expected	O = CHI$observed E = CHI$expected
# Execute and print	O E
# Standardized test on Observed vs. Expected	(O-E)^2/E
# Two steps to obtain Pearson's residuals	
# Count Pearson's residuals	RES= CHI$res
# Execute and print	RES
# Three steps to do Correspondence Analysis from a pivot table	
# Load data into R	name <- read.table(file.choose(), header=T, row.names= 1)
# Do the analysis	MASS_analysis <- corresp (name, nf= 2)
# Plot the results	plot (MASS_analysis)

A data frame loaded into R can be displayed after typing in the executable command; in this case, *df*. Although, one can use any other name (see Figure 10).

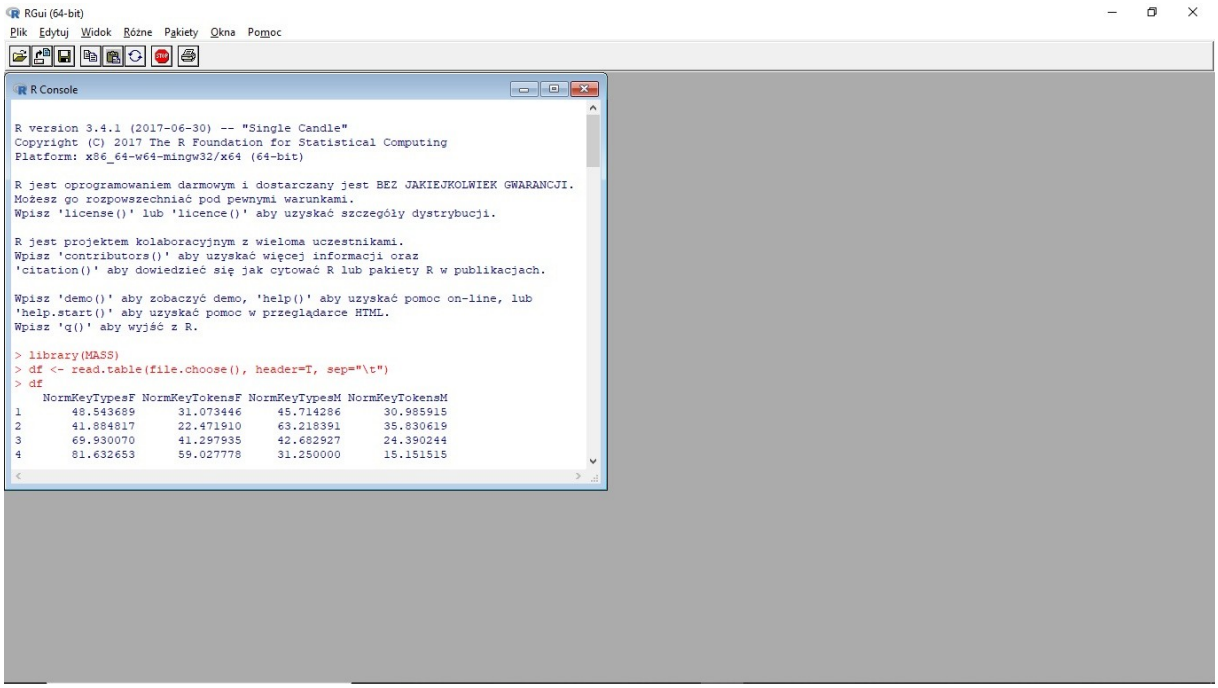

Figure 10 R Gui –Data load

The *summary (data_name)* function displays *minimum, maximum, median,* and *mean* values for all columns in the input file (see Figure 11).

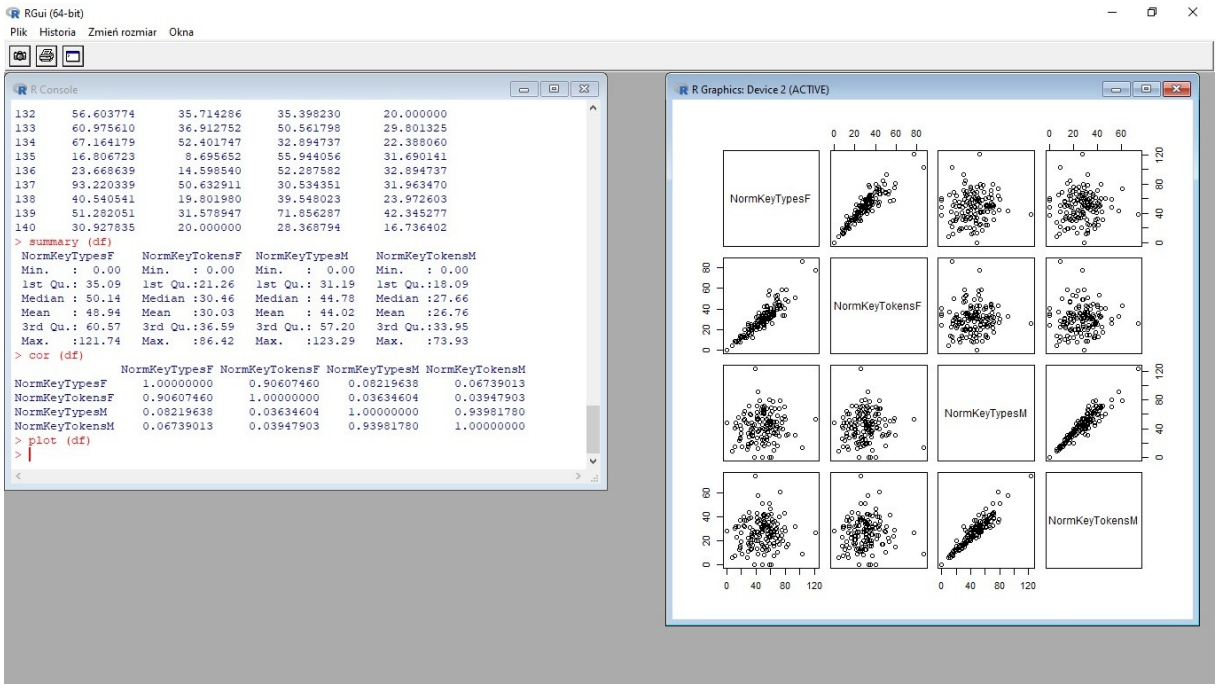

Figure 11 R Gui – Correspondence analysis execution

Correlation analysis can be obtained by loading the data into R and executing the *cor(data_name)* command. To visualize the outcome in a graph form, there is a need to type

in: *plot(data_name)*. The program will then display a second window *R Graphics Device.*

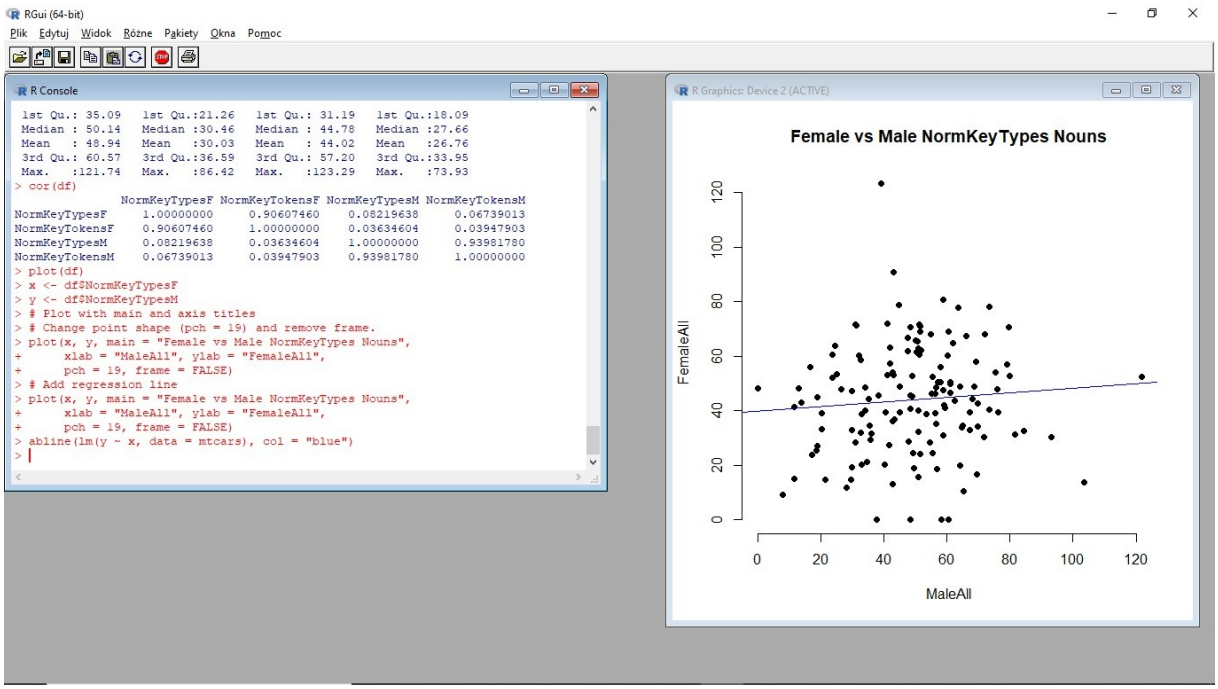

Figure 12 R Gui – Plot with main axis titles and the regression line

A *Plot with main and axis titles* function to view one specific rectangle (see Figure 12) or the all of the correlations graph (as in Figure 11) might be used. *Change point shape and remove frame* and *Add regression line* functions may construct a reasonably more precise graph (see Figure 17 above). Right-click on the *Graphics device* allows saving the graph file in .bmp or meta file. Examples of such correspondence analysis can be seen in Dunat (2021).

Chapter 3 presented the overall information about the data of the prepared corpus, pointed out the hypotheses, as well as, explained the methods of division and analysis of the data used in this dissertation. Although the methods used here might seem simple, they can be a very time-consuming process. The next part of this dissertation looks at the research data from the perspective of computational linguistics. It consists of nine sections which explore how the corpus data is divided quantitatively, according to parts of speech, academic fields, and academic word list vocabulary. It presents the software background to functions analysis and investigates how academic language functions are quantitatively distributed in the researched abstracts, in connection with academic fields, gender and speaker origin in the following three sections: vocabulary of the research process functions, vocabulary of evaluation functions, and vocabulary of analysis functions.

Chapter Four
Corpus Data Analysis

The analysis aims to detect possible interdisciplinary variations in the occurrence and usage of academic language functions, based on Martin (1976), which includes the Vocabulary of Research Process (VORP), Vocabulary of Analysis (VOA), and Vocabulary of Evaluation (VOE). For this purpose, 480 abstracts were selected. VORP consists of seven functions: formulating, investigating, presenting the methodology, describing the experiment, simulating the situation/developing a model, drawing conclusions, and reporting results. VOA comprises two functions: organizing the sequence of presenting information and reinforcing the vocabulary of the research process. Finally, VOE includes three functions: reviewing, evaluating, and criticizing (see Table 7).

All the labels mentioned above are used not only to segregate the academic vocabulary in the analysis sections 4.1 to 4.6, but also to categorize academic language functions presented in 4.8 and 4.9. Academic Vocabulary analysis is done on the basis of *New Academic Wordlists* by Coxhead (2000). The corpus data is selected from five books of abstracts chosen for this research. Furthermore, the data is divided into three academic fields, as shown in Table 11. The total of Tokens and Types in the whole corpus consisting of the whole corpus amounts to 294,780 and 91,097, respectively.

Table 11 Corpus division into academic fields

Academic field	%	Abstracts number
Computational Statistics	**20.83**	**100**
Humanities	**37.5**	**180**
Biological and Health Sciences	**41.66**	**200**
Total	**100**	**480**

Table 11 presents the sub-corpus division in terms of the number of abstracts. The biggest field, Biological and Health Sciences, comprises 200 abstracts (41.66%). The humanities sub-corpus constitutes 37.5% and has 180 texts. Lastly, the Computational Statistics field is represented by 100 samples of abstracts (20.83%).

Table 12 Academic division by gender

Academic field	Group	%	Female	%	Male	%	Total
Computational Statistics	40	8.33	20	4.16	40	8.33	**100**
Humanities	60	12.5	60	12.5	60	12.5	**180**
Biological and Health Sciences	100	20.83	60	12.5	40	8.33	**200**
Total	**200**	41.66	**140**	29.16	**140**	29.16	**480**

Table 12 shows the distribution of abstracts by gender. Thus, the sub-corpus of Computational Statistics has 20 texts of female authorship (4.16%),and 40 male texts (8.33%). The Humanities features an equal representation with 60 texts from both genders (12.5% each). The Biological and Health sciences sub-corpus consists of 60 female-authored and 40 male-authored texts, respectively 12.5% and 8.33% of the total number of abstracts.

Table 13 Academic division by speaker origin

Academic field	Native	%	Non-native	%	Total
Computational Statistics	50	10.41	50	10.41	**100**
Humanities	80	16.66	100	20.83	**180**
Biological and Health Sciences	110	22.91	90	18.75	**200**
Total	**240**	50	**240**	50	**480**

Table 13 illustrates the speaker's origin in the sub-corpora. The Computation Statistics (CS) field consists of 50 texts (10.41%) equal in terms of quantity for both native and non-native writers (see Section 3.5). There are 80 native (16.66%) and 100 non-native writers (20.83%) of abstracts in the Humanities (H) field. The last field, Biology and Health (BH), is represented by 110 native (22.91%) and 90 non-native (18.75%) texts.

Table 14 Gender division by speaker origin

Gender	Native	%	Non-native	%	Total
Group	100	20.83	100	20.83	**200**
Female	70	14.58	70	14.58	**140**
Male	70	14.58	70	14.58	**140**
Total	**240**	50	**240**	50	**480**

Table 14 presents a combined parameter of speaker origin and gender in the corpus. There are 70 female abstracts for native and non-native groups, respectively (14.58%). The same situation occurs in the male sub-corpus with 70 examples (14.58%) for native and non-native writers. A hundred abstracts (20.83%), coded as multi-authored (Group), were selected for both native and non-native writers in the corpus. The total number of native and non-native texts is 240, which constitutes 50% of the entire corpus for each category.

Table 15 Gender division through speaker type

Gender	Non-scholar	%	Scholar	%	Total
Group	32	6.66	168	35	**200**
Female	23	4.79	117	24.37	**140**
Male	32	6.66	108	22.5	**140**
Total	**87**	18.12	**392**	81.87	**480**

Table 15 shows speaker type division in gender sub-corpora. The criteria for labeling

scholar/non-scholar categories were presented in Section 3.5. The category was not used in the analysis due to the vast quantitative differences, as previously mentioned in Section 3.5. There are 87 texts (18.12%) in total, written by non-scholar writers: 32 for multi-authored (group) and male (M texts) writers (6.66%), and 23 (4.79%) for females (F texts). The total of 392 texts (81.87%) for scholar writers: 168 (35%) for a group, 117 (24.37%) for females, and 108 (22.5%) for males.

Table 16 Corpus divisions

Academic Field	Group	%	Female	%	Male	%
Computational statistics	**40**	**8.33**	**20**	**4.16**	**40**	**8.33**
Native	20	4.16	10	2.08	20	4.16
non-native	20	4.16	10	2.08	20	4.16
Humanities	**60**	**12.5**	**60**	**12.5**	**60**	**12.5**
Native	20	4.16	30	6.25	30	6.25
non-native	40	8.33	30	6.25	30	6.25
Biology and Health Sciences	**100**	**20.83**	**60**	**12.5**	**40**	**8.33**
Native	60	12.5	30	6.25	20	4.16
non-native	40	8.33	30	6.25	20	4.16
Total	**200**	41.66	**140**	29.16	**140**	**29.16**

The corpus has been normalized, as much as it was possible to level the number of the texts written in accordance with the authors' origin and gender. Table 16 presents the divisions of field, nativity, and gender. Thus, Computational statistics has 20 examples of F texts (4.16%) and 40 examples of M texts (8.33%), both equally divided into native and non-native writers. Accordingly, Biology and Health Sciences have 30 (6.25%) and 20 (4.16%) examples of F and M texts, equally in each female and male group. The Humanities field has equally 30 (6.25%) samples of abstracts for each nativity and gender division. Thus, there is a total of 140 samples of abstracts of both native and non-native authors of both genders.

Table 17 Average length of the researched abstracts

Academic field	Data type	Mean value for Native		Mean value for Non-native		Total	
Computational Statistics	Types	135.44		139.58		13751	
	Tokens	369.46		399		38423	
Humanities	Types	188.91		238.40		38953	
	Tokens	640.27		914.25		142647	
Biology and Health	Types	179.62		162.12		34350	
	Tokens	599.52		530.68		113710	
		F	**M**	**F**	**M**	**F**	**M**
Computational Statistics	Types	136.10	127.80	127.50	130.95	2636	5175
	Tokens	379.90	354.45	352.70	371.70	7326	14523
Humanities	Types	182.50	199.90	214.56	217.36	11912	12518
	Tokens	614.90	680.03	835.76	778.16	43520	43746

Biology and Health	Types	196.03	191	157.10	146.45	10594	6749
	Tokens	688.06	670.80	522.93	448.95	36330	22395
		Mean value for all Females		**Mean value for all Males**		**Total**	
Computational Statistics	Types	131.80		168.25		7811	
	Tokens	366.30		559.87		21849	
Humanities	Types	198.53		129.37		24430	
	Tokens	725.33		363.07		87266	
Biology and Health	Types	176.56		208.63		17343	
	Tokens	605.50		729.10		58725	

Table 17 presents the average length of abstracts in the corpus. It can be observed that Non-native authors write longer abstracts in the fields of CS and H fields; the mean token values are 399 words and 914.25 words, respectively. In the field of BH one can see the opposite trend with native authors writing longer abstracts, namely, 599.52 words, which is 68.84 more than the value for the non-native counterpart. What is interesting, the trend seems not to be the same if one looks at the gender division in comparison to speaker origin in the CS field. Thus, Native F write longer texts than non-native F, there is a difference of 27.20 words. Furthermore, on average, F authors write more extended abstracts in the H field, whilst M authors write longer abstracts in the CS and BH fields, which seems not to correspond to speaker origin division. As one can observe, native M authors write longer abstracts in the H field, namely the average difference amounts to 65.13 words. Additionally, non-native F authors in the CS and BH fields, as well as native F in the BH field write longer abstracts. Although the differences are not big in the first two cases, as they amount to 25.45 and 17.26 words, they seem to be substantial in the BH field, where the difference between speaker origin and gender amounts to 73.89 words.

4.1 Nouns

The number of all nouns used in the corpus is presented in Figures 13 and 14. The mean value for the whole corpus amounts to 106.79 nouns. The mean frequency of occurrence of nouns used in M and F abstracts amounts to 110.90 nouns and 98.69 nouns, respectively. For group abstracts the mean frequency of occurrence of nouns amounts to 109.59. In the following figures, the vertical axis corresponds to the number of texts, while the horizontal axis corresponds to the number of specified parts of speech.

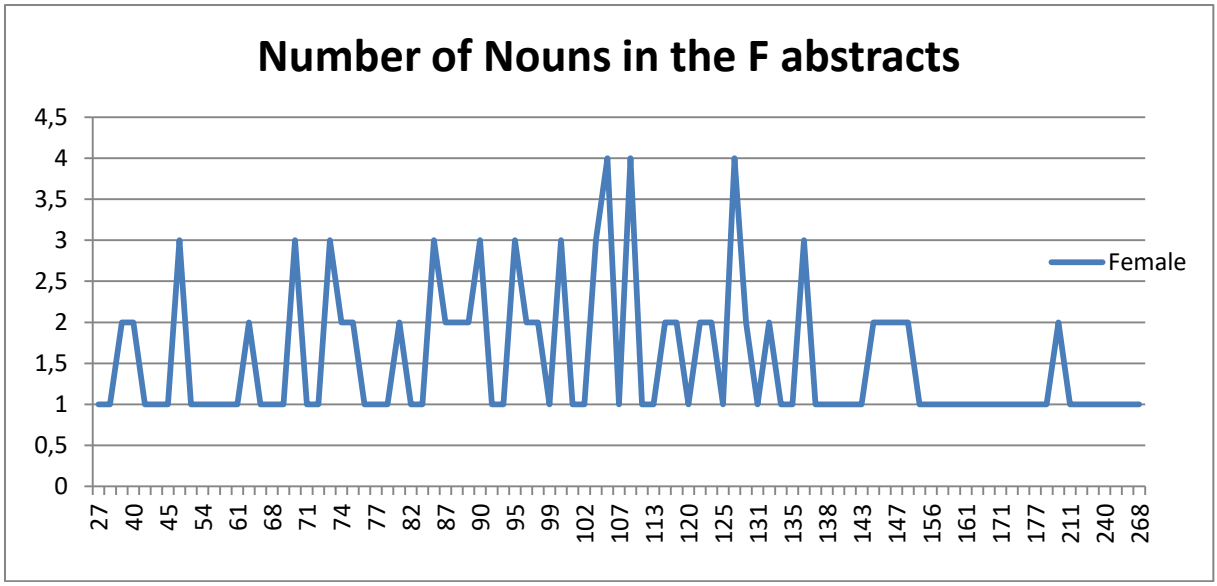

Figure 13 Number of nouns in the F abstracts

Figure 13 demonstrates that women use between 27 and 268 nouns per text. There are four texts which use the following number of nouns: 105, 109, and 126, in each of the three peaks. There are eight peaks on the graph, showing the number of nouns used: 46, 70, 73, 86, 90, 95, 100, and 104, each peak represents three abstracts. Moreover, there are twenty-one instances in which two texts have used the following numbers of nouns: 38, 40, 63, 74, 75, 81, 87-89, 97-98, 118-119, 123-124, 130, 132, 144, 145, 147-148, and 210.

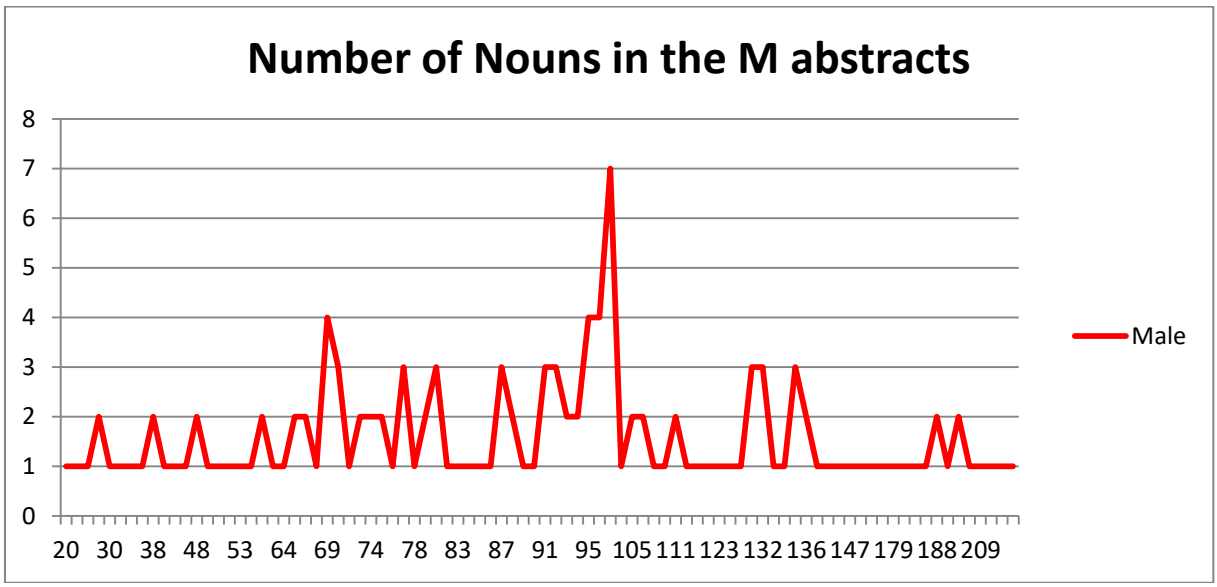

Figure 14 Number of nouns in the M abstracts

Figure 14 shows the number of nouns used by men throughout the corpus. Men use between 20 and 230 nouns. The highest peak occurs in seven texts in which 97 nouns can be found. There are three peaks displayed on the graph: 69, 95, and 96 nouns. Each peak represents four abstracts, that use the presented number of nouns. Three texts used the following number of

nouns: 70, 77, 80, 87, 91, 92, 130, 132, and 135. Two texts using the same number of nouns can be found on the graph, for each of the following number of nouns: 28, 38, 48, 59 66, 67, 73-75, 79, 88, 93-94, 105, 107, 111, 136, 188, and 195 nouns marked.

4.2 Verbs

Figures in this section present the number of verbs used in the corpus. The mean value for verbs in the whole corpus totals 32.19 verbs. The mean value for F abstract amounts to 32.53, while for M abstract it is 30.56. The mean value for group (multi-authored) abstract amounts to 33.09 verbs.

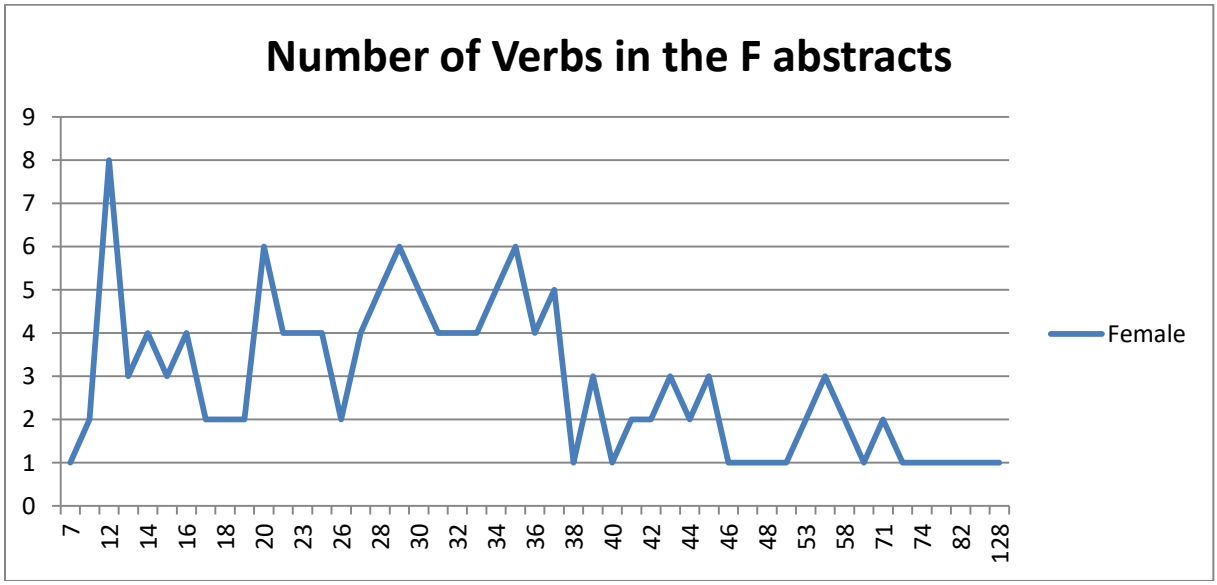

Figure 15 Number of verbs in the F abstracts

Figure 15 shows that women use between 7 and 83 verbs in their texts. There is one abstract in which 128 verbs were used. The highest peak in the number of texts with an equal number of verbs can be observed for 12 verbs, present in eight F texts. Further, six texts with an equal number of examples can be observed as three peaks in the graph. Each peak is formed by six abstracts and the number of verbs amounts to 20, 29, and 35. In five texts verbs are distributed as follows: 28, 34, and 37. The following numbers of verbs: 14, 16, 22, 23, 25, 27, 31-33, 39, 43, 45, and 54, each representing a peak on the graph, can be found in four texts. There are also 11 instances of two texts with the same number of verbs used, i.e., 8, 17-19, 26, 41-42, 44, 53, 58, and 71.

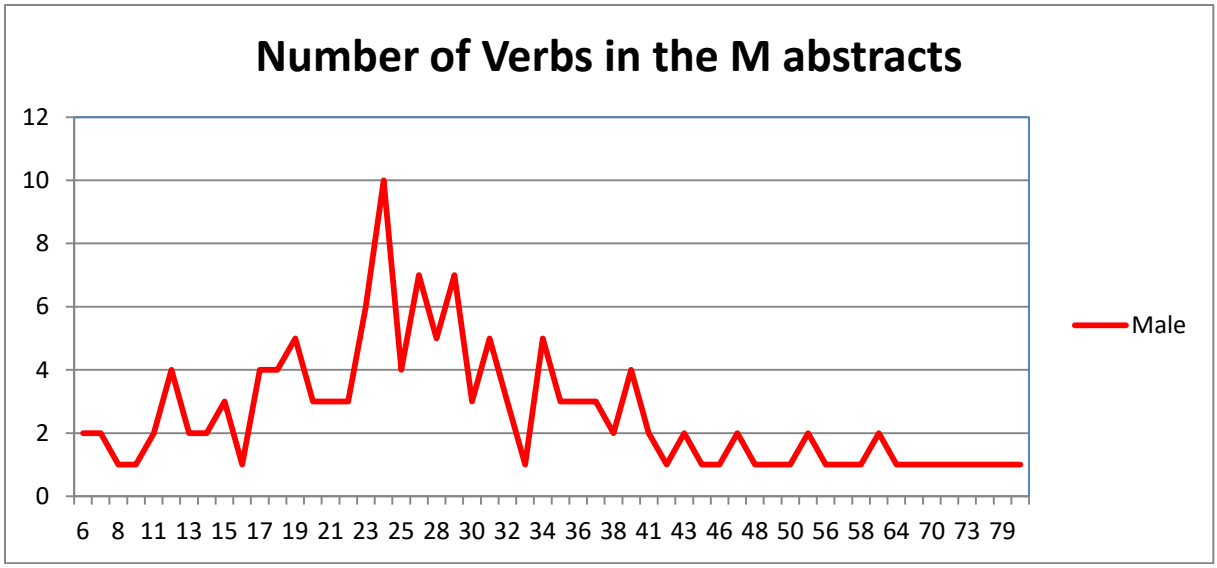

Figure 16 Number of verbs in the M abstracts

Figure 16 demonstrates the number of verbs used by men in their abstracts. The use varies between 6 to 79 verbs per text. Ten texts use 24 verbs, which constitutes the highest number of texts in one group. Seven texts use 27 and 29 verbs, while six texts use 23 ones. Five texts peak three times and use 19, 31, and 34 verbs. Furthermore, three texts use the following number of verbs: 15, 20-22, 30, and 35-37. Finally, two texts use the following numbers of verbs: 6-7, 11,13-14, 38, 41, 43, 47, 52 and 60.

4.3 Adjectives

The following figures present the number of adjectives used by both sexes throughout the corpus. The mean value for adjectives in the whole corpus amounts to 31.23 adjectives; for F abstracts it is 30.12 adjectives. The mean value for M abstracts amounts to 30.25 and the mean value for group abstracts equals 32.69.

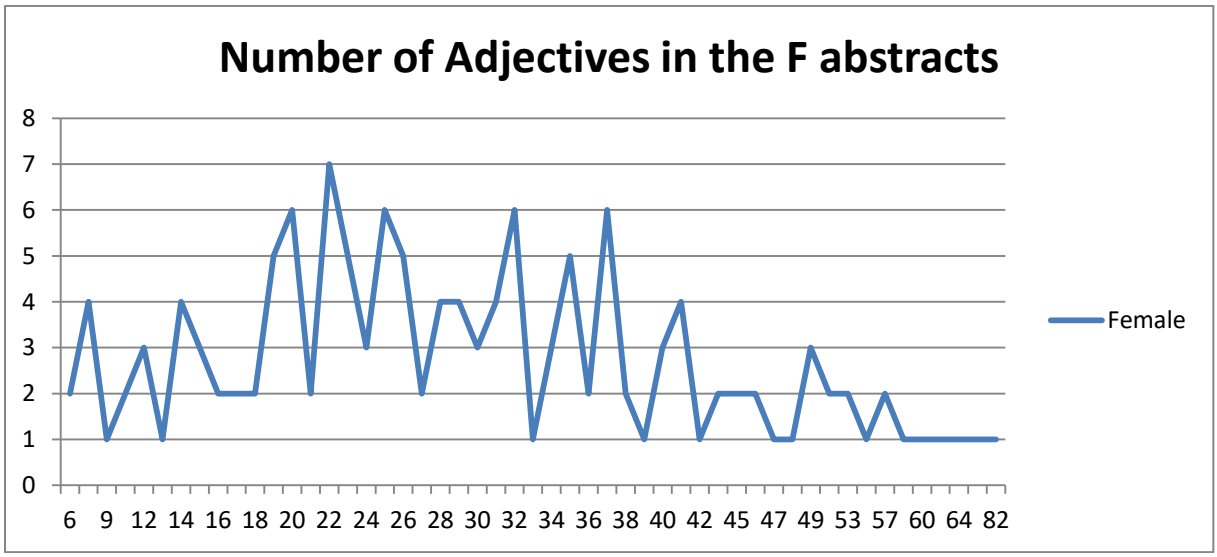

Figure 17 Number of adjectives in the F abstracts

Figure 17 illustrates the number of adjectives used by women in abstracts in the corpus. The use varies between 6 and 67, and there is one text with as many as 82 adjectives used in it. Seven texts use 22 adjectives. Six texts use 20, 25, 32, and 37 adjectives. Moreover, five texts use four different quantities of adjectives: 19, 23, 26, and 35. These two groups (of five and six texts in the graph) constitute the biggest set in the corpus, as can be observed in Figure 17 above (19 to 37 adjectives). The following numbers of adjectives: 8, 14, 28, 29, 31, and 41 were found to be used in four texts each. Finally, three texts were found for each of the following numbers of adjectives: 12, 15, 24, 30, 34, 40, and 49.

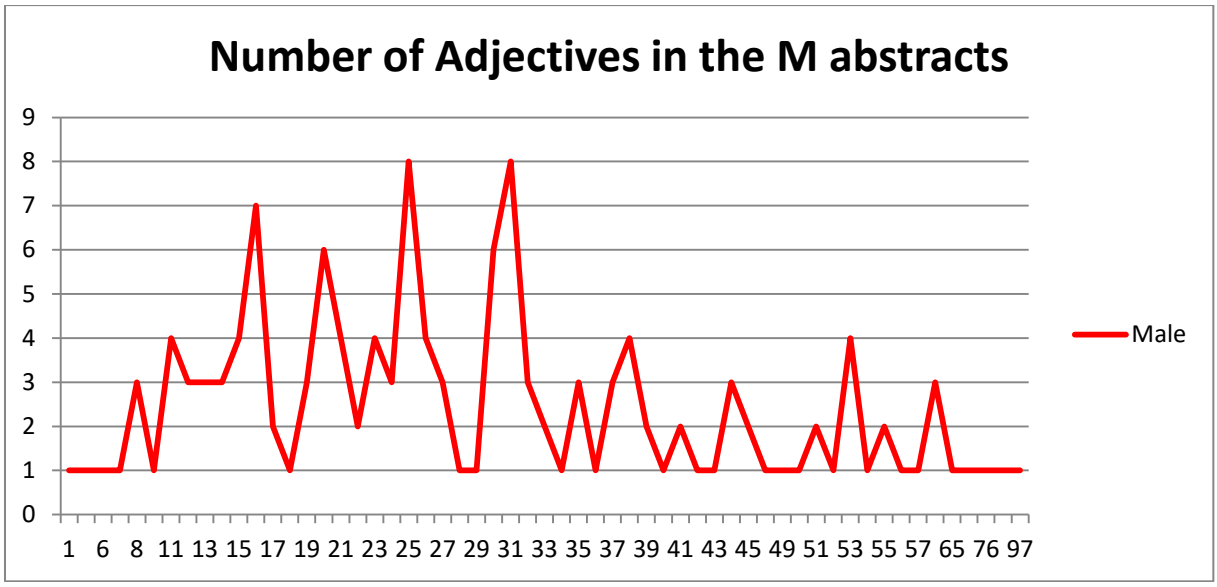

Figure 18 Number of adjectives in the M abstracts

Figure 18 presents adjectives used by men in the researched abstracts. There are two groups, each consisting of eight texts, where the number of adjectives has been found to be 25 and 31. Next, seven texts have 16 adjectives each. For a group of six texts, the frequency of adjective usage is 20 and 30. What is interesting is the fact that there are no examples that form a peak for the five similar texts, as observed in Figure 18. Four texts were found with the following numbers of adjectives: 11, 15, 21, 23, 26, 38, and 53. Next, three texts were found to have the sets of 8, 12, 13, 14, 19, 24, 27, 35, 37, 44, and 58 adjectives. Two texts were found for each of the following numbers of adjectives: 17, 22, 33, 41, 45, 51, and 55. The following numbers of adjectives do not occur in any of the texts: 2-4, 9, 47, 48, 59-64, 66, 69-75, 77-84, and 86-96. The highest number of adjectives used in a M abstract is 97.

4.4 Adverbs

Figures 19 and 20 visualize the use of all adverbs by both gender groups in the corpus. The

mean value for adverbs in the whole corpus amounts to 8.78. The mean value for F abstracts totals 8.57, while for M abstracts, it is 8.97. The mean value for adverbs for group abstracts amounts to 8.80.

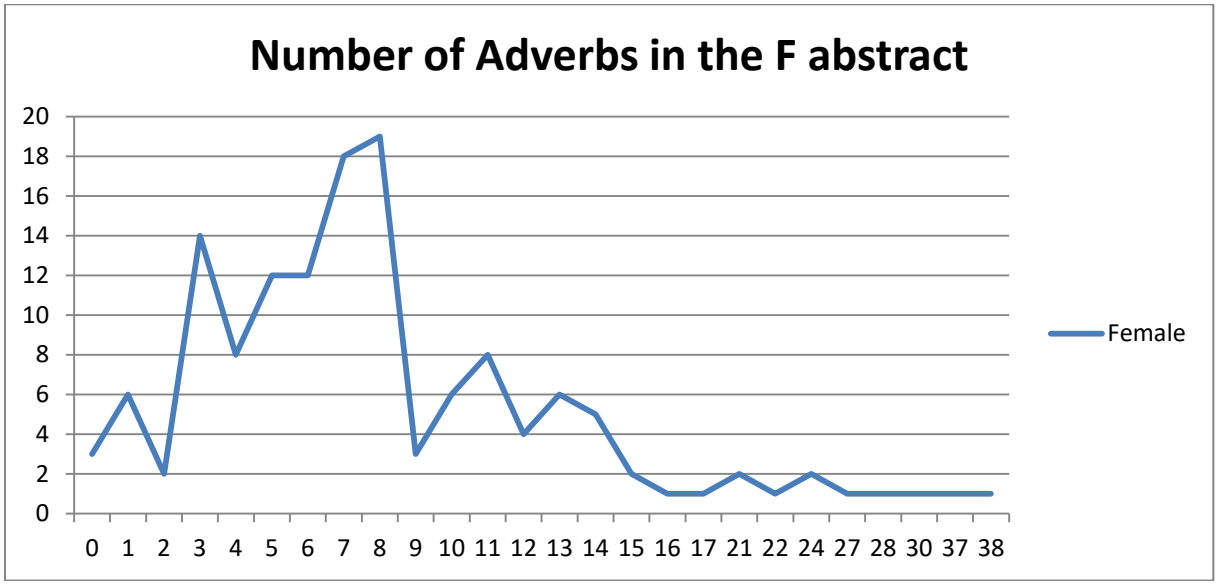

Figure 19 Number of adverbs in the F abstracts

Figure 19 shows the number of adverbs in F abstracts. In 19 abstracts, eight adverbs have been observed. Next, 18 texts use seven adverbs, and 14 texts use three adverbs. In addition, five and six adverbs are present in 12 texts each. Moreover, eight texts use 4 and 11 adverbs. Six texts use 1, 10, and 13 adverbs. Four texts use 12 adverbs. Three texts use none, and nine, adverbs. Finally, a pair of texts was identified corresponding to each of these numbers: 2, 15, 21, and 24 adverbs.

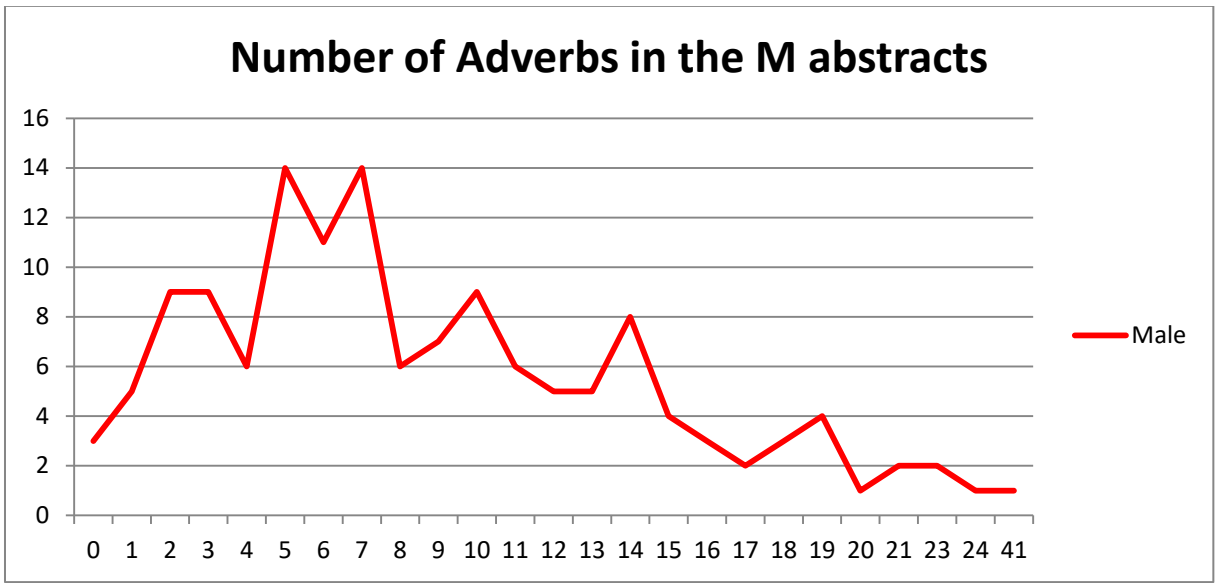

Figure 20 Number of adverbs in the M abstracts

Figure 18 shows the number of adverbs in M abstracts in the researched corpus. There are 14 texts in which men use 5 and 7 adverbs. Next, in eleven texts, six adverbs occur. There are three peaks of nine texts, with 2, 3, and 10 adverbs. Eight texts use 14 adverbs. Furthermore, seven texts have been found to have nine adverbs. Six texts use 4, 8, and 11 adverbs. Five texts use respectively: 1, 12, and 13 adverbs. Moreover, four texts employ 15 and 19 adverbs. There are three texts which make use of no adverbs. Three texts use 16 adverbs, and another three use 18 adverbs. Finally, 17, 21, and 23 adverbs were used in two texts each.

4.5 Academic Word Lists division

The following sections present the findings in reference to Academic Word Lists. The data of the corpus was sifted through ProtAnt using AWL vocabulary list prepared beforehand. The method used, step by step, has been described in section 3.5. The presented keywords refer to actual words found in the abstracts constituting the corpus, which was dictated by the limitations of the software used. It is worth noting that verb forms are considered separate types in this research because they appeared in such forms on the AWL by Coxhead (2000).

4.5.1 Nouns

Table 18 AWL keyword nouns for gender division

No.	Keyword Noun	Freq. occurrence for Females	%	Keyword Noun	Freq. occurrence for Males	%
1	analysis	49	3.48%	analysis	48	3.50%
2	research	38	2.70%	research	42	3.06%
3	approach	28	1.99%	context	26	1.89%
4	process	28	1.99%	design	26	1.89%
5	context	24	1.70%	approach	24	1.75%
6	environment	23	1.63%	framework	22	1.60%
7	focus	22	1.56%	process	20	1.46%
8	community	21	1.49%	focus	20	1.46%
9	impact	18	1.28%	project	20	1.46%
10	intervention	18	1.28%	role	19	1.38%
11	design	17	1.21%	impact	19	1.38%
12	project	17	1.21%	assessment	16	1.17%
13	method	16	1.14%	function	16	1.17%
14	role	15	1.07%	range	16	1.17%
15	implementation	15	1.07%	environment	14	1.02%
16	evaluation	14	0.99%	participation	14	1.02%
17	framework	14	0.99%	interaction	14	1.02%
18	access	14	0.99%	access	14	1.02%
19	communication	14	0.99%	technology	13	0.95%
20	culture	12	0.85%	area	12	0.87%
	Total key types/ key tokens	**349/1408**			**321/1373**	

Table 18 displays the nouns found in abstracts related to gender. The analysis reveals that the two most frequently used nouns for both genders are *analysis* at 3.48% for females and 3.50% for males, followed by *research* at 2.70% for females and 3.06% for males. The first notable difference in noun usage arises with *approach*, which ranks third in the female sub-corpus at 1.99%, while it ranks fifth in the male sub-corpus at 1.75%. Additionally, *design* occupies the fourth position in the male sub-corpus at 1.89%, but is ranked twelfth in the female sub-corpus with only 1.21%. *Process* holds the fourth position in the female abstracts at 1.99%, whereas it appears seventh in the male abstracts at 1.46%. Lastly, *context*, the fifth noun in the female sub-corpus at 1.70%, ranks third in the male sub-corpus at 1.89%.

Table 19 AWL keyword nouns for native writer division

No.	Keyword Noun	Freq. occurrence for Native Females	%	Keyword Noun	Freq. occurrence for Native Males	%
1	analysis	26	3.64%	analysis	30	4.13%
2	research	20	2.80%	research	18	2.48%
3	community	17	2.38%	context	17	2.34%
4	approach	15	2.10%	design	15	2.06%
5	intervention	15	2.10%	role	14	1.93%
6	focus	14	1.96%	approach	13	1.79%
7	process	13	1.82%	impact	12	1.65%
8	context	11	1.54%	focus	11	1.51%
9	environment	10	1.40%	participation	11	1.51%
10	role	10	1.40%	function	10	1.38%
11	evidence	9	1.26%	process	10	1.38%
12	project	9	1.26%	interaction	10	1.38%
13	promotion	9	1.26%	project	10	1.38%
14	impact	8	1.12%	area	9	1.24%
15	identity	7	0.98%	evidence	9	1.24%
16	method	7	0.98%	framework	9	1.24%
17	evaluation	7	0.98%	access	9	1.24%
18	range	7	0.98%	intervention	9	1.24%
19	framework	7	0.98%	assessment	8	1.10%
20	goal	7	0.98%	range	8	1.10%
	Total key types/ key tokens	**256/715**			**249/727**	

Table 19 presents the nouns used in the female (NF) and male (NM) corpora within the native speakers (N) group. Both female and male authors share the same top two most frequent nouns: *analysis* and *research*. The first noun accounts for 3.64% of the NF sub-corpus and 4.13% of the NM sub-corpus, while *research* represents 2.80% for the NF writers and 2.48% for NM writers. In the NF abstracts, the nouns *community*, *approach*, and *intervention* comprise 2.38%, 2.10%, and 2.10% of the total examples, respectively. Conversely, in the

NM abstracts, *context*, *design*, and *role* account for 2.34%, 2.06%, and 1.93%. Notably, *community* does not appear among the top twenty most frequent words in the NM sub-corpus, while both *approach* and *intervention* do. *Approach* ranks sixth in the NM sub-corpus at 1.79%, and *intervention* is positioned nineteenth at 1.10%. Additionally, *design* does not appear among the top twenty words in the NF sub-corpus. However, *role* is placed fifth in the NM list at 1.93% and tenth in the NF list at 1.40%.

Table 20 AWL keyword nouns for non-native writer division

No.	Keyword Noun	Freq. occurrence for Non-native Females	%	Keyword Noun	Freq. occurrence for Non-native Males	%
1	analysis	23	3.32%	research	24	3.72%
2	research	18	2.60%	analysis	18	2.79%
3	process	15	2.16%	framework	13	2.01%
4	approach	13	1.88%	approach	11	1.70%
5	context	13	1.88%	design	11	1.70%
6	environment	13	1.88%	environment	10	1.55%
7	design	12	1.73%	process	10	1.55%
8	impact	10	1.44%	project	10	1.55%
9	implementation	10	1.44%	context	9	1.39%
10	area	9	1.30%	focus	9	1.39%
11	method	9	1.30%	assessment	8	1.24%
12	focus	8	1.15%	source	8	1.24%
13	technology	8	1.15%	range	8	1.24%
14	project	8	1.15%	culture	7	1.08%
15	assessment	7	1.01%	impact	7	1.08%
16	culture	7	1.01%	text	7	1.08%
17	evaluation	7	1.01%	technology	7	1.08%
18	framework	7	1.01%	function	6	0.93%
19	access	7	1.01%	period	6	0.93%
20	communication	7	1.01%	challenge	6	0.93%
	Total key types/ key tokens	**268/693**			**244/646**	

Table 20 shows the most frequently occurring nouns in the non-native (NN) F and M sub-corpora. First of all, *analysis* (3.32%) and *research* (2.60%) take the first and second places on the NNF list, with a slight change on the NNM list, i.e., *research* is listed first, and *analysis* second, with 3.72% and 2.79% respectively; *environment* has the same position on both lists. In the NNF sub-corpus, they represent 1,88%, and in the NNM sub-corpus. They constitute 1.70% and 1.55% of the total AWL nouns used in the corpus. *Implementation* (1.44%), *area* (1.30%), *method* (1.30%), *evaluation* (1.01%), *access* (1.01%), and *communication* (1.01.%) do not occur among the first twenty nouns on NNM list. On the other hand, *source* (1.24%), *range* (1.24%), *text* (1.08%), *function* (0.93%), *period* (0.93%),

and *challenge* (0.93%) are those that occur on the NNM wordlist but are absent in the NNF first twenty most frequent nouns wordlist.

4.5.2 Verbs

Table 21 AWL keyword verbs for gender division

No.	Keyword Verb	Freq. occurrence for Females	%	Keyword Verb	Freq. occurrence for Males	%
1	conducted	26	2.40%	assess	13	1.37%
2	found	22	2.03%	assessed	13	1.37%
3	analysed	15	1.84%	found	12	1.27%
4	assessed	15	1.38%	identify	11	1.16%
5	identified	15	1.38%	evaluate	11	1.16%
6	designed	14	1.38%	enhance	11	1.16%
7	create	13	1.29%	focused	11	1.16%
8	identify	13	1.20%	required	11	1.16%
9	focused	13	1.20%	create	10	1.06%
10	perceived	13	1.20%	participate	10	1.06%
11	investigate	12	1.20%	demonstrate	10	1.06%
12	involved	12	1.11%	designed	10	1.06%
13	promote	11	1.11%	analysed	9	0.95%
14	evaluate	10	1.01%	promote	8	0.85%
15	analyzed	10	0.92%	established	8	0.85%
16	indicated	10	0.92%	implemented	8	0.85%
17	creating	10	0.92%	affect	7	0.74%
18	achieve	9	0.92%	contribute	7	0.74%
19	created	9	0.83%	conducted	7	0.74%
20	participated	9	0.83%	derived	7	0.74%
	Total key types/ key tokens	**354/1085**			**346/946**	

Table 21 displays the first 20 verbs used throughout the corpus by both the F and M authors. Firstly, the verbs *assessed,* and *found* occur on both sides of the wordlist. The first of the verbs, *assessed,* occupies the second position on the M wordlist (13; 1.37%) and the fourth on the F wordlist (15; 1.38%). It is worth to observe the frequency of occurrence in this instance. The second of the verbs, *found*, has the third position on the males' wordlist (12; 1.27%) and the second on the F wordlist (26; 2.40%). Next, the first verb on the F list, i.e., *conducted* (2.40%), does not occur on the opposite list. Similarly, the verbs: *identified, perceived, investigate, involved, indicated, creating, achieve,* and *participated* show the same tendency. *Analysed* (1.38%) has the third position on the F wordlist, whereas it is 13th on the M wordlist (0.95%). *Assess* ranks first on the M wordlist. It does not occur on the F wordlist, nor do the following verbs: *enhance, required, participate, demonstrate, established, implemented, affect,* and *contribute*.

Table 22 AWL keyword verbs for native writer division

No.	Keyword Verb	Freq. occurrence for Native Females	%	Keyword Verb	Freq. occurrence for Native Males	%
1	conducted	16	2.83%	assessed	10	1.93%
2	assessed	11	1.94%	assess	8	1.55%
3	found	9	1.59%	required	8	1.55%
4	analysed	8	1.41%	create	6	1.16%
5	perceived	8	1.41%	identify	6	1.16%
6	assess	7	1.24%	affect	6	1.16%
7	create	7	1.24%	construct	6	1.16%
8	identify	7	1.24%	participate	6	1.16%
9	focused	7	1.24%	contribute	6	1.16%
10	participated	7	1.24%	demonstrate	6	1.16%
11	involved	7	1.24%	enhance	6	1.16%
12	promote	6	1.06%	analysed	6	1.16%
13	implementing	6	1.06%	found	6	1.16%
14	identified	6	1.06%	mediated	6	1.16%
15	designed	6	1.06%	enable	5	0.97%
16	targeted	6	1.06%	participated	5	0.97%
17	investigate	5	0.88%	investigating	5	0.97%
18	reveal	5	0.88%	derived	5	0.97%
19	indicated	5	0.88%	established	5	0.97%
20	implemented	5	0.88%	conducted	4	0.77%
	Total key types/ key tokens	**257/566**			**255/517**	

Table 22 presents the top 20 most frequent keyword verbs in the native female (NF) and native male (NM) sub-corpora. The verb *conducted* leads the NF sub-corpus with 16 occurrences, accounting for 2.83% of all verbs. It appears in the 20th position in the NM sub-corpus, constituting only 0.77%. The verb *required* is exclusive to the NM list, occupying the third position with a frequency of 1.55%. The top verb in the NM sub-corpus, *assessed*, has a frequency of 1.93%, which closely aligns with its NF counterpart at 1.94%, where it ranks third. In the NF wordlist, *found* is the third most frequent verb, appearing at 1.59%, while it ranks 13th in the NM list with a frequency of 1.16%. The verb *create* is fifth in the NM list (1.16%) and seventh in the NF list (1.24%). *Analysed*, which ranks fourth in the NF sub-corpus (1.41%), is positioned 12th in the NM list at 1.16%. The NF list includes verbs that do not appear on the NM list, filling positions 11th through 20th. In contrast, the NM list features several unique verbs, including *affect*, *construct*, *participate*, *contribute*, *demonstrate*, and *enhance* (ranked 7th to 12th), as well as *mediated* and *enable* (ranked 15th and 16th). Additionally, *investigating*, *derived*, and *established* occupy the 18th to 20th positions on the NM list, alongside the previously mentioned *required*. It is important to note that the verb listed as 17th on both wordlists differs in form, which results in it being counted as a separate

entry.

Table 23 AWL keyword verbs for non-native writer division

No.	Keyword Verb	Freq. occurrence for Non-native Females	%	Keyword Verb	Freq. occurrence for Non-native Males	%
1	found	13	2.50%	evaluate	7	1.63%
2	conducted	10	1.93%	focused	7	1.63%
3	identified	9	1.73%	achieve	6	1.40%
4	evaluate	8	1.54%	found	6	1.40%
5	designed	8	1.54%	designed	6	1.40%
6	investigate	7	1.35%	assess	5	1.17%
7	analysed	7	1.35%	identify	5	1.17%
8	analyzed	7	1.35%	enhance	5	1.17%
9	create	6	1.16%	generated	5	1.17%
10	identify	6	1.16%	implemented	5	1.17%
11	enhance	6	1.16%	indicate	5	1.17%
12	focused	6	1.16%	create	4	0.93%
13	creating	6	1.16%	participate	4	0.93%
14	focuses	6	1.16%	demonstrate	4	0.93%
15	achieve	5	0.96%	ensure	4	0.93%
16	affect	5	0.96%	promote	4	0.93%
17	ensure	5	0.96%	transform	4	0.93%
18	promote	5	0.96%	analyzed	4	0.93%
19	indicated	5	0.96%	analysing	4	0.93%
20	involved	5	0.96%	demonstrated	4	0.93%
	Total key types/ key tokens	**247/519**			**233/429**	

Table 23 shows the most frequent verbs for the NN authors. *Found* is first in the NNF sub-corpus (2,50%) and fifth in the NNM sub-corpus (1.40%). *Conducted* constitutes 1.93% of examples in the NNF sub-corpus, and is not present on the NNM list of the 20 most frequent keyword verbs. *Identified* (1.72%); *creating, focuses* (both 1.16%); *affect, indicated,* and *involved* (all constituting 0.96%) show the same tendency. *Indicated* is used by M in the base form of the verb, see the 12th position (1.17%). The same situation occurs for *identified,* see the third position on the NNF list (1.73%). *Identify* (1.17%), though, is eighth on the NNM wordlist. *Evaluate* (1.63%), which occupies the first position in the NNM sub-corpus, is fourth in the NNF sub-corpus (1.54%). *Focused* (1.63%) is the second most frequent verb in the NNM sub-corpus, whereas on the NNF wordlist, it is listed as 12th (1.16%). *Achieve* occupies the third position with 1.40%. However, it does not occur on the NNF wordlist. Similarly, verbs listed below are absent on the NNF wordlist: *assess, generate, implemented* (all three with 1.17%); *participate, demonstrate, ensure,* and *transform* (all four with 0.93%). Additionally, there is an interesting occurrence of the lexeme *analyze*, on the NNF list, which

has a total frequency of 14 and constitutes 2.70% (seven *analysed*, and seven *analyzed*). Thus, the lexeme *analyze* could be the most frequently used in the NNF sub-corpus, but it is not as the word forms are treated as separate types. Furthermore, verbs *analyzed*, and *analyzing*, with 0.93% on the NNM list and frequency of 4 each, occupy 18th and 19th positions. If treated as forms of the same lexeme, they would come second on the NNM list. The same applies to *demonstrate*, and *demonstrated*, which occupy the 14th and 20th positions on the NNM list.

4.5.3 Adjectives

Table 24 AWL keywords adjectives for gender division

No.	Keyword Adjective	Freq. occurrence for Females	%	Keyword Adjective	Freq. occurrence for Males	%
1	potential	19	2.90%	potential	24	3.42%
2	available	18	2.74%	available	23	3.28%
3	significant	18	2.74%	individual	22	3.13%
4	specific	17	2.59%	specific	22	3.13%
5	cultural	16	2.44%	physical	17	2.42%
6	complex	15	2.29%	traditional	15	2.14%
7	physical	15	2.29%	significant	14	1.99%
8	positive	13	1.98%	relevant	12	1.71%
9	traditional	11	1.68%	primary	11	1.57%
10	academic	11	1.68%	major	10	1.42%
11	environmental	10	1.52%	appropriate	10	1.42%
12	similar	10	1.52%	positive	10	1.42%
13	primary	10	1.52%	previous	10	1.42%
14	qualitative	10	1.52%	overall	10	1.42%
15	individual	9	1.37%	professional	10	1.42%
16	relevant	9	1.37%	subsequent	10	1.42%
17	negative	9	1.37%	unique	10	1.42%
18	overall	9	1.37%	complex	9	1.28%
19	major	8	1.22%	interactive	9	1.28%
20	interactive	8	1.22%	academic	9	1.28%
	Total key types/ key tokens	**175/656**			**169/702**	

Table 24 illustrates keyword adjectives used in the corpus. Both females' and males' most frequently used AWL keywords are *potential* and *available*. They occupy the first and the second positions. In the F and M sub-corpora, their respective percentages are as follows: the first adjective 2.90 (F) and 3.42 (M); the second adjective follows: 2.74 (F) and 3.28 (M). Next, *significant* (2.47%) occupies the third position on the F wordlist, and the seventh on the M wordlist with 1.99%. *Specific* accounts for 2.59% in the F sub-corpus and 3.13% in the M sub-corpus; it occupies fourth position on both lists. *Major* is listed as 19th and tenth, with

1.22% and 1.42% respectively. *Interactive* occupies the 20th (F) and the 19th (M) positions, representing percentages of 1.22 and 1.28, respectively. *Academic* constitutes 1.68% and 1.28% of all AWL adjectives used in the corpus and can be found in the tenth (F sub-corpus) and 20th (M sub-corpus) positions. *Individual* is third in the M sub-corpus; it does not occur on the F list. Other adjectives that do not appear on the F list are: *appropriate, previous, professional, subsequent,* and *unique.* On the other hand, *cultural, environmental, similar, qualitative,* and *negative* are present in the F sub-corpus and absent from the M list of the most frequently used adjectives. Additionally, both groups use negative adjectives (*non-prefixed*). On the F wordlist, negatives represent a total percentage of 3.20, and on the M list they account for 2.42%. This part of the presented data may suggest more frequent negative collocations in the F abstracts.

Table 25 AWL keyword adjectives for native writer division

No.	Keyword Adjective	Freq. occurrence for Native Females	%	Keyword Adjective	Freq. occurrence for Native Males	%
1	specific	13	4.04%	physical	17	4.59%
2	cultural	11	3.42%	potential	12	3.24%
3	physical	10	3.11%	individual	11	2.97%
4	complex	9	2.80%	specific	11	2.97%
5	available	8	2.48%	under	8	2.16%
6	significant	8	2.48%	traditional	8	2.16%
7	potential	8	2.48%	available	7	1.89%
8	individual	7	2.17%	significant	7	1.89%
9	similar	7	2.17%	relevant	7	1.89%
10	primary	6	1.86%	professional	7	1.89%
11	relevant	6	1.86%	subsequent	7	1.89%
12	overall	6	1.86%	previous	6	1.62%
13	positive	5	1.55%	primary	6	1.62%
14	statistical	5	1.55%	overall	6	1.62%
15	academic	5	1.55%	psychological	6	1.62%
16	preliminary	5	1.55%	appropriate	5	1.35%
17	distinct	4	1.24%	positive	5	1.35%
18	integrated	4	1.24%	interactive	5	1.35%
19	professional	4	1.24%	dynamic	5	1.35%
20	challenging	4	1.24%	empirical	5	1.35%
	Total key types/ key tokens	**123/322**			**127/370**	

Table 25 presents keyword adjectives in the sub-corpus of the N authors. Notably, the first two adjectives differ between the lists. For females, the top two adjectives are *specific* (4.04%) and *cultural* (3.42%). The adjective *specific* appears in the fourth position on the M wordlist, with a frequency of 2.97%, while *cultural* does not appear on the NM wordlist.

Physical (3.11%), the third adjective in the female (NF) sub-corpus, ranks first in the NM sub-corpus with a frequency of 4.59%. *Potential* (3.24%) is the second most common adjective in the NM wordlist and occupies the seventh position in the NF sub-corpus at 2.48%. Moreover, adjectives found exclusively among the top 20 in the NM sub-corpus include *so, under, traditional, subsequent, previous, psychological, appropriate, interactive, dynamic, and empirical*. In contrast, adjectives exclusive to the NF sub-corpus and absent from the males' wordlist are: *cultural, complex, similar, statistical, academic, preliminary, distinct, integrated, and challenging*. Finally, negative adjectives (those with the non- prefix) appear in both lists, with a slightly higher frequency for the NM authors (2.97%) compared to the NF list (2.80%).

Table 26 AWL keyword list adjectives for non-native writer

No.	Keyword Adjective	Freq. occurrence for Non-native Females	%	Keyword Adjective	Freq. occurrence for Non-native Males	%
1	potential	11	3.29%	available	16	4.82%
2	available	10	2.99%	potential	12	3.61%
3	significant	10	2.99%	individual	11	3.31%
4	positive	8	2.40%	specific	11	3.31%
5	traditional	8	2.40%	major	7	2.11%
6	environmental	7	2.10%	significant	7	2.11%
7	complex	6	1.80%	complex	7	2.11%
8	under	6	1.80%	under	7	2.11%
9	negative	6	1.80%	traditional	7	2.11%
10	academic	6	1.80%	unique	6	1.81%
11	qualitative	6	1.80%	economic	5	1.51%
12	major	5	1.50%	similar	5	1.51%
13	cultural	5	1.50%	appropriate	5	1.51%
14	physical	5	1.50%	positive	5	1.51%
15	virtual	5	1.50%	primary	5	1.51%
16	specific	4	1.20%	relevant	5	1.51%
17	primary	4	1.20%	negative	5	1.51%
18	strategic	4	1.20%	academic	5	1.51%
19	initial	4	1.20%	diverse	5	1.51%
20	interactive	4	1.20%	environmental	4	1.20%
	Total key types/ key tokens	**133/334**			**124/332**	

Table 26 displays key adjectives used by the NN authors of both sexes. The adjective *potential* appears on both lists with frequencies of 3.29% and 3.61% for the NNF and NNM, respectively. *Available*, with a frequency of 2.99%, ranks second in the female (F) sub-corpus and first in the NNM wordlist with a 4.82% occurrence ratio. The adjective *individual* ranks third on the NNM wordlist at 3.31%. *Significant* (2.99%) and *specific* (3.31%) occupy the

third and fourth positions on the NNF and NNM wordlists. *Complex* and *under* hold the same positions on both lists, namely seventh and eighth. *Academic* and *major* are listed as the tenth (1.80%) and 12th (1.50%) on the NNF list and the 18th (1.51%) and fifth (2.11%) on the males' list. *Environmental*, the last of the commonly present adjectives, occupies the sixth and 20th positions with frequencies of 2.10% and 1.20%, respectively. Other adjectives that occur only on the NNF list among the top 20 keywords are *negative, qualitative, cultural, physical, virtual, strategic, initial,* and *interactive*. Conversely, in the NNM sub-corpus, this trend is illustrated by adjectives such as: *individual, unique, economic, similar, appropriate, relevant,* and *diverse*. Lastly, negative adjectives (those with the non- prefix) are twice as frequent in the NNF sub-corpora, appearing with a percentage of 3.59%, compared to 1.81% in the NNM sub-corpora.

4.5.4 **Adverbs**

Table 27 AWL keyword adverbs for gender division

No.	Keyword Adverb	Freq. occurrence for Females	%	Keyword Adverb	Freq. occurrence for Males	%
1	significantly	12	10.71%	specifically	8	6.20%
2	positively	7	6.25%	significantly	7	5.43%
3	negatively	7	6.25%	previously	6	4.65%
4	prior	7	6.25%	primarily	5	3.88%
5	potentially	6	5.36%	qualitatively	5	3.88%
6	traditionally	4	3.57%	positively	4	3.10%
7	statistically	4	3.57%	potentially	4	3.10%
8	normally	3	2.68%	adequately	4	3.10%
9	previously	3	2.68%	hence	4	3.10%
10	primarily	3	2.68%	prior	4	3.10%
11	approximately	3	2.68%	subsequently	4	3.10%
12	randomly	3	2.68%	automatically	4	3.10%
13	thereby	3	2.68%	normally	3	2.33%
14	specifically	2	1.79%	negatively	3	2.33%
15	finally	2	1.79%	statistically	3	2.33%
16	adequately	2	1.79%	ultimately	3	2.33%
17	internally	2	1.79%	consistently	2	1.55%
18	accurately	2	1.79%	theoretically	2	1.55%
19	globally	2	1.79%	finally	2	1.55%
20	radically	2	1.79%	constantly	2	1.55%
	Total key types/ key tokens	**45/112**			**58/129**	

Table 27 lists the first twenty keyword adverbs for both sexes. *Significantly* ranks first in the F sub-corpus with a frequency of 10.71% and second in the M wordlist with 5.43%. *Positively*, the second most frequent adverb on the F list at 6.25%, occupies the sixth position on the M

list with a frequency of 3.10%. *Negatively* occurs exclusively for females, accounting for 6.25%. *Prior* is far more frequent for females, occupying the fourth position, compared to the tenth position for males. The top position for males, *specifically* (6.20%), is also listed on the F wordlist but ranks 15th with a frequency of 1.79%. *Previously* and *primarily* appear next to each other on both lists and are more frequently used by males, ranking fourth and fifth with frequencies of 4.65% and 3.88%, respectively. For females, these adverbs occupy the tenth and 11th positions, both with a frequency of 2.68%. Adverbs that occur only in the female wordlist include *traditionally, approximately, randomly, thereby, adequately, internally,* and *globally*. Adverbs exclusive to the male wordlist are *qualitatively, hence, subsequently, automatically, ultimately, consistently,* and *theoretically.*

Table 28 AWL keyword adverbs for native writer division

No.	Keyword Adverb	Freq. occurrence for Native Females	%	Keyword Adverb	Freq. occurrence for Native Males	%
1	significantly	8	11.94%	significantly	5	7.04%
2	positively	6	8.96%	specifically	5	7.04%
3	prior	6	8.96%	positively	3	4.23%
4	negatively	4	5.97%	previously	3	4.23%
5	statistically	4	5.97%	negatively	3	4.23%
6	traditionally	3	4.48%	prior	3	4.23%
7	primarily	2	2.99%	qualitatively	3	4.23%
8	adequately	2	2.99%	consistently	2	2.82%
9	approximately	2	2.99%	theoretically	2	2.82%
10	consistently	1	1.49%	potentially	2	2.82%
11	specifically	1	1.49%	primarily	2	2.82%
12	structurally	1	1.49%	physically	2	2.82%
13	theoretically	1	1.49%	reliably	2	2.82%
14	normally	1	1.49%	adequately	2	2.82%
15	potentially	1	1.49%	subsequently	2	2.82%
16	previously	1	1.49%	ultimately	2	2.82%
17	strategically	1	1.49%	predominantly	2	2.82%
18	conventionally	1	1.49%	randomly	2	2.82%
19	exclusively	1	1.49%	similarly	1	1.41%
20	physically	1	1.49%	finally	1	1.41%
	Total key types/ key tokens	**36/67**			**40/71**	

Table 28 displays twenty adverbs for the groups of the N authors. For both sexes, the first adverb is *significantly*. It is more frequent for the NF (11.94%) than for NM (7.04%). *Positively*, with 8.96%, is the second in the NF sub-corpus and the third in the NM sub-corpus (4.23%). *Specifically* is the second on the NM wordlist with 7.04%, but it is not present in the NF wordlist. *Previously* (4.23%) occupies the fourth position in the NM sub-corpus and the

17th position in the NF sub-corpus (1.49%). *Negatively* is listed as the fifth in the NM sub-corpus and the fourth in the NF sub-corpus, with frequencies of 4.23% and 5.97%, respectively. *Statistically* (5.97%) is the fifth adverb in the NF sub-corpus and does not occur in the NM wordlist. Other adverbs which do not occur in the NM wordlist are: *traditionally* (4.48%), *approximately* (2.99%), *structurally* (1.49%), *normally* (1.49%), *strategically*, *conventionally*, and *exclusively* (all with 1.49%). On the other hand, adverbs that occur only in the NM wordlist are *qualitatively* (4.23%) and adverbs occupying positions from 14 to 20.

Table 29 AWL keyword adverbs for non-native writer

No.	Keyword Adverb	Freq. occurrence for Non-native Females	%	Keyword Adverb	Freq. occurrence for Non-native Males	%
1	potentially	5	11.11%	specifically	3	5.17%
2	significantly	4	8.89%	previously	3	5.17%
3	negatively	3	6.67%	primarily	3	5.17%
4	finally	2	4.44%	hence	3	5.17%
5	normally	2	4.44%	statistically	3	5.17%
6	previously	2	4.44%	automatically	3	5.17%
7	randomly	2	4.44%	significantly	2	3.45%
8	thereby	2	4.44%	normally	2	3.45%
9	periodically	1	2.22%	potentially	2	3.45%
10	specifically	1	2.22%	exclusively	2	3.45%
11	appropriately	1	2.22%	adequately	2	3.45%
12	inappropriately	1	2.22%	internally	2	3.45%
13	positively	1	2.22%	subsequently	2	3.45%
14	primarily	1	2.22%	eventually	2	3.45%
15	traditionally	1	2.22%	thereby	2	3.45%
16	constantly	1	2.22%	qualitatively	2	3.45%
17	apparently	1	2.22%	creatively	1	1.72%
18	approximately	1	2.22%	structurally	1	1.72%
19	internally	1	2.22%	finally	1	1.72%
20	prior	1	2.22%	positively	1	1.72%
	Total key types/ key tokens	**29/45**			**36/58**	

Table 29 illustrates the most frequent adverbs for the NN authors in the corpus. The first adverb on the NNM wordlist is *specifically* (5.17%), which is also present in the NNF sub-corpus (2.22%, 11th position). *Potentially* (11.11%) is the first adverb in the NNF sub-corpus and appears as the ninth in the NNM sub-corpus (3.45%). *Significantly* (8.98%) occupies the second position on the NNF wordlist but does not occur on the NNM wordlist. *Negatively* (6.67%) and *randomly* (4.44%), along with *periodically*, *appropriately*, *inappropriately*, *traditionally*, *constantly*, *apparently*, *approximately*, and *prior* (all with 2.22%) occur only in the NNF sub-corpus. On the other hand, *primarily* (5.17%, the third position), present on the

NNM wordlist, does not occur in the NNF wordlist. Other adverbs specific to the NNM authors only are: *hence*, *statistically*, *automatically* (all three with 5.17%), *potentially* (3.45%), as well as *exclusively*, *adequately*, *internally*, *subsequently*, *qualitatively* (all five with 3.45%), and *further*, *creatively*, and *structurally* (both with 1.72%). *Normally* is the sixth in the NNF sub-corpus (4.44%) and the eighth in the NNM sub-corpus (3.45%). *Thereby* is the ninth in the NNF sub-corpus and 15th in the NNM sub-corpus, with frequencies of 4.44% and 3.45%, respectively. The last two adverbs on the NNM wordlist, *finally* and *positively* (both 1.72%), occur respectively in the fifth (4.44%) and 14th (2.22%) positions on the NNF wordlist.

4.6 Academic Field Division

The following section presents the findings combining academic field, speaker origin and gender factors. The data of the corpus was sorted with ProtAnt software. The method used has been described in section 3.5. The presented keywords refer to the words found in abstracts that constitute the corpus.

4.6.1 Nouns

Table 30 Frequency of nouns used by females in all academic fields

No.	Keyword Noun	Freq. in Comp. Statistics	%	Keyword Noun	Freq. in Humanities	%	Keyword Noun	Freq. in Biology &Health	%
1	analysis	6	4.76%	research	25	3.41%	analysis	19	3.47%
2	access	5	3.97%	analysis	24	3.27%	intervention	18	3.28%
3	approach	4	3.17%	approach	18	2.45%	research	13	2.37%
4	framework	4	3.17%	process	16	2.18%	process	12	2.19%
5	environment	3	2.38%	focus	16	2.18%	promotion	10	1.82%
6	community	3	2.38%	context	15	2.04%	evidence	9	1.64%
7	challenge	3	2.38%	environment	13	1.77%	method	9	1.64%
8	transition	3	2.38%	communication	11	1.50%	community	9	1.64%
9	version	3	2.38%	project	11	1.50%	context	8	1.46%
10	inference	3	2.38%	role	10	1.36%	consumption	8	1.46%
11	concept	2	1.59%	design	10	1.36%	impact	8	1.46%
12	estimation	2	1.59%	interaction	10	1.36%	environment	7	1.28%
13	formula	2	1.59%	community	9	1.23%	survey	7	1.28%
14	function	2	1.59%	implementation	9	1.23%	energy	7	1.28%
15	method	2	1.59%	theory	8	1.09%	index	7	1.28%
16	source	2	1.59%	evaluation	8	1.09%	approach	6	1.09%
17	aspect	2	1.59%	impact	8	1.09%	area	6	1.09%
18	culture	2	1.59%	framework	8	1.09%	income	6	1.09%
19	impact	2	1.59%	technology	8	1.09%	period	6	1.09%
20	institution	2	1.59%	assessment	7	0.95%	design	6	1.09%
Total key types/ key tokens		**77/126**			**258/734**			**206/548**	

Table 30 presents the first twenty most frequent nouns used by F across the studied academic fields. The first similar noun in all fields is *analysis*. It occupies the first position in Computational Statistics (CS) and Biology & Health (BH) fields, accounting for 4.76% and 3.47%, respectively. In Humanities (H), *analysis* comes second and represents 3.27% of this sub-corpus. Next, *approach* is the third on the lists of CS and H, with 3.17% and 2.45%, respectively, while for BH it occupies the 16th position (1.09%). The third similar noun, *environment*, is the fifth most frequent in CS (2.38%), the seventh in H (1.77%), and the 12th in BH (1.09%). *Community* is, in the same order of the academic fields, the sixth (2.38%) in CS, 13th (1.23%) in H, and eighth (1.64%) in BH. *Impact* has a frequency of 1.59% in CS (19th position), 1.09% in H (17th position), and 1.46% in BH (11th position). The nouns that occur only on the lists of the first twenty in H and BH sub-corpora are *research*, *process*, and *context*. *Research* holds the first (3.41%) and the third (2.37%) positions. *Process* is the fourth (2.18% and 2.19%) in both fields. *Context* is the sixth (2.04%) and the ninth (1.46%). Finally, the only noun that occurs in both CS and BH is *method*, occupying the 16th position (1.59%) in CS and the seventh position (1.64%) in BH. Interestingly, nouns with a negative prefix (*non-*) are present in all three fields. In CS, they constitute 2.38% of the AWL key nouns; in H, it is 1.50%; and in BH, the percentage of overall key tokens is 1.28%.

Table 31 Frequency of nouns used by males in all academic fields

No.	Keyword Noun	Freq. in Comp. Statistics	%	Keyword Noun	Freq. in Humanities	%	Keyword Noun	Freq. in Biology &Health	%
1	analysis	17	6.77%	research	29	3.83%	analysis	11	3.02%
2	method	8	3.19%	context	21	2.77%	impact	11	3.02%
3	function	6	2.39%	analysis	20	2.64%	design	10	2.75%
4	code	6	2.39%	approach	15	1.98%	research	9	2.47%
5	approach	5	1.99%	design	15	1.98%	intervention	9	2.47%
6	distribution	5	1.99%	framework	15	1.98%	evidence	8	2.20%
7	framework	5	1.99%	role	14	1.85%	participation	8	2.20%
8	processing	4	1.59%	process	13	1.72%	intensity	7	1.92%
9	research	4	1.59%	project	13	1.72%	assessment	6	1.65%
10	community	4	1.59%	environment	12	1.58%	function	6	1.65%
11	output	4	1.59%	focus	12	1.58%	focus	6	1.65%
12	project	4	1.59%	range	10	1.32%	culture	5	1.37%
13	version	4	1.59%	technology	10	1.32%	range	5	1.37%
14	inference	4	1.59%	access	10	1.32%	interaction	5	1.37%
15	area	3	1.20%	concept	9	1.19%	extraction	5	1.37%
16	procedure	3	1.20%	assessment	8	1.06%	isolation	5	1.37%
17	process	3	1.20%	policy	8	1.06%	duration	5	1.37%
18	response	3	1.20%	impact	8	1.06%	approach	4	1.10%
19	source	3	1.20%	interaction	8	1.06%	period	4	1.10%

20	theory	3	1.20%	communication	7	0.92%	process	4	1.10%
Total key types/ key tokens		**132/251**			**244/758**			**168/364**	

Table 31 demonstrates twenty nouns used most frequently by M across the studied disciplines. *Analysis* is the first in Computational Statistics (CS) and Biology & Health (BH), with frequencies of 6.77% and 3.02%, respectively. It is also present on the Humanities (H) wordlist with 2.64% (third position). *Research* (3.83%) is the first in the H sub-corpus, the ninth (1.59%) in the CS sub-corpus, and the fourth (2.47%) in the BH sub-corpus. *Process* is the last noun present on all lists, ranking 18th (1.20%) in CS, eighth (1.72%) in H, and 20th (1.10%) in BH. The following nouns are present in both the H and BH lists: *focus, range, assessment, impact*, and *interaction*. *Focus* and *range* occupy the same positions in both fields; *focus* is 11th (1.58% and 1.65%), and *range* is 13th (1.32% and 1.37%). *Assessment* is 16th in the H sub-corpus (1.06%) and ninth in the BH sub-corpus (1.65%). *Impact* is second on the BH wordlist (3.02%) and 18th on the H wordlist (1.06%). The last of the above-mentioned nouns, *interaction*, is 19th in the H sub-corpus and 14th in the BH sub-corpus, accounting for 1.06% and 1.37%, respectively. Nouns common to the fields of CS and H are *framework* and *project*. *Framework* is fifth on the H wordlist (1.98%) and seventh on the CS wordlist (1.99%). *Project* ranks ninth on the H list (1.72%) and 12th in CS (1.59%). Additionally, nouns with negative prefixes (*non-*) can be found in the fields of CS and H, with frequencies of 1.45% and 1.59%, respectively. Interestingly, they are not present in the BH field. The last noun similarly used in the two fields is *function*, occurring in CS (third position, 2.39%) and BH (tenth position, 1.65%).

Table 32 Frequency of nouns used by native females in all academic fields

No.	Keyword Noun	Freq. in Comp. Statistics	%	Keyword Noun	Freq. in Humanities	%	Keyword Noun	Freq. in Biology &Health	%
1	analysis	4	6.45%	analysis	14	4.23%	intervention	15	4.39%
2	access	3	4.84%	research	10	3.02%	research	10	2.92%
3	approach	2	3.23%	focus	9	2.72%	process	9	2.63%
4	environment	2	3.23%	approach	8	2.42%	promotion	9	2.63%
5	method	2	3.23%	community	8	2.42%	analysis	8	2.34%
6	community	2	3.23%	non-(prefix)	7	2.11%	evidence	8	2.34%
7	framework	2	3.23%	context	6	1.81%	community	7	2.05%
8	transition	2	3.23%	identity	6	1.81%	evaluation	6	1.75%
9	inference	2	3.23%	role	6	1.81%	index	6	1.75%
10	concept	1	1.61%	interaction	6	1.81%	approach	5	1.46%
11	evidence	1	1.61%	range	5	1.51%	context	5	1.46%
12	formula	1	1.61%	project	5	1.51%	environment	5	1.46%

13	function	1	1.61%	process	4	1.21%	strategy	5	1.46%
14	policy	1	1.61%	structure	4	1.21%	outcome	5	1.46%
15	response	1	1.61%	theory	4	1.21%	energy	5	1.46%
16	role	1	1.61%	variation	4	1.21%	availability	4	1.17%
17	source	1	1.61%	task	4	1.21%	method	4	1.17%
18	aspect	1	1.61%	communication	4	1.21%	period	4	1.17%
19	culture	1	1.61%	target	4	1.21%	consumption	4	1.17%
20	institution	1	1.61%	environment	3	0.91%	design	4	1.17%
Total key types/ key tokens		**50/62**			**152/311**			**164/342**	

Table 32 comprises nouns used by the NF authors in the corpus. There are four nouns present in all three fields. *Analysis* occupies the first position on the lists of Computational Statistics (CS) (6.45%) and Humanities (H) (4.23%); it is the fifth on the Biology & Health (BH) wordlist (2.34%). The second noun, *approach*, occupies the third, fourth, and tenth positions, with percentages of 3.23%, 2.42%, and 1.46%, respectively in CS, H, and BH. *Environment* is the fourth in CS (3.23%), 20th in H (0.91%), and 12th in BH (1.46%). The last of these nouns, *community*, is the sixth in CS (3.23%), fifth in H (2.42%), and seventh in BH (2.05%). Additionally, there are two nouns present only on the H and BH wordlists: *research* and *process*. *Research* occupies the second position in both fields, with frequencies of 3.02% and 2.92%, respectively. *Process* occupies the third position in BH (2.63%) and the 13th position in H (1.21%). Furthermore, *role* occurs in both CS and H fields, occupying the 16th position in CS (1.61%) and the ninth position in H (1.81%). *Method* is the last noun present on the list of at least two fields, appearing in CS (fifth, 3.23%) and BH (17th, 1.17%).

Table 33 Frequency of nouns used by native males in all academic fields

No.	Keyword Noun	Freq. in Comp. Statistics	%	Keyword Noun	Freq. in Humanities	%	Keyword Noun	Freq. in Biology &Health	%
1	analysis	6	5.31%	analysis	14	3.75%	analysis	10	4.15%
2	method	5	4.42%	context	14	3.75%	design	9	3.73%
3	approach	3	2.65%	research	11	2.95%	intervention	9	3.73%
4	area	3	2.65%	role	10	2.68%	participation	8	3.32%
5	distribution	3	2.65%	approach	7	1.88%	evidence	7	2.90%
6	function	3	2.65%	project	7	1.88%	research	7	2.90%
7	framework	3	2.65%	interaction	6	1.61%	impact	7	2.90%
8	technique	3	2.65%	access	6	1.61%	function	5	2.07%
9	code	3	2.65%	concept	5	1.34%	focus	5	2.07%
10	inference	3	2.65%	policy	5	1.34%	duration	5	2.07%
11	estimation	2	1.77%	process	5	1.34%	assessment	4	1.66%
12	process	2	1.77%	design	5	1.34%	range	4	1.66%
13	role	2	1.77%	focus	5	1.34%	survey	4	1.66%
14	community	2	1.77%	impact	5	1.34%	proportion	4	1.66%

15	element	2	1.77%	framework	5	1.34%	isolation	4	1.66%
16	feature	2	1.77%	technology	5	1.34%	intensity	4	1.66%
17	survey	2	1.77%	area	4	1.07%	approach	3	1.24%
18	outcome	2	1.77%	theory	4	1.07%	context	3	1.24%
19	equivalent	2	1.77%	range	4	1.07%	process	3	1.24%
20	accuracy	2	1.77%	communication	4	1.07%	theory	3	1.24%
Total key types/ key tokens		**75/113**			**177/373**			**124/241**	

Table 33 displays the most frequently used nouns for the NM used in the corpus. Only three nouns occur in all three fields: *analysis*, *approach*, and *process*. *Analysis* occupies the first position in all fields with 5.31% in CS, 3.75% in H, and 4.15% in BH, respectively. The second noun, *approach*, is the third for CS (2.65%), the fifth for H (1.88%), and the 17th for BH (1.24%). *Process* is 12th on the first two lists, CS and H, with 1.77% and 1.34%, respectively. It occupies the 19th position on the BH wordlist with a frequency of 1.24%. Next, there are 12 nouns present on two of the word lists. *Area* is the fourth in CS (2.65%) and 17th in H (1.07%) sub-corpus. *Framework* is the seventh (2.65%) and 15th (1.34%). *Role* is 13th and the fourth, respectively, for CS (1.77%) and H (2.68%) fields. The following two nouns are present on the C and BH wordlists; consequently, *function* is the sixth (2.65%) and the eighth (2.07%). *Survey* ranks 17th (1.77%) and 13th (1.66%), respectively, in CS and BH sub-corpus. Next, six nouns belong to both H and BH. Thus, *research* is the third and the sixth, with 2.95% for H and 2.90% for BH. *Design* occupies the second position on the BH (3.73%) wordlist and 12th on the H (1.34%) wordlist. Thirdly, *focus* is the ninth and 13th, respectively, for BH (2.07%) and H (1.34%) fields. *Impact* is listed in the 14th and the seventh positions, with 1.34% and 2.90%, on the H and BH wordlists. Lastly, *theory* occupies the 18th and 20th positions. The last similar noun is *range*, which occupies the 19th place in H with 1.07%, and the 12th place with 1.66% in the BH field.

Table 34 Frequency of nouns used by non-native females in all academic fields

No.	Keyword Noun	Freq. in Comp. Statistics	%	Keyword Noun	Freq. in Humanities	%	Keyword Noun	Freq. in Biology &Health	%
1	analysis	2	3.13%	research	15	3.55%	analysis	11	5.34%
2	approach	2	3.13%	process	12	2.84%	area	6	2.91%
3	estimation	2	3.13%	analysis	10	2.36%	method	5	2.43%
4	framework	2	3.13%	approach	10	2.36%	panel	5	2.43%
5	access	2	3.13%	environment	10	2.36%	consumption	4	1.94%
6	challenge	2	3.13%	context	9	2.13%	impact	4	1.94%
7	generation	2	3.13%	design	9	2.13%	objective	4	1.94%
8	version	2	3.13%	implementation	8	1.89%	context	3	1.46%
9	simulation	2	3.13%	evaluation	7	1.65%	distribution	3	1.46%

10	analyst	1	1.56%	focus	7	1.65%	income	3	1.46%
11	concept	1	1.56%	technology	7	1.65%	occurrence	3	1.46%
12	context	1	1.56%	communication	7	1.65%	process	3	1.46%
13	distribution	1	1.56%	assessment	6	1.42%	research	3	1.46%
14	environment	1	1.56%	project	6	1.42%	evaluated	3	1.46%
15	formula	1	1.56%	impact	5	1.18%	survey	3	1.46%
16	function	1	1.56%	framework	5	1.18%	link	3	1.46%
17	interpretation	1	1.56%	perspective	5	1.18%	concentration	3	1.46%
18	issue	1	1.56%	innovation	5	1.18%	intervention	3	1.46%
19	processing	1	1.56%	majority	4	0.95%	depression	3	1.46%
20	similarity	1	1.56%	method	4	0.95%	environment	2	0.97%
Total key types/ key tokens		**54/64**			**193/423**			**118/206**	

Table 34 shows the most frequent nouns for the NNF writers across the studied academic fields. Two of the most frequent nouns occur in all three fields. The first of them, *analysis*, occupies the first position on two of the wordlists: CS and BH (3.13% and 5.34%). On the H wordlist, it holds the third position (2.36%). The second noun present on all of the word lists is *context*. It is listed as 12th, sixth, and eighth, with percentages of 1.56%, 2.13%, and 1.46%, respectively. One noun occurs on the CS and BH lists, i.e., *distribution*. It is 13th in CS (1.56%) and the ninth in BH (1.46%) sub-corpus. Finally, three similar nouns are present in the CS and H pair: *approach*, *framework*, and *environment*. Firstly, *approach* occupies the second and the fourth positions, respectively, with 3.13% for CS and 2.36% for H. Secondly, *framework* is positioned as the fourth (3.13% in CS) and 16th (1.18% in H). The last of the three, *environment*, is 14th in the CS (1.56%) wordlist and fifth in the H (2.36%) wordlist. Finally, there are four similar nouns on the lists of H and BH fields. *Research* is listed as the first in H (3.55%) and 13th in BH (1.46%). *Process* is the second (2.84%) and 12th (1.46%). Thirdly, *impact* is 15th on the H wordlist and sixth on the BH list, representing 1.18% and 1.94%, respectively. The last of the nouns, *method*, is the third and 20th, respectively, for BH and H, with 2.43% and 0.95%.

Table 35 Frequency of nouns used by non-native males in all academic fields

No.	Keyword Noun	Freq. in Comp. Statistics	%	Keyword Noun	Freq. in Humanities	%	Keyword Noun	Freq. in Biology &Health	%
1	analysis	11	7.97%	research	18	4.68%	culture	5	4.07%
2	research	4	2.90%	design	10	2.60%	impact	4	3.25%
3	function	3	2.17%	framework	10	2.60%	region	4	3.25%
4	method	3	2.17%	environment	9	2.34%	period	3	2.44%
5	processing	3	2.17%	approach	8	2.08%	structure	3	2.44%
6	code	3	2.17%	process	8	2.08%	concentration	3	2.44%
7	output	3	2.17%	context	7	1.82%	resolution	3	2.44%

8	project	3	2.17%	focus	7	1.82%	detection	3	2.44%
9	version	3	2.17%	analysis	6	1.56%	intensity	3	2.44%
10	approach	2	1.45%	range	6	1.56%	assessment	2	1.63%
11	context	2	1.45%	project	6	1.56%	distribution	2	1.63%
12	distribution	2	1.45%	assessment	5	1.30%	economy	2	1.63%
13	procedure	2	1.45%	text	5	1.30%	occurrence	2	1.63%
14	response	2	1.45%	technology	5	1.30%	research	2	1.63%
15	source	2	1.45%	challenge	5	1.30%	source	2	1.63%
16	theory	2	1.45%	network	5	1.30%	variation	2	1.63%
17	community	2	1.45%	motivation	5	1.30%	evaluated	2	1.63%
18	computer	2	1.45%	concept	4	1.04%	contribution	2	1.63%
19	text	2	1.45%	role	4	1.04%	interaction	2	1.63%
20	framework	2	1.45%	source	4	1.04%	contrast	2	1.63%
Total key types/ key tokens		**95/138**			**177/385**			**83/123**	

Table 35 presents the most frequent nouns of the NNM used in the corpus. Only two nouns occur on the lists of all three fields: *research* and *source*. The first of the two is positioned first for H (4.68%), second for CS (2.90%), and 14th for BH (1.63%). The other is set as 15th for BH (1.63%), 16th for CS (1.45%), and 20th for H (1.04%). Next, six nouns occur on the lists of the two studied fields. Four of those nouns are present on the lists of CS and H. Thus, *analysis* is the first (7.97%) in the CS sub-corpus and the ninth (1.56%) in the H sub-corpus. *Project* is the eighth on the CS wordlist. It constitutes 2.17% of the overall AWL tokens and occupies the 11th position with 1.56% on the H wordlist. *Context* is the seventh and 12th, respectively, for H (1.82%) and CS (1.45%). *Approach* occupies the fifth position on the H list with a percentage of 2.08. It holds the 11th position on the CS list with 1.45%. *Text*, the last co-occurring noun for the CS and H fields, has 1.45% and 1.30%, occupying the 20th and 13th positions. *Framework* is present on the lists of CS and H, representing 1.45% and 2.60%, with 20th (CS) and third (H) positions. *Distribution* occurs in the sub-corpora of CS and BH, with 1.45% (13th position) and 1.63% (11th position), respectively. Lastly, *assessment* is present on the wordlists of the BH and H fields; it occupies the tenth and 12th positions, with percentages of 1.63% and 1.30%.

4.6.2 **Verbs**

Table 36 Frequency of verbs used by females in all academic fields

No.	Keyword Verb	Freq. in Comp. Statistics	%	Keyword Verb	Freq. in Humanities	%	Keyword Verb	Freq. in Biology &Health	%
1	requires	4	5.13%	found	10	1.82%	conducted	15	3.27%
2	create	3	3.85%	creating	10	1.82%	assessed	12	2.61%
3	demonstrating	3	3.85%	conducted	10	1.82%	found	12	2.61%
4	accessing	3	3.85%	create	9	1.64%	identified	10	2.18%

5	distribute	2	2.56%	analysed	9	1.64%	indicated	9	1.96%
6	contribute	2	2.56%	focuses	9	1.64%	perceived	8	1.74%
7	interact	2	2.56%	identify	8	1.46%	evaluate	7	1.53%
8	enable	2	2.56%	ensure	7	1.28%	promote	7	1.53%
9	exceed	2	2.56%	investigate	7	1.28%	analyzed	7	1.53%
10	detect	2	2.56%	enhance	7	1.28%	involved	7	1.53%
11	validated	2	2.56%	focused	7	1.28%	designed	7	1.53%
12	triggered	2	2.56%	designed	7	1.28%	assess	6	1.31%
13	editing	2	2.56%	achieve	6	1.09%	analysed	6	1.31%
14	created	2	2.56%	integrate	6	1.09%	evaluated	6	1.31%
15	required	2	2.56%	investigates	6	1.09%	focused	5	1.09%
16	exposed	2	2.56%	establish	5	0.91%	participated	5	1.09%
17	analyze	1	1.28%	implementing	5	0.91%	selected	5	1.09%
18	identify	1	1.28%	created	5	0.91%	linked	5	1.09%
19	achieve	1	1.28%	identified	5	0.91%	registered	5	1.09%
20	construct	1	1.28%	involved	5	0.91%	implemented	5	1.09%
Total key types/ key tokens		**57/78**			**259/548**			**199/459**	

Table 36 presents the most commonly used verbs by F authors across various academic disciplines. Three verbs are consistent across all fields: *analyze*, *analysed*, and *identify*. The verb *analyze* ranks 17th on the CS list, accounting for 1.28% of usage. The verb *analysed* ranks fifth in the Humanities (H) field (1.64%) and ninth in Biology and Health (BH) (1.53%). *Identify* appears as the seventh verb on the H list (1.46%) and 18th on the CS list (1.28%). Additionally, *identified* is listed as the fourth most frequent verb on the BH list (2.18%) and 19th on the H list (0.91%). One key verb type occurs in the sub-corpus of CS and H, i.e., *create*. It occupies the second position on the CS list (3.85%) and the fourth position on the H list (1.64%). *Created* occupies the 14th and 18th positions, with 2.56% and 0.91%, respectively. Additionally, *creating* occurs only on the H wordlist and holds the second position with a percentage of 1.82. It is worth noting that it is equal in frequency to the verbs *found* and *conducted*. Furthermore, four verbs commonly occur in the sub-corpora of H and BH. The first verb, *found*, occupies the first and third positions, with 1.82% (H) and 2.61% (BH), respectively. The second verb, *conducted*, occupies the first position in the BH (3.27%) sub-corpus and the third in the H (1.82%) sub-corpus. The third verb, *focused*, is 12th (1.28% for H) and 15th (1.09% for BH). Additionally, *focuses* is present in the H wordlist in the sixth position with 1.64%. The last of the verbs, *designed*, occupies the 12th and 11th positions, respectively; it constitutes 1.28% in H and 1.53% in BH. Interestingly, the past participle appears in only 6 examples from the CS wordlist and 7 examples from the H wordlists. In contrast, on the BH wordlist, there are as many as 16 examples of the past participle used.

Table 37 Frequency of verbs used by males in all academic fields

No.	Keyword Verb	Freq. in Comp. Statistics	%	Keyword Verb	Freq. in Humanities	%	Keyword Verb	Freq. in Biology &Health	%
1	demonstrate	6	3.66%	participate	8	1.57%	assessed	10	3.66%
2	create	4	2.44%	identify	7	1.38%	assess	7	2.56%
3	assess	3	1.83%	designed	7	1.38%	found	6	2.20%
4	evaluate	3	1.83%	contribute	6	1.18%	enable	5	1.83%
5	implement	3	1.83%	promote	6	1.18%	analyzed	5	1.83%
6	generate	3	1.83%	analysed	6	1.18%	focused	5	1.83%
7	enhance	3	1.83%	focused	6	1.18%	published	5	1.83%
8	visualize	3	1.83%	found	6	1.18%	evaluate	4	1.47%
9	generated	3	1.83%	involved	6	1.18%	mediate	4	1.47%
10	illustrate	3	1.83%	required	6	1.18%	mediated	4	1.47%
11	derive	2	1.22%	implemented	6	1.18%	investigating	4	1.47%
12	identify	2	1.22%	achieve	5	0.98%	identified	4	1.47%
13	involve	2	1.22%	affect	5	0.98%	conducted	4	1.47%
14	require	2	1.22%	interact	5	0.98%	investigated	4	1.47%
15	ensure	2	1.22%	enhance	5	0.98%	exposed	4	1.47%
16	interact	2	1.22%	transform	5	0.98%	targeted	4	1.47%
17	validate	2	1.22%	emerged	5	0.98%	indicates	4	1.47%
18	convert	2	1.22%	designing	5	0.98%	demonstrate	3	1.10%
19	creating	2	1.22%	established	5	0.98%	implement	3	1.10%
20	conducting	2	1.22%	create	4	0.78%	enhance	3	1.10%
Total key types/ key tokens		**112/164**			**247/509**			**152/273**	

Table 37 lists verbs used by M across all studied disciplines. The verb *identify* is present in CS (1.22%) and H (1.38%) fields, occupying the 12th and the second positions, respectively. For the BH field, it occurs as a past participle (12th position and 1.47%). Three identical verbs are present on BH and H lists. The first one, *focused*, is the sixth and seventh, with 1.83% and 1.18%, respectively. The second verb, *found*, is the third for BH with 2.20% and the eighth for H, with a frequency of 1.18%. *Analysed* is the third verb present in these fields, occupying the sixth position in H (1.18%) and the fifth position in BH (1.83%). The two versions of the lexeme might suggest the origin of the writers. Additionally, three verbs are consistent across the fields of CS and BH: *demonstrate*, *assess*, and *implement*. The first, *demonstrate*, occurs as the most frequent verb on the CS list (3.66%), while it is 18th on the BH list (1.10%). The second verb, *assess*, is the third most frequent (1.83%) in CS and the second (2.56%) in BH. There is also the verb *implement*, which is set as the fifth (1.83%) and 20th (1.10%). *Implemented* occupies the 11th (1.18%) position on the H wordlist. Next, three verbs are present on the CS and H wordlists: *enhance*, *create*, and *involve*. The first verb, *enhance*, is set as the seventh (1.83%) and 15th (0.98%). *Create*, as listed above, is classified as 20th on

the H list (0.98%) and second on the CS list (1.22%). *Involve* is present on the CS wordlist and *involved* is present on the H wordlist; respectively, they constitute 1.22% (13th position) and 1.18% (ninth position). Finally, the past participle is used more frequently in the BH sub-corpus (with a total of 11 instances) than in the H (8 instances) and CS (1 instance) sub-corpora. This trend is similar to the F sub-corpus.

Table 38 Frequency of verbs used by native females in all academic fields

No.	Keyword Verb	Freq. in Comp. Statistics	%	Keyword Verb	Freq. in Humanities	%	Keyword Verb	Freq. in Biology &Health	%
1	create	2	4.88%	analysed	5	1.92%	conducted	10	3.77%
2	demonstrating	2	4.88%	found	5	1.92%	assessed	9	3.40%
3	accessing	2	4.88%	conducted	5	1.92%	assess	6	2.26%
4	requires	2	4.88%	investigates	5	1.92%	promote	6	2.26%
5	distribute	1	2.44%	create	4	1.54%	designed	6	2.26%
6	identify	1	2.44%	identify	4	1.54%	perceived	6	2.26%
7	achieve	1	2.44%	reveal	4	1.54%	indicated	5	1.89%
8	construct	1	2.44%	creating	4	1.54%	adjusted	5	1.89%
9	contribute	1	2.44%	investigated	4	1.54%	participated	4	1.51%
10	interact	1	2.44%	interpret	3	1.15%	revealed	4	1.51%
11	specify	1	2.44%	focused	3	1.15%	found	4	1.51%
12	enable	1	2.44%	participated	3	1.15%	identified	4	1.51%
13	generate	1	2.44%	implementing	3	1.15%	involved	4	1.51%
14	exceed	1	2.44%	involved	3	1.15%	implemented	4	1.51%
15	motivate	1	2.44%	selected	3	1.15%	targeted	4	1.51%
16	transform	1	2.44%	constitute	3	1.15%	estimate	3	1.13%
17	detect	1	2.44%	indicates	3	1.15%	obtain	3	1.13%
18	minimize	1	2.44%	occurs	3	1.15%	participate	3	1.13%
19	focused	1	2.44%	focuses	3	1.15%	investigate	3	1.13%
20	validated	1	2.44%	reveals	3	1.15%	analysed	3	1.13%
Total key types/ key tokens		**37/41**			**166/260**			**142/265**	

Table 38 presents NF verbs used in the researched abstracts, divided by discipline. One key verb type is present in all fields, i.e., *identify*. It is situated as the sixth on two of the lists: CS (2.44%) and H (1.54%), while in the BH sub-corpus it occupies the 12th position (1.51%) as a past participle. Furthermore, two verbs occur on the CS and H wordlists: *create* and *focused*. Correspondingly, *create* is listed as the first (4.88%) and fifth (1.54%) most frequent verb. The second verb, *focused*, is situated as 19th and 11th, respectively, with percentages of 2.44% in CS and 1.15% in the *H* sub-corpus. Next, six lexemes are commonly present on the H and BH wordlists. The first key verb type, *found*, is situated in the second and 11th positions, with 1.92% for H and 1.51% for BH, respectively. The second verb, *conducted*, occupies the first position in BH (3.77%) and the third in H (1.92%) sub-corpus. The third key type,

participated, is listed as the ninth on the BH and 12th on the H wordlists, with percentages of 1.51% and 1.15%, respectively. *Investigate* occurs as a past participle: ninth in the H sub-corpus and 19th in the BH sub-corpus. It is also displayed on the H list in the present, third-person, singular form as *investigates* (the fourth position, 1.92%). *Implement* appears in gerund form on the H wordlist (13th position, 1.15%) and as a past participle on the BH wordlist (14th position, 1.51%). The last key verb type, *indicate*, is shown on the H list in the present, third-person, singular form, situated in the 17th position with 1.15%. Meanwhile, *indicated* holds the seventh position and accounts for 1.89% of the total examples in the BH sub-corpus. In total, past participle verbs are more frequent in the BH sub-corpus (12 examples) than in the H (6 examples) or CS (3 examples) sub-corpora.

Table 39 Frequency of verbs used by native males in all academic fields

No.	Keyword Verb	Freq. in Comp. Statistics	%	Keyword Verb	Freq. in Humanities	%	Keyword Verb	Freq. in Biology &Health	%
1	demonstrate	3	3.33%	contribute	6	2.39%	assessed	9	5.11%
2	create	2	2.22%	affect	4	1.59%	assess	6	3.41%
3	derive	2	2.22%	construct	4	1.59%	enable	4	2.27%
4	identify	2	2.22%	participate	4	1.59%	found	4	2.27%
5	convert	2	2.22%	interact	4	1.59%	mediated	4	2.27%
6	visualize	2	2.22%	analysed	4	1.59%	conducted	4	2.27%
7	derived	2	2.22%	required	4	1.59%	targeted	4	2.27%
8	obtained	2	2.22%	enhance	3	1.20%	indicates	4	2.27%
9	surveyed	2	2.22%	focused	3	1.20%	implement	3	1.70%
10	illustrate	2	2.22%	participated	3	1.20%	mediate	3	1.70%
11	analyze	1	1.11%	emerged	3	1.20%	investigating	3	1.70%
12	assess	1	1.11%	designing	3	1.20%	required	3	1.70%
13	formulate	1	1.11%	emerging	3	1.20%	perceived	3	1.70%
14	interpret	1	1.11%	established	3	1.20%	published	3	1.70%
15	involve	1	1.11%	constitutes	3	1.20%	enhanced	3	1.70%
16	require	1	1.11%	involves	3	1.20%	create	2	1.14%
17	construct	1	1.11%	occurs	3	1.20%	identify	2	1.14%
18	evaluate	1	1.11%	reveals	3	1.20%	require	2	1.14%
19	invest	1	1.11%	analyse	2	0.80%	affect	2	1.14%
20	participate	1	1.11%	assume	2	0.80%	demonstrate	2	1.14%
Total key types/ key tokens		**76/90**			**157/251**			**109/176**	

Table 39 shows the first twenty most frequent verbs used by the NM authors across researched academic fields. There is one key verb type present on all lists, i.e., *require*, which occupies the 16th position in CS (1.11%), the 18th position in BH (1.14%), and the 12th position in H sub-corpora (*required*, with 1.59%). Additionally, there are two verbs present on CS and BH wordlists: *create* and *assess*. The first verb, *create*, occupies the second position

in the CS sub-corpus (2.22%) and the 16th position in the BH sub-corpus (1.14%). *Assess* is 13th in CS (1.11%) and the second in BH (3.41%). Furthermore, three verbs appear on both CS and H wordlists. The first, *construct*, occupies the third position on the H list (1.59%) and 17th position on the CS list (1.11%). The second verb, represented in two forms of the same lexeme *analyze* (US spelling) and *analyse* (UK spelling), ranks 11th (1.11%) on the CS list and 19th (0.80%) on the H list. Notably, the past participle *analysed* is present on the H list in the sixth position (1.59%). The third verb, appearing in two forms of the lexeme *involve,* holds the 15th position in the CS sub-corpus (1.11%) and the 16th position in H (*involves*, 1.20%). Finally, there are two similar verbs on H and BH lists: *affect* and *required*. The first verb, *affect*, represents 1.59% in H (the second position) and 1.14% in BH (the 19th position). *Required* is ranked seventh in H sub-corpus (1.59%) and 12th in BH sub-corpus (1.70%). Interestingly, the overall frequency of the past participle is slightly higher in the BH sub-corpus (7 words) compared to H sub-corpus (6 words) or CS sub-corpora (2 words).

Table 40 Frequency of verbs used by non-native females in all academic fields

No.	Keyword Verb	Freq. in Comp. Statistics	%	Keyword Verb	Freq. in Humanities	%	Keyword Verb	Freq. in Biology &Health	%
1	requires	2	5.41%	designed	7	2.43%	found	8	4.12%
2	analyze	1	2.70%	creating	6	2.08%	identified	6	3.09%
3	create	1	2.70%	focuses	6	2.08%	evaluate	5	2.58%
4	distribute	1	2.70%	create	5	1.74%	analyzed	5	2.58%
5	evaluate	1	2.70%	achieve	5	1.74%	conducted	5	2.58%
6	contribute	1	2.70%	ensure	5	1.74%	indicated	4	2.06%
7	demonstrate	1	2.70%	investigate	5	1.74%	selected	4	2.06%
8	interact	1	2.70%	enhance	5	1.74%	registered	4	2.06%
9	investigate	1	2.70%	found	5	1.74%	isolated	4	2.06%
10	enable	1	2.70%	conducted	5	1.74%	occur	3	1.55%
11	enhance	1	2.70%	identify	4	1.39%	affect	3	1.55%
12	exceed	1	2.70%	integrate	4	1.39%	analysed	3	1.55%
13	recover	1	2.70%	promote	4	1.39%	assessed	3	1.55%
14	detect	1	2.70%	analysed	4	1.39%	occurred	3	1.55%
15	visualize	1	2.70%	focused	4	1.39%	detected	3	1.55%
16	assessed	1	2.70%	establish	3	1.04%	involved	3	1.55%
17	validated	1	2.70%	implement	3	1.04%	structured	3	1.55%
18	facilitated	1	2.70%	participating	3	1.04%	evaluated	3	1.55%
19	triggered	1	2.70%	integrating	3	1.04%	located	3	1.55%
20	seeking	1	2.70%	created	3	1.04%	indicates	3	1.55%
Total key types/ key tokens		**36/37**			**170/288**			**118/194**	

Table 40 illustrates the most frequent NNF verbs used in abstracts within the corpus. One key verb type, represented in two forms of the same lexeme *analyze,* is present across all research

fields. It is listed in its base form in the CS sub-corpus (second position, 2.70%), as well as the past participle form in the H (14th position, 1.39%) and BH sub-corpora (fourth and 12th positions). The presence of the past participle form of the verb *analyze* (in both US and UK spelling) constitutes 4.13% of the total. Furthermore, three verbs are present on the CS and H wordlists. *Create* occupies the third and fourth positions, with 2.70% in CS and 1.74% in H, respectively. *Investigate* ranks seventh in H (1.74%) and ninth in CS (2.70%). *Enhance* holds the 11th and eighth positions on the CS (2.70%) and H (1.74%) lists, respectively. Additionally, two verbs are present on the CS and BH wordlists. The first verb, *evaluate*, ranks third in BH (2.58%) and fifth in CS (2.70%). It is also present in the past participle form, occupying the 18th position (1.55%) in the BH sub-corpus. The second verb, *assessed*, is listed as 16th (2.70%) in CS and 13th (1.55%) in BH. Lastly, two verbs are present on both the H and BH wordlists. Firstly, *found* occupies the ninth position on the H list (1.74%) and the first on the BH list (4.12%). Secondly, *conducted* ranks fifth in BH (2.58%) and tenth in H (1.74%). Overall, the use of past participles in the NNF sub-corpus is more frequent in the BH field (16 instances) compared to the CS (5 instances) or H (3 instances) fields.

Table 41 Frequency of verbs used by non-native males in all academic fields

No.	Keyword Verb	Freq. in Comp. Statistics	%	Keyword Verb	Freq. in Humanities	%	Keyword Verb	Freq. in Biology &Health	%
1	demonstrate	3	4.05%	identify	5	1.94%	focused	4	4.12%
2	assess	2	2.70%	achieve	5	1.94%	evaluate	3	3.09%
3	create	2	2.70%	designed	5	1.94%	analyzed	3	3.09%
4	evaluate	2	2.70%	participate	4	1.55%	demonstrated	3	3.09%
5	interact	2	2.70%	promote	4	1.55%	exposed	3	3.09%
6	validate	2	2.70%	transform	4	1.55%	occur	2	2.06%
7	implement	2	2.70%	found	4	1.55%	generated	2	2.06%
8	generate	2	2.70%	involved	4	1.55%	inhibited	2	2.06%
9	enhance	2	2.70%	implemented	4	1.55%	found	2	2.06%
10	generated	2	2.70%	indicate	4	1.55%	identified	2	2.06%
11	creating	2	2.70%	ensure	3	1.16%	affected	2	2.06%
12	processed	2	2.70%	acquired	3	1.16%	evaluated	2	2.06%
13	transformed	2	2.70%	focused	3	1.16%	obtained	2	2.06%
14	automated	2	2.70%	analysing	3	1.16%	located	2	2.06%
15	estimate	1	1.35%	ensuring	3	1.16%	published	2	2.06%
16	involve	1	1.35%	created	3	1.16%	investigated	2	2.06%
17	require	1	1.35%	structured	3	1.16%	released	2	2.06%
18	ensure	1	1.35%	conducted	3	1.16%	assess	1	1.03%
19	specify	1	1.35%	identifies	3	1.16%	involve	1	1.03%
20	facilitate	1	1.35%	indicates	3	1.16%	respond	1	1.03%
Total key types/ key tokens		**58/74**			**164/258**			**74/97**	

Table 41 illustrates the most commonly used verbs by the NNM authors in the corpus. Notably, only one verb, *involve*, appears on all lists, ranking 16th and 19th in the CS and BH lists, respectively. In the Humanities, *involved* holds the eighth position and constitutes 1.55% of the total examples. Additionally, three verbs are included in both the CS and BH wordlists. *Evaluate* is ranked second in the BH sub-corpus (3.09%) and fourth in the CS sub-corpus (2.70%). The verb *generated* is sixth in the BH list (2.06%) and 11th in the CS list (2.70%). Its base form, *generate*, appears in the eighth position on the CS list (2.70%). The third verb, *assess*, ranks second (2.70%) in the CS list and 18th (1.03%) in the BH sub-corpus. Furthermore, two verbs appear in both the H and BH wordlists. *Focused* is first in the BH list (4.12%) and 13th in the H list (1.16%). The second verb, *found*, is seventh in the H list (1.55%) and ninth in the BH list (2.06%). Lastly, the verb *ensure* is included in both the CS and H lists, ranking 18th (1.35%) and 13th (1.16%), respectively. In summary, the past participle form is utilized more frequently in the BH field (16 times) compared to the H field (8 times) and the CS field (4 times).

4.6.3 Adjectives

Table 42 Frequency of adjectives used by females in all academic fields

No.	Keyword Adjective	Freq. in Comp. Statistics	%	Keyword Adjective	Freq. in Humanities	%	Keyword Adjective	Freq. in Biology &Health	%
1	complex	4	6.25%	cultural	12	3.53%	potential	14	5.56%
2	similar	3	4.69%	academic	11	3.24%	physical	14	5.56%
3	interactive	3	4.69%	available	10	2.94%	environmental	10	3.97%
4	dimensional	3	4.69%	specific	10	2.94%	significant	10	3.97%
5	challenging	3	4.69%	traditional	9	2.65%	specific	7	2.78%
6	available	2	3.13%	significant	8	2.35%	positive	7	2.78%
7	cultural	2	3.13%	qualitative	8	2.35%	primary	7	2.78%
8	statistical	2	3.13%	professional	7	2.06%	negative	7	2.78%
9	capable	2	3.13%	positive	6	1.76%	available	6	2.38%
10	classic	2	3.13%	innovative	6	1.76%	major	6	2.38%
11	dynamic	2	3.13%	individual	5	1.47%	complex	6	2.38%
12	analytic	1	1.56%	complex	5	1.47%	overall	6	2.38%
13	analytical	1	1.56%	potential	5	1.47%	relevant	5	1.98%
14	consistent	1	1.56%	under	5	1.47%	individual	4	1.59%
15	derivative	1	1.56%	technical	5	1.47%	initial	4	1.59%
16	major	1	1.56%	integrated	5	1.47%	analytical	3	1.19%
17	theoretical	1	1.56%	diverse	5	1.47%	economic	3	1.19%
18	appropriate	1	1.56%	virtual	5	1.47%	similar	3	1.19%
19	normal	1	1.56%	called	5	1.47%	final	3	1.19%
20	insecure	1	1.56%	similar	4	1.18%	previous	3	1.19%
Total key types/ key tokens		**45/64**			**131/340**			**102/252**	

Table 42 shows the most frequently occurring adjectives in the F abstracts within the corpus. Firstly, one adjective is used across all three fields. It is listed as the sixth most frequent adjective in the CS sub-corpus (3.13%), the third in H (2.94%), and the ninth in BH (2.38%). Secondly, six adjectives are present on two of the lists. Three of these adjectives are commonly used in both the H and BH sub-corpora: *complex*, *specific*, and *potential*. *Complex* occupies the 11th position in BH (2.38%) and the 13th position in H (1.47%). *Specific* is the fourth most frequent adjective on the H list and the fifth on the BH list, with percentages of 2.94% and 2.78%, respectively. The third adjective, *potential*, is listed as 14th in H (1.47%) and first in BH (5.56%). Moreover, two adjectives are commonly used in both the CS and H sub-corpora: *cultural* and *similar*. *Cultural* is listed in the seventh position in CS (3.13%) and the first position in H (3.24%). *Similar* occupies the second position in CS (4.69%) and the 20th position in H (1.18%). The last of the six adjectives, *major*, is present in the CS and BH fields, with percentages of 1.56% and 2.38%, respectively, occupying the 16th and 10th positions. Additionally, adjectives with the negative prefix *non-* are present across all fields, with total percentages of 3.24% in H, 4.69% in CS, and 2.78% in BH. The key types with the *non-* prefix found in the F sub-corpus include *non-programmers*, *non-referring*, *non-manual* (found four times in one abstract), *non-white*, *non-Japanese*, *non-native* (found three times in two abstracts), *non-grammatical*, *non-toxic* (found in three abstracts), *non-blooming*, *non-linear*, *non-stabilized*, *non-formal*, and *non-teachers* (found twice in one abstract).

Table 43 Frequency of adjectives used by males in all academic fields

No.	Keyword Adjective	Freq. in Comp. Statistics	%	Keyword Adjective	Freq. in Humanities	%	Keyword Adjective	Freq. in Biology &Health	%
1	available	12	8.96%	specific	18	4.60%	physical	16	9.04%
2	called	6	4.48%	individual	12	3.07%	potential	10	5.65%
3	individual	5	3.73%	potential	11	2.81%	subsequent	7	3.95%
4	complex	4	2.99%	traditional	11	2.81%	environmental	6	3.39%
5	positive	4	2.99%	significant	10	2.56%	individual	5	2.82%
6	potential	3	2.24%	professional	10	2.56%	primary	5	2.82%
7	interactive	3	2.24%	available	9	2.30%	psychological	5	2.82%
8	negative	3	2.24%	empirical	8	2.05%	significant	4	2.26%
9	accessible	3	2.24%	appropriate	7	1.79%	specific	4	2.26%
10	dimensional	3	2.24%	relevant	7	1.79%	previous	4	2.26%
11	statistical	3	2.24%	academic	7	1.79%	sufficient	4	2.26%
12	challenging	3	2.24%	unique	7	1.79%	overall	4	2.26%
13	classical	3	2.24%	consistent	6	1.53%	mental	4	2.26%
14	hierarchical	3	2.24%	major	6	1.53%	diverse	4	2.26%
15	analytic	2	1.49%	theoretical	6	1.53%	economic	3	1.69%
16	distributional	2	1.49%	previous	6	1.53%	distinct	3	1.69%

17	major	2	1.49%	interactive	6	1.53%	positive	3	1.69%
18	appropriate	2	1.49%	obvious	6	1.53%	relevant	3	1.69%
19	primary	2	1.49%	qualitative	6	1.53%	traditional	3	1.69%
20	relevant	2	1.49%	similar	5	1.27%	innovative	3	1.69%
Total key types/ key tokens		**64/134**			**130/391**			**80/177**	

Table 43 lists male writers' most frequently used adjectives in the corpus. Firstly, two adjectives are present across all fields: *individual* and *potential*. *Individual* ranks second in the H sub-corpus (3.07%), third in the CS sub-corpus (3.73%), and fifth in the BH sub-corpus (2.82%). *Potential* is the second most frequent adjective in the BH sub-corpus (5.65%), fourth in the H sub-corpus (2.81%), and sixth in the CS sub-corpus (2.24%). Secondly, several key adjectives are common to both the CS and H fields: *available*, *interactive*, *major*, *appropriate*, and *relevant*. *Available* ranks first on the CS list (8.96%) and eighth on the H list (2.30%). *Interactive* appears seventh on the CS list and 19th on the H list, with percentages of 2.24% and 1.53%, respectively. *Major* ranks 14th in the H sub-corpus and 17th in the CS sub-corpus, with 1.49% and 1.53% of total key adjectives used, respectively. *Appropriate* occupies the 18th position in the CS sub-corpus (1.49%) and the ninth in the H sub-corpus (1.79%). Finally, *relevant* ranks 20th in the CS sub-corpus (1.49%) and 10th in the H sub-corpus (1.79%). Thirdly, certain adjectives are prevalent in both the H and BH fields: *specific*, *traditional*, *significant*, and *previous*. *Specific* occupies the first position in the H sub-corpus (4.60%) and the ninth position in the BH sub-corpus (2.26%). *Traditional* ranks fifth in the H sub-corpus (2.81%) and 19th in the BH sub-corpus (1.69%). *Significant* appears sixth in the H sub-corpus (2.56%) and eighth in the BH sub-corpus (2.26%). The adjective *previous* is listed as the tenth most frequent in the BH sub-corpus (2.26%) and 16th in the H sub-corpus (1.53%). Lastly, two adjectives are common to both the CS and BH wordlists: *positive* and *primary*. *Positive* ranks fifth in the CS sub-corpus (2.99%) and 17th in the BH sub-corpus (1.68%), while *primary* is listed 19th in the CS sub-corpus (1.49%) and sixth in the BH sub-corpus (2.82%).

Table 44 Frequency of adjectives used by native females in all academic fields

No.	Keyword Adjective	Freq. in Comp. Statistics	%	Keyword Adjective	Freq. in Humanities	%	Keyword Adjective	Freq. in Biology &Health	%
1	similar	2	5.71%	cultural	9	2.67%	physical	10	3.97%
2	complex	2	5.71%	specific	7	2.08%	potential	7	2.78%
3	interactive	2	5.71%	academic	5	1.48%	specific	6	2.38%
4	dimensional	2	5.71%	available	4	1.19%	overall	5	1.98%
5	statistical	2	5.71%	significant	4	1.19%	individual	4	1.59%
6	challenging	2	5.71%	complex	4	1.19%	significant	4	1.59%
7	dynamic	2	5.71%	diverse	4	1.19%	relevant	4	1.59%

8	analytical	1	2.86%	migrant	4	1.19%	available	3	1.19%
9	available	1	2.86%	individual	3	0.89%	environmental	3	1.19%
10	derivative	1	2.86%	similar	3	0.89%	major	3	1.19%
11	appropriate	1	2.86%	distinct	3	0.89%	complex	3	1.19%
12	cultural	1	2.86%	positive	3	0.89%	primary	3	1.19%
13	insecure	1	2.86%	primary	3	0.89%	preliminary	3	1.19%
14	select	1	2.86%	professional	3	0.89%	economic	2	0.79%
15	unreliable	1	2.86%	statistical	3	0.89%	similar	2	0.79%
16	adequate	1	2.86%	empirical	3	0.89%	positive	2	0.79%
17	annual	1	2.86%	contextual	2	0.59%	restrictive	2	0.79%
18	approximate	1	2.86%	theoretical	2	0.59%	initial	2	0.79%
19	internal	1	2.86%	previous	2	0.59%	negative	2	0.79%
20	external	1	2.86%	traditional	2	0.59%	reliable	2	0.79%
Total key types/ key tokens		28/35			130/337			102/252	

Table 44 presents the most frequently used adjectives in the NF abstracts within the researched academic fields. Firstly, three adjectives are consistently found across all fields: *similar*, *complex*, and *available*. The first adjective, *similar,* occupies the first, tenth, and 15th positions, with percentages of 5.71%, 0.89%, and 0.79% for CS, H, and BH, respectively. *Complex* is listed as second, sixth, and 11th, with 5.71% in CS and 1.19% in both H and BH. The adjective *available* occupies the fourth, eighth, and ninth positions in H (1.19%), BH (1.19%), and CS (2.86%). Additionally, the adjective *cultural* appears in both the CS and H wordlists: *cultural*. It ranks first in H (2.67%) and 12th in CS (2.86%). Moreover, six adjectives are commonly present on both the H and BH lists: *specific* (second and third), *significant* (fifth and sixth), *individual* (ninth and fifth), *similar* (tenth and 15th), *positive* (12th and 16th), and *primary* (13th and 12th). *Specific* constitutes 2.08% in H and 2.38% in BH. The percentage of *significant* is 1.19% in H and 1.59% in BH. *Individual*, *similar*, *positive*, and *primary* have the same share in the H field, i.e., 0.89%, whereas in BH they constitute 1.59%, 0.79%, 0.79%, and 1.19%, respectively. Overall, it is worth noting that adjectives with the prefix *non-* constitute 2.86% and 2.67% of the total key adjective examples in the CS and H fields, respectively. All nine examples of *non-* prefixed adjectives in the *NF* sub-corpus include: *non-programmers*, *non-referring*, *non-manual* (found four times in one abstract), *non-white*, *non-native*, and *non-grammatical*.

Table 45 Frequency of adjectives used by native males in all academic fields

No.	Keyword Adjective	Freq. in Comp. Statistics	%	Keyword Adjective	Freq. in Humanities	%	Keyword Adjective	Freq. in Biology &Health	%
1	available	3	4.76%	specific	8	4.08%	physical	16	14.41%
2	statistical	3	4.76%	professional	7	3.57%	potential	7	6.31%

3	hierarchical	3	4.76%	individual	5	2.55%	subsequent	5	4.50%
4	called	3	4.76%	significant	5	2.55%	psychological	5	4.50%
5	analytic	2	3.17%	relevant	5	2.55%	individual	4	3.60%
6	individual	2	3.17%	empirical	5	2.55%	previous	4	3.60%
7	medical	2	3.17%	available	4	2.04%	overall	4	3.60%
8	dynamic	2	3.17%	appropriate	4	2.04%	mental	4	3.60%
9	analytical	1	1.59%	cultural	4	2.04%	specific	3	2.70%
10	derivative	1	1.59%	potential	4	2.04%	primary	3	2.70%
11	distributional	1	1.59%	traditional	4	2.04%	traditional	3	2.70%
12	environmental	1	1.59%	interactive	4	2.04%	sufficient	3	2.70%
13	major	1	1.59%	consistent	3	1.53%	innovative	3	2.70%
14	complex	1	1.59%	distinct	3	1.53%	environmental	2	1.80%
15	distinct	1	1.59%	technical	3	1.53%	significant	2	1.80%
16	normal	1	1.59%	global	3	1.53%	positive	2	1.80%
17	positive	1	1.59%	unique	3	1.53%	relevant	2	1.80%
18	potential	1	1.59%	qualitative	3	1.53%	academic	2	1.80%
19	primary	1	1.59%	contextual	2	1.02%	fundamental	2	1.80%
20	traditional	1	1.59%	economic	2	1.02%	contemporary	2	1.80%
Total key types/ key tokens		47/63			94/196			52/111	

Table 45 presents the most frequently used adjectives in the NM abstracts within the corpus. Three adjectives are consistently found across all studied fields: *individual*, *potential*, and *traditional*. *Individual* occupies the sixth, third, and fifth positions, respectively, with 3.17% in CS, 2.55% in H, and 3.60% in BH. *Potential* is listed as 18th (1.59%) in CS, 10th (2.04%) in H, and second (6.31%) in BH. *Traditional* occupies the 20th position in CS, and the 11th position in both H and BH, accounting for 1.59% in CS, 2.04% in H, and 2.70% in BH. Furthermore, two adjectives are common to both the CS and BH fields: *environmental* and *positive*. *Environmental* ranks 12th in CS (1.59%) and 14th in BH (1.80%). *Positive* occupies the 16th position in BH (1.80%) and the 17th in CS (1.69%). Next, two adjectives are common to the CS and H fields. *Available* occupies the first position in CS (4.76%) and the seventh in H (2.04%). The second adjective, *distinct*, ranks 14th in H (1.53%) and 16th in CS (1.59%). Lastly, three key adjectives are present on both the H and BH lists: *specific*, *significant*, and *relevant*. On the H list, these adjectives rank first (4.08%), fourth (2.55%), and fifth (2.55%), respectively. On the BH list, they occupy the ninth (2.70%), 15th (1.80%), and 17th (1.80%) positions. Moreover, it is worth noting that adjectives with the prefix *non-* are present in both the H and BH sub-corpora, constituting 4.08% in H and 1.80% in BH. All 10 examples of *non*-prefixed adjectives are listed below, each occurring only once: *non-inferiority*, *non-authoritarian*, *non-Americans*, *non-competition*, *non-existent*, *non-formal*, *non-jocular*, *non-professionals*, *non-interactive*, and *non-English*.

Table 46 Frequency of adjectives used by non-native females in all academic fields

No.	Keyword Adjective	Freq. in Comp. Statistics	%	Keyword Adjective	Freq. in Humanities	%	Keyword Adjective	Freq. in Biology &Health	%
1	complex	2	6.90%	traditional	7	3.83%	environmental	7	5.74%
2	analytic	1	3.45%	available	6	3.28%	potential	7	5.74%
3	available	1	3.45%	academic	6	3.28%	significant	6	4.92%
4	consistent	1	3.45%	qualitative	6	3.28%	positive	5	4.10%
5	major	1	3.45%	virtual	5	3.28%	negative	5	4.10%
6	similar	1	3.45%	significant	4	2.73%	primary	4	3.28%
7	theoretical	1	3.45%	potential	4	2.19%	physical	4	3.28%
8	cultural	1	3.45%	professional	4	2.19%	analytical	3	2.46%
9	normal	1	3.45%	innovative	4	2.19%	available	3	2.46%
10	traditional	1	3.45%	specific	3	2.19%	major	3	2.46%
11	interactive	1	3.45%	appropriate	3	1.64%	complex	3	2.46%
12	accessible	1	3.45%	cultural	3	1.64%	evident	2	1.64%
13	attributable	1	3.45%	institutional	3	1.64%	final	2	1.64%
14	dimensional	1	3.45%	positive	3	1.64%	previous	2	1.64%
15	integrated	1	3.45%	strategic	3	1.64%	initial	2	1.64%
16	parallel	1	3.45%	interactive	3	1.64%	mental	2	1.64%
17	challenging	1	3.45%	technical	3	1.64%	psychological	2	1.64%
18	capable	1	3.45%	technological	3	1.64%	unique	2	1.64%
19	classic	1	3.45%	integrated	3	1.64%	random	2	1.64%
20	classical	1	3.45%	challenging	3	1.64%	beneficial	1	0.82%
Total key types/ key tokens		27/29			97/183			67/122	

Table 46 shows the most frequently used adjectives by the NNF in the corpus. Firstly, one adjective appears across three lists, namely, *available*. It ranks third, second, and ninth in the CS (3.45%), H (3.28%), and BH (2.46%) sub-corpora, respectively. Secondly, four adjectives are shared between the CS and BH lists. The first of these, *complex*, holds the first position in CS (6.90%) and the 12th in BH (2.46%). The second, *major*, ranks fifth in CS and 11th in BH, with percentages of 3.45% and 2.46%, respectively. Additionally, both forms of *analytic* and *analytical* are present: *analytic* appears in the CS list at the second position (3.45%), while *analytical* is found in the BH list at the eighth position (2.46%). Thirdly, five adjectives occur on both the CS and H lists: *cultural*, *traditional*, *interactive*, *integrated*, and *challenging*. Specifically, *cultural* ranks eighth in CS (3.45%) and 12th in H (1.64%). Meanwhile, *traditional* is first in H (3.83%) and tenth in CS (3.45%). The adjective *interactive* occupies the 12th position in CS (3.45%) and the 17th in H (1.64%). Furthermore, *integrated* is listed as 15th in CS (3.45%) and 19th in H (1.64%). The last of these five adjectives, *challenging*, appears in the 17th position in CS (3.45%) and the 20th in H (1.64%). Lastly, three most frequent keyword adjectives found in both the H and BH sub-corpora are: *significant*, *potential*, and *positive*. Thus, *significant* ranks sixth in H (2.73%) and third in BH (4.92%).

The adjectives, *potential* and *positive*, rank seventh (2.19%) and 14th (1.64%) in H, and second (5.74%) and fourth (4.10%) in BH, respectively. It is also noteworthy that *non-* prefixed adjectives are commonly present across all fields studied. Collectively, they account for 6.90%, 2.19%, and 4.92% of the AWL adjectives used in the sub-corpora. The 12 examples include *non-Japanese*, *non-toxic* (found in three abstracts), *non-blooming*, *non-linear*, *non-stabilized*, *non-formal*, *non-native*, and *non-teachers* (each found twice in one abstract).

Table 47 Frequency of adjectives used by non-native males in all academic fields

No.	Keyword Adjective	Freq. in Comp. Statistic	%	Keyword Adjective	Freq. in Humanities	%	Keyword Adjective	Freq. in Biology &Health	%
1	available	9	12.68%	specific	10	5.13%	environmental	4	6.06%
2	individual	3	4.23%	individual	7	3.59%	economic	3	4.55%
3	complex	3	4.23%	potential	7	3.59%	distinct	3	4.55%
4	positive	3	4.23%	traditional	7	3.59%	potential	3	4.55%
5	challenging	3	4.23%	available	5	3.59%	under	3	4.55%
6	called	3	4.23%	significant	5	2.56%	diverse	3	4.55%
7	appropriate	2	2.82%	academic	5	2.56%	available	2	3.03%
8	potential	2	2.82%	major	4	2.56%	major	2	3.03%
9	relevant	2	2.82%	theoretical	4	2.05%	significant	2	3.03%
10	textual	2	2.82%	previous	4	2.05%	complex	2	3.03%
11	interactive	2	2.82%	obvious	4	2.05%	primary	2	3.03%
12	negative	2	2.82%	flexible	4	2.05%	dominant	2	3.03%
13	accessible	2	2.82%	unique	4	2.05%	negative	2	3.03%
14	dimensional	2	2.82%	consistent	3	2.05%	subsequent	2	3.03%
15	classical	2	2.82%	similar	3	1.54%	exposing	2	3.03%
16	comprehensive	2	2.82%	appropriate	3	1.54%	individual	1	1.52%
17	distributional	1	2.82%	under	3	1.54%	periodic	1	1.52%
18	major	1	1.41%	adequate	3	1.54%	similar	1	1.52%
19	similar	1	1.41%	overall	3	1.54%	specific	1	1.52%
20	structural	1	1.41%	professional	3	1.54%	abnormal	1	1.52%
Total key types/ key tokens		40/71			95/195			44/66	

Table 47 shows the twenty most frequently used adjectives in the researched fields by the NNM writers. Firstly, five adjectives appear on all the lists: *available*, *individual*, *potential*, *major*, and *similar*. In the CS field, *available* ranks first with 12.68%, whereas in the H and BH fields, it occupies the fifth (3.59%) and seventh (3.03%) positions, respectively. *Individual* is the second most frequent adjective in both CS (4.13%) and H (3.59%) fields; additionally, it holds the 16th position in the BH field (1.52%). Next, *potential* ranks eighth in CS (2.82%), third in H (3.59%), and fourth in BH (4.55%). The adjective, *major*, occupies the eighth position in both H (2.56%) and BH (3.03%) sub-corpora, whereas in the CS sub-corpus

it ranks 18th (1.41%). The last of the five adjectives, *similar*, appears in the 15th (1.54%), 18th (1.52%), and 19th (1.41%) positions in H, BH, and CS, respectively. Secondly, there are two adjectives common to both CS and BH lists: *complex* and *negative*. *Complex* ranks third (4.23%) in CS and tenth (3.03%) in BH. Meanwhile, *negative* occupies the 12th position in CS and 13th in BH, with frequencies of 2.82% and 3.03%, respectively. Thirdly, one adjective is present on both the CS and H lists, namely, *appropriate*. It ranks seventh in CS (2.82%) and 16th in H (1.54%). Lastly, three adjectives appear in both H and BH sub-corpora: *specific*, *significant*, and *under*. In the H field, they occupy the first (5.13%), sixth (2.56%), and 17th (1.54%) positions, respectively. Conversely, in the BH field, they are listed as 19th (1.52%), ninth (3.03%), and fifth (4.55%). Additionally, *non*-prefixed adjectives constitute 4.23% of the total AWL adjectives used in CS and 1.54% in H sub-corpora. All six examples of the non-prefixed adjectives found in the corpus are *non-smoothed*, *non-bioinformaticians*, *non-parametric*, *non-academic*, *non-traditional*, and *non-native*.

4.6.4 **Adverbs**

The tables in Section 4.6.4 are incomplete due to the limited number of *keyword adverbs* identified in the corpus data (refer to the total key types/key tokens numbers at the bottom of the subsequent tables). The descriptions of adverbs in Section 4.5.4 provide a detailed account of the 20 most frequently occurring adverbs, presenting the data exclusively in relation to the gender factor. However, when both the academic field and gender are considered, the distribution of the data is presented in Tables 48-53.

Table 48 Frequency of adverbs used by females in all academic fields

No.	Keyword Adverb	Freq. in Comp. Statistics	%	Keyword Adverb	Freq. in Humanities	%	Keyword Adverb	Freq. in Biology &Health	%
1	internally	2	25.00%	significantly	3	6.82%	significantly	9	15.25%
2	prior	2	25.00%	approximately	3	6.82%	positively	7	11.86%
3	likewise	2	25.00%	finally	2	6.82%	negatively	6	10.17%
4	traditionally	1	12.50%	previously	2	4.55%	potentially	5	8.47%
5	automatically	1	12.50%	primarily	2	4.55%	statistically	4	6.78%
6				traditionally	2	4.55%	prior	3	5.08%
7				prior	2	4.55%	randomly	3	5.08%
8				accurately	2	4.55%	normally	2	3.39%
9				globally	2	4.55%	thereby	2	3.39%
10				radically	2	4.55%	periodically	1	3.39%
11				consistently	1	4.55%	specifically	1	1.69%
12				specifically	1	2.27%	previously	1	1.69%
13				structurally	1	2.27%	primarily	1	1.69%
14				theoretically	1	2.27%	strategically	1	1.69%

15		appropriately	1	2.27%	traditionally	1	1.69%	
16		inappropriately	1	2.27%	constantly	1	1.69%	
17		normally	1	2.27%	physically	1	1.69%	
18		potentially	1	2.27%	sufficiently	1	1.69%	
19		conventionally	1	2.27%	adequately	1	1.69%	
20		exclusively	1	2.27%	apparently	1	1.69%	
Total key types/ key tokens	5/8		29/44			26/59		

Table 48 presents examples of adverbs utilized by the F authors across the academic fields in the corpus. Firstly, two adverbs are present in all fields: *prior* and *traditionally*. *Prior* ranks second, seventh, and sixth in the CS, H, and BH lists, respectively, constituting 25%, 4.55%, and 5.08%. The second adverb, *traditionally*, occupies the fourth (12.5%), sixth (4.55%), and 15th (1.69%) positions. Secondly, six key type adverbs are present in both the H and BH lists, namely, *significantly*, *previously*, *primarily*, *specifically*, *normally*, and *potentially*. *Significantly* ranks first in both fields, with 6.82% in H and 15.25% in BH. The second adverb, *previously*, ranks fourth in H and 12th in BH, with percentages of 4.55% and 1.69%, respectively. The same frequency has been observed for *primarily*, which ranks fifth and 13th. *Specifically* accounts for 2.27% in H (12th position) and 1.69% in BH (11th position). *Normally* ranks 17th in H (2.27%) and eighth in BH (3.39%). Finally, the last adverb, *potentially*, has the same percentage as *normally* in the H field (2.27%, 18th position), while it ranks fourth in the BH wordlist (8.47%).

Table 49 Frequency of adverbs used by males in all academic fields

No.	Keyword Adverb	Freq. in Comp. Statistics	%	Keyword Adverb	Freq. in Humanities	%	Keyword Adverb	Freq. in Biology &Health	%
1	previously	3	14.29%	specifically	5	7.81%	significantly	3	7.32%
2	automatically	3	14.29%	significantly	4	6.25%	specifically	3	7.32%
3	normally	2	9.52%	hence	3	6.25%	positively	3	7.32%
4	reliably	2	9.52%	qualitatively	3	4.69%	potentially	3	7.32%
5	subsequently	2	9.52%	theoretically	2	4.69%	primarily	3	7.32%
6	interactively	1	4.76%	previously	2	3.13%	negatively	3	7.32%
7	adequately	1	4.76%	primarily	2	3.13%	consistently	2	4.88%
8	prior	1	4.76%	constantly	2	3.13%	thereby	2	4.88%
9	statistically	1	4.76%	exclusively	2	3.13%	similarly	1	2.44%
10	successively	1	4.76%	technologically	2	3.13%	structurally	1	2.44%
11	predominantly	1	4.76%	adequately	2	3.13%	finally	1	2.44%
12	randomly	1	4.76%	internally	2	3.13%	previously	1	2.44%
13	visually	1	4.76%	prior	2	3.13%	physically	1	2.44%
14	qualitatively	1	4.76%	subsequently	2	3.13%	insufficiently	1	2.44%
15				ultimately	2	3.13%	adequately	1	2.44%
16				eventually	2	3.13%	apparently	1	2.44%

17		creatively	1	3.13%	hence	1	2.44%
18		finally	1	1.56%	prior	1	2.44%
19		normally	1	1.56%	statistically	1	2.44%
20		positively	1	1.56%	objectively	1	2.44%
Total key types/ key tokens	14/21		41/64			27/41	

Table 49 illustrates the most frequently used adverbs used by the M authors in the corpus. Firstly, three adverbs are commonly present on all lists: *previously*, *adequately*, and *prior*. The first of these, *prior*, holds the first (CS), sixth (H), and 12th (BH) positions, with percentages of 14.29%, 3.13%, and 2.44%, respectively. *Adequately* occupies the seventh (4.76%), 11th (3.13%), and 15th (2.44%) positions. The third adverb, *prior*, is listed as eighth in CS, 13th in H, and 18th in BH sub-corpora, with percentages of 4.76%, 3.13%, and 2.44%, respectively. Secondly, three adverbs are present in the CS and H sub-corpora: *normally*, *subsequently*, and *qualitatively*. *Normally* ranks third in CS (9.52%) and 15th in H (1.55%). *Subsequently* is listed as fifth (9.52%) in CS and 14th (3.13%) in H. The last of the three, *qualitatively*, occupies the 14th position in CS (4.76%) and the fourth position in H (4.69%). Thirdly, the number of the same key type adverbs present on the H and BH wordlists is the highest. These include *specifically*, *significantly*, *primarily*, *adequately*, *finally*, and *positively*. *Specifically* and *significantly* are the most frequently used in both fields, occupying the first (7.81%) and second (6.25%) positions in the H sub-corpus. In the BH sub-corpora, the figures for *specifically*, *significantly*, *positively*, *potentially*, *primarily*, and *negatively* are all 7.32%, allowing them to be listed equally as first. *Primarily* holds the seventh position in the H sub-corpus (3.13%) and the fifth (or first) position in the BH sub-corpus (7.32%). *Adequately* ranks 11th in H (3.13%) and 15th in BH (2.44%) of the total examples in the sub-corpora. The last two adverbs, *finally* and *positively*, occupy the 18th and 20th positions in the H field each constituting 1.56%., In the BH field, they are ranked 11th (2.44%) and third (or first, with 7.32%). Lastly, one key type adverb is present on the CS and BH wordlists, namely, *statistically*. It occupies the ninth (4.76%) and 19th (2.44%) positions, respectively.

Table 50 Frequency of adverbs used by native females in all academic fields

No.	Keyword Adverb	Freq. in Comp. Statistics	%	Keyword Adverb	Freq. in Humanities	%	Keyword Adverb	Freq. in Biology &Health	%
1	traditionally	1	20.00%	significantly	3	15.00%	positively	6	15.79%
2	internally	1	20.00%	approximately	2	10.00%	significantly	5	13.16%
3	prior	1	20.00%	prior	2	10.00%	negatively	4	10.53%
4	automatically	1	20.00%	consistently	1	10.00%	statistically	4	10.53%
5	likewise	1	20.00%	structurally	1	5.00%	prior	3	7.89%

6		theoretically	1	5.00%	specifically	1	2.63%
7		previously	1	5.00%	normally	1	2.63%
8		primarily	1	5.00%	potentially	1	2.63%
9		traditionally	1	5.00%	primarily	1	2.63%
10		conventionally	1	5.00%	strategically	1	2.63%
11		exclusively	1	5.00%	traditionally	1	2.63%
12		adequately	1	5.00%	physically	1	2.63%
13		hence	1	5.00%	sufficiently	1	2.63%
14		accurately	1	5.00%	adequately	1	2.63%
15		globally	1	5.00%	marginally	1	2.63%
16		radically	1	5.00%	plus	1	2.63%
17					randomly	1	2.63%
18					thematically	1	2.63%
19					thereby	1	2.63%
20					visually	1	2.63%
Total key types/ key tokens	5/5		16/20			21/38	

Table 50 displays the adverbs most frequently used in the NF abstracts across the various academic fields studied. There are two groups of adverbs: those present in all fields and those present only in the H and BH fields. The first group consists of two adverbs: *traditionally* and *prior*. The percentage of the first adverb, *traditionally*, is 20%, 5%, and 2.63%, respectively, in CS (first position), H (ninth position), and BH (11th position). The second adverb, *prior*, holds the third position in both the CS and H lists, with percentages of 20% and 10%, respectively. In the BH field, *prior* accounts for 7.89% of the examples and occupies the fifth position on the list. The second group of adverbs, common to the H and BH lists, includes three key type adverbs: *significantly*, *primarily*, and *adequately*. *Significantly* ranks first in H (15%) and second in BH (13.16%). *Primarily* occupies the eighth position in H (5%) and ninth position in BH (2.63%). Lastly, *adequately* is listed as 12th in H (5%) and 14th in BH (2.63%).

Table 51 Frequency of adverbs used by native males in all academic fields

No.	Keyword Adverb	Freq. in Comp. Statistics	%	Keyword Adverb	Freq. in Humanities	%	Keyword Adverb	Freq. in Biology &Health	%
1	previously	2	16.67%	significantly	3	9.09%	negatively	3	13.04%
2	reliably	2	16.67%	specifically	3	9.09%	consistently	2	8.70%
3	prior	1	8.33%	theoretically	2	9.09%	significantly	2	8.70%
4	subsequently	1	8.33%	ultimately	2	6.06%	specifically	2	8.70%
5	successively	1	8.33%	qualitatively	2	6.06%	positively	2	8.70%
6	automatically	1	8.33%	finally	1	6.06%	similarly	1	4.35%
7	predominantly	1	8.33%	normally	1	3.03%	potentially	1	4.35%
8	randomly	1	8.33%	positively	1	3.03%	primarily	1	4.35%
9	visually	1	8.33%	potentially	1	3.03%	physically	1	4.35%

10	qualitatively	1	8.33%	previously	1	3.03%	insufficiently	1	4.35%
11				primarily	1	3.03%	adequately	1	4.35%
12				alternatively	1	3.03%	prior	1	4.35%
13				considerably	1	3.03%	objectively	1	4.35%
14				constantly	1	3.03%	dramatically	1	4.35%
15				physically	1	3.03%	predominantly	1	4.35%
16				technologically	1	3.03%	randomly	1	4.35%
17				adequately	1	3.03%	intrinsically	1	4.35%
18				approximately	1	3.03%			
19				hence	1	3.03%			
20				prior	1	3.03%			
Total key types/ key tokens		10/12			26/33			17/23	

Table 51 lists the most frequently used adverbs by the NM authors in the researched academic fields. Firstly, one adverb is present on all academic field lists, namely, *prior*. It occupies the third, 20th, and 12th positions, with percentages of 8.33%, 3.03%, and 4.35%, respectively, in the CS, H, and BH sub-corpora. Secondly, one adverb (*predominantly*) is commonly used in abstracts within the CS (seventh position, 8.33%) and BH (15th position, 4.35%) fields. Thirdly, two adverbs, *previously* and *qualitative*, are present on the CS and H wordlists. *Previously* occupies the first and tenth positions, respectively, with 16.67% in CS and 3.03% in H sub-corpora. In contrast, the second adverb, *qualitative*, is listed as fifth in H (6.06%) and tenth in CS (8.33%) sub-corpora. Finally, seven key type adverbs are present on the H and BH lists: *significantly*, *specifically*, *positively*, *potentially*, *primarily*, *physically*, and *adequately*. In H, they occupy the first and second positions (both 9.09%), as well as the eighth, ninth, 11th, 15th, and 17th positions (3.03% each). In BH, these adverbs occupy the third, fourth, and fifth positions (each at 8.70%), as well as the seventh, eighth, ninth, and 11th positions (each at 4.35%).

Table 52 Frequency of adverbs used by non-native females in all academic fields

No.	Keyword Adverb	Freq. in Comp. Statistics	%	Keyword Adverb	Freq. in Humanities	%	Keyword Adverb	Freq. in Biology &Health	%
1	internally	1	33.33%	finally	2	10.00%	significantly	4	21.05%
2	prior	1	33.33%	specifically	1	5.00%	potentially	4	21.05%
3	likewise	1	33.33%	appropriately	1	5.00%	negatively	2	10.53%
4				inappropriately	1	5.00%	randomly	2	10.53%
5				normally	1	5.00%	periodically	1	5.26%
6				potentially	1	5.00%	normally	1	5.26%
7				previously	1	5.00%	positively	1	5.26%
8				primarily	1	5.00%	previously	1	5.26%
9				traditionally	1	5.00%	constantly	1	5.26%
10				negatively	1	5.00%	apparently	1	5.26%

11		approximately	1	5.00%	thereby	1	5.26%
12		professionally	1	5.00%			
13		accurately	1	5.00%			
14		globally	1	5.00%			
15		intensively	1	5.00%			
16		radically	1	5.00%			
17		thematically	1	5.00%			
18		thereby	1	5.00%			
19		virtually	1	5.00%			
20							
Total key types/ key tokens	3/3		19/20			11/19	

Table 52 presents examples of adverbs used in the NNF abstracts. Similar adverbs are observed only in the H and BH lists. The adverbs commonly present in these two academic fields are *normally*, *potentially*, *previously*, and *negatively*. The first adverb, *normally*, occupies the fifth and sixth positions, with percentages of 5% and 5.26%, respectively. The second adverb, *potentially*, is listed as second in BH (21.05%) and sixth in H (5%) sub-corpora. The third adverb, *previously*, occupies the seventh position in H (5%) and the eighth position in BH (5.26%). The last adverb, *negatively*, constitutes 5% of the total examples in H (in the tenth position) and 10.53% in BH (third position) sub-corpora.

Table 53 Frequency of adverbs used by non-native males in all academic fields

No.	Keyword Adverb	Freq. in Comp. Statistics	%	Keyword Adverb	Freq. in Humanities	%	Keyword Adverb	Freq. in Biology &Health	%
1	normally	2	22.22%	specifically	2	6.45%	potentially	2	11.76%
2	automatically	2	22.22%	exclusively	2	6.45%	primarily	2	11.76%
3	previously	1	11.11%	hence	2	6.45%	thereby	2	11.76%
4	interactively	1	11.11%	internally	2	6.45%	significantly	1	5.88%
5	adequately	1	11.11%	eventually	2	6.45%	specifically	1	5.88%
6	statistically	1	11.11%	creatively	1	6.45%	structurally	1	5.88%
7	subsequently	1	11.11%	significantly	1	3.23%	finally	1	5.88%
8				previously	1	3.23%	positively	1	5.88%
9				primarily	1	3.23%	previously	1	5.88%
10				regionally	1	3.23%	apparently	1	5.88%
11				constantly	1	3.23%	hence	1	5.88%
12				sufficiently	1	3.23%	statistically	1	5.88%
13				technologically	1	3.23%	ultimately	1	5.88%
14				adequately	1	3.23%	qualitatively	1	5.88%
15				prior	1	3.23%			
16				professionally	1	3.23%			
17				statistically	1	3.23%			
18				subsequently	1	3.23%			

19			overseas	1	3.23%		
20			somewhat	1	3.23%		
Total key types/ key tokens	7/9			26/31		14/17	

Table 53 demonstrates the adverbs present in the NNM abstracts within the corpus. Firstly, two adverbs are present on all lists: *previously* and *statistically*. *Previously* occupies the third, eighth, and ninth positions, with percentages of 11.11% in CS, 3.23% in H, and 5.88% in BH sub-corpora. *Statistically*, the second adverb, occupies the sixth, 17th, and 12th positions, with percentages of 11.11% in CS, 3.23% in H, and 5.88% in BH, respectively. Secondly, one adverb, *subsequently*, is present on the CS and BH lists. It occupies the seventh and 18th positions, with 11.11% and 3.23%, respectively. Lastly, four key type adverbs are present in the H and BH sub-corpora: *primarily*, *significantly*, *specifically*, and *hence*. In the BH field, they occupy the second (11.76%), fourth, fifth, and 12th positions (each at 3.23%). In Humanities, they are listed as the ninth and seventh positions (both at 3.23%), and the first and third positions (both at 6.45%).

Table 53 concludes the first segment of the results presentation, specifically the AWL vocabulary section of this corpus study. The subsequent sections detail the outcomes and results derived from the computational analysis of academic language functions. Section 4.7 provides an overview of the computational software employed for this analysis. Following this, Section 4.8 presents an excerpt from the corpus, comprising selected example sentences that form part of the data used as a basis for software labeling, along with definitions of each function and the methodological elements considered during the data labeling process. The quantitative analysis of academic language functions in conference abstracts, along with Pearson's Chi-squared Tests, is subsequently discussed in Section 4.9. The final section presents all conclusions drawn from the study.

4.7 AntMover software

It is a freeware software program that analyzes text structures and labels corpus data. It is easy to use, as all its functions are intuitively placed in the program's graphical user interface. As Anthony (2016) claims, "AntMover is also designed to be a completely general learning system and so knowledgeable users can modify almost all aspects of the system to suit their particular needs." (p. 1). This vital feature is possible after some adjustments to the program files. Anthony (2016) underlines that the AntMover program uses PERL 5.6.1 programming language and modules available in Comprehensive Perl Archive Network (CPAN).

Figure 21 AntMover software

After changes are made in the corpus data files and folders installation in the AntMover directory (Anthony, 2016), the program can be used to suit the researchers' needs. It can, for example, be used to label the data according to previously prepared functions.

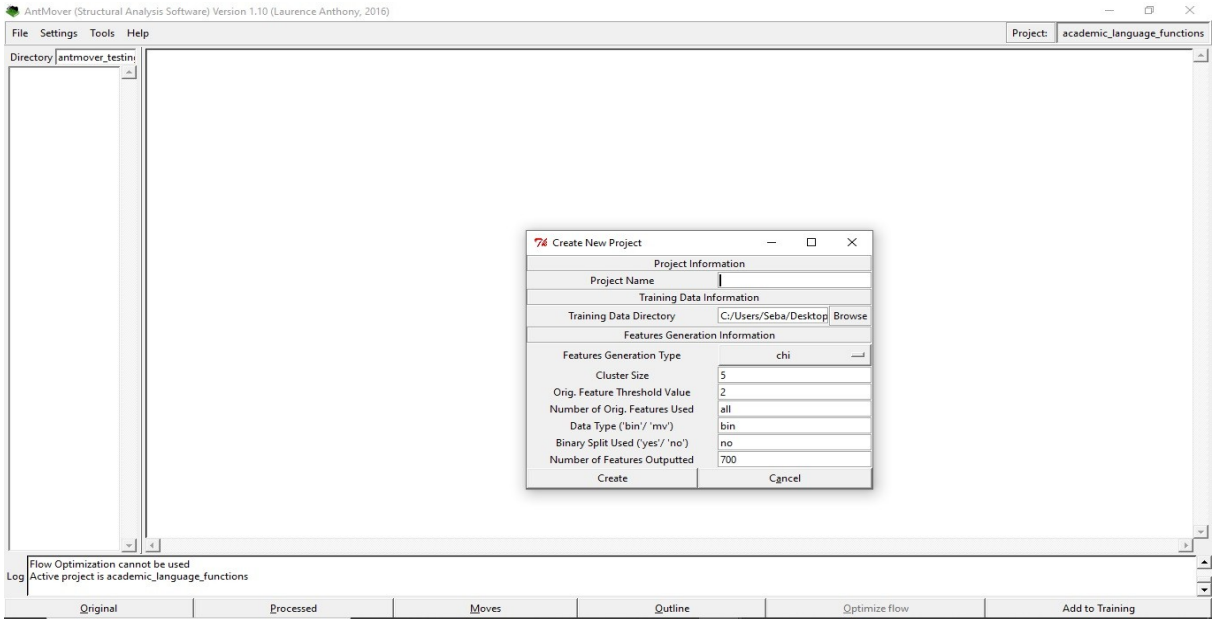

Figure 22 AntMover Project creation window

A new project can be created using the following program options: features generation type, cluster size, original feature threshold value, number of original features used, data type; one can enable or disable the binary split option, or set the number of maximum feature output. *The File>Open File* or *File>Open Directory* option can load the data stored in the .txt extension (UTF-8 coded). Afterwards, data is processed into sentences and can be controlled using the program interface buttons on the bottom (see Figure 23). A specific text can be

displayed by clicking the data name on the left side window. Then the following tools can be used to view and manipulate the chosen text: *Original, Processed, Moves, Outline, Optimize flow,* and *Add to Training*.

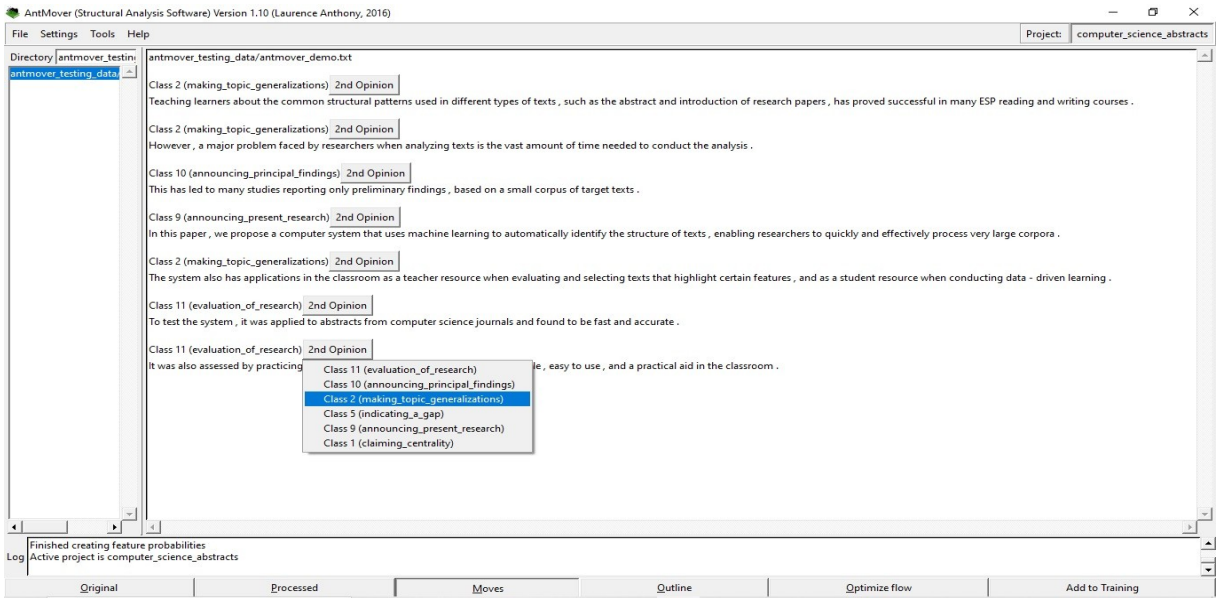

Figure 23 AntMover text labeling in Moves window

A label of any of the sentences in a viewed text can be changed by clicking the *second option* button in the *Moves* window and choosing it from the previously prepared list. The labeled text from *Moves* or *Outline* sheets can be saved as a .txt file in *File>Save Output* option. The *Optimize flow* option can be used after retraining the AntMover system, having enough training examples in the database (Anthony, 2016).

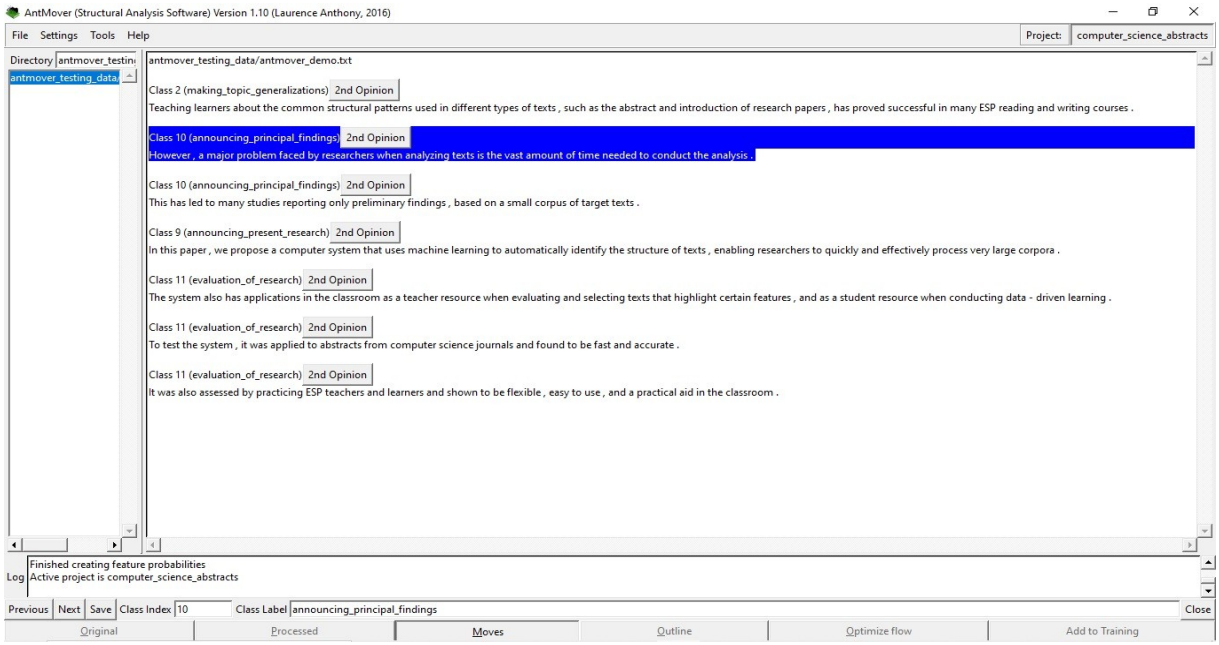

Figure 24 AntMover Add to Training window

One can retrain the system by changing the training data files in the AntMover folder or adding specific sentence examples in the *Add to Training* option (see Figure 24). Moreover, additional new labels may be added through *Class Index* and *Class Label* fields at the bottom of the *Add to Training* window.

The AntMover system was retrained for the purposes of this research. The naming of the labels was changed and new label categories were added according to Table 7, based on Martin (1976). The outcome of AntMover software labeling, saved as .txt files, has the form of two lines. In the first line the stored information comprises class, number, function name (given in brackets); in the second line the example sentence is displayed.

The data coded manually was used to form a basis (training data), which then helped to find corresponding, matching sentences to each studied function. This procedure was done having in mind the most frequently used academic vocabulary found in the corpus and presented in the previous sections. The following section presents the excerpt from the corpus, viz., several examples of the sentences used as a basis for the *training data* divided into sections, corresponding to the studied academic language functions.

4.8 Function examples (excerpt from the corpus)

As previously mentioned, the labeling system is based on Martin (1976), who presented three sets of academic vocabulary: *academic vocabulary of the research process*, *evaluation*, and *analysis*. Each of these sets can be utilized in academic settings, such as teaching students, and can be related to specific actions undertaken while communicating within academia. Each set is associated with particular parts of speech, namely, the vocabulary of the research process is linked with nouns and verbs, the vocabulary of evaluation with adjectives and adverbs, and the vocabulary of analysis with two-word verbs, high-frequency verbs, and nouns.

Based on these three sets and the parts of speech represented in each, academic language functions can be formed. In the first set, driven by nouns and verbs, one can identify functional elements such as *formulating*, *investigating*, *presenting methodology*, *designing experiments*, *simulating situations/developing models*, *reporting results*, and *drawing conclusions*. The second set includes *reviewing*, *evaluating*, and *criticizing*, all functions linked with adjectives and adverbs. The third set, associated with two-word verbs, high-frequency verbs, and nouns, encompasses functions such as *organizing the order of information* and *reinforcing the research process*.

As presented in Section 3.5, the AntMover program data has been modified to suit the corpus data handling appropriately. A sentence has been used as the unit of coding. 25% of the corpus has been coded manually, while 75% has been coded automatically by AntMover. The software has worked on the training data loaded into the program, i.e., the 25% manually coded examples. The automatically coded data has been manually rechecked afterwards.

Finding definitions for academic language functions has proven challenging. Many scholars refer to them as commonly understood elements, listing various names of academic language functions (e.g., Snow and Uccelli, 2009; Lin, 2016; Heinke, 2018). These functions are also referred to as *cognitive cross-content language functions* (Mohan, 1986), *cognitive strategies* (O'Malley & Chamot, 1990), *cognitive dimension components* (Scarcella, 2003), or *cognitive discourse functions* (Dalton-Puffer, 2013). Others provide similar lists of functions, labeling them according to various theoretical frameworks (O'Malley & Chamot, 1994; Herrel & Jordan, 2001; Chamot & Robbins, 2005; Clyne, 2006; District-Wide Academic Support Teams, 2010). None of the aforementioned studies can be directly compared to Martin's (1976) framework. Furthermore, Martin does not define the functional elements as separate units. Therefore, the following sections include dictionary definitions at the beginning of each section, along with the researcher's definitions for each academic language function presented.

The following sections present excerpts from the sub-corpus of labeled sentences. Each Class corresponds to the number of academic language functions in this study (see Table 7). The function names are additionally marked by their abbreviated versions in brackets next to the Class. Each example sentence presented below is part of the studied conference abstracts, identified by a .txt number, illustrating the use of the respective academic function examples in the given text. All tagged material presented in the sections below employs the labeling of the TreeTagger Tag Set and was prepared using TagAnt software (see Anthony, 2015).

4.8.1 **Formulating**

The verb *formulate* can be defined as follows: "to create something" (Cambridge Dictionary), "to put into a systematized statement of expression" (Merriam-Webster Dictionary), or "to express in a precise form, state definitely or systematically" (Dictionary.com). Based on this definitions, I propose the following definition for the F1 *Formulating* function: it can be defined as as the methodical creation or preparation, and systematic expression, of statements that describe the elements of a study, such as its focus or the articulation of what the study is about, utilizing academic language.

The methodological aspect of labeling the F1 *Formulating* function involved investigating how F and M authors construct coherent and structured arguments, hypotheses, or theoretical frameworks within their abstracts. The key elements examined included identifying statements that outline the main argument or hypothesis of the text, using phrases such as *this study argues*, *we propose*, *the hypothesis is*, or *our main argument*. Attention was also given to the organization of ideas, focusing on the use of headings, subheadings, and paragraph structures. Additionally, elements such as building arguments with supporting evidence, examples, and logical reasoning were also taken into consideration.

Sections 4.5.1, 4.5.2, 4.6.1 and 4.6.2 presented the vocabulary that was accounted for while tagging the F1 *Formulating* function. Some of the nouns and verbs found in the examples realizing this function are *study, package, classroom, paper, project, talk; addresses, allows, approach, examines, is, was, create, develop,* and *discuss.*

The excerpt from the corpus is presented below; each example sentence is followed by its tagged version.

Class 1 (formulating):

- "In this talk **we will discuss some examples** of how Shiny has been used at Allstate to empower business users and create an organizational appetite for data science." (36.txt)
  In_IN this_DT talk_NN we_PP will_MD discuss_VV some_DT examples_NNS of_IN how_WRB Shiny_NP has_VHZ been_VBN used_VVN at_IN Allstate_NP to_TO empower_VV business_NN users_NNS and_CC create_VV an_DT organizational_JJ appetite_NN for_IN data_NN science_NN ._SENT
- "**We show how** the OpenML **package allows** R users to easily search , download and upload machine learning datasets ." (72.txt)
  We_PP show_VVP how_WRB the_DT OpenML_NP package_NN allows_VVZ R_NN users_NNS to_TO easily_RB search_VV ,_, download_NN and_CC upload_NN machine_NN learning_VVG datasets_NNS ._SENT
- "**We examined the species composition** of Gambierdiscus in Japanese surface coastal waters and at sites below 15 m using a quantitative PCR assay ." (484.txt)
  We_PP examined_VVD the_DT species_NN composition_NN of_IN Gambierdiscus_NP in_IN Japanese_JJ surface_NN coastal_JJ waters_NNS and_CC at_IN sites_NNS below_IN 15_CD m_NN using_VVG a_DT quantitative_JJ PCR_NP assay_NN ._SENT
- "**The key aim of this project is** to incentivise higher education students to engage with a Year 1 Computer Systems module through a blended and flipped classroom approach ." (104.txt)
  The_DT key_JJ aim_NN of_IN this_DT project_NN is_VBZ to_TO incentivise_VV higher_JJR education_NN students_NNS to_TO engage_VV with_IN a_DT Year_NN 1_CD Computer_NP Systems_NPS module_NN through_IN a_DT blended_VVN and_CC flipped_VVN classroom_NN approach_NN ._SENT
- "**The purpose of this study was** to examine the acute effects of a reduced sitting day on executive function in pre - school - aged children ." (253.txt)
  The_DT purpose_NN of_IN this_DT study_NN was_VBD to_TO examine_VV the_DT acute_JJ effects_NNS of_IN a_DT reduced_VVN sitting_VVG day_NN on_IN executive_JJ function_NN in_IN pre_NN -_: school_NN -_: aged_VVN children_NNS ._SENT
- "**The aim of the** joint **research project** Capital4Health - " " Capabilities for active lifestyles " " **is** to develop capabilities in different population groups and for professionals ." (294.txt)
  The_DT aim_NN of_IN the_DT joint_JJ research_NN project_NN Capital4Health_NN -_: "_" "_`` Capabilities_NP for_IN active_JJ lifestyles_NNS "_" "_" is_VBZ to_TO develop_VV capabilities_NNS in_IN different_JJ population_NN groups_NNS and_CC for_IN professionals_NNS ._SENT
- **"The aim of the paper is** to document and explain the variation found in the service encounter openings ." (332.txt)

The_DT aim_NN of_IN the_DT paper_NN is_VBZ to_TO document_VV and_CC explain_VV the_DT variation_NN found_VVD in_IN the_DT service_NN encounter_NN openings_NNS ._SENT

- "**This study addresses the challenge** of inclusion in mainstream schools of learners with developmental and attention deficits and **examines the potential** of a digital structuring tool , MobilizeMe , to scaffold this process , including the impact and implications associated with the implementation ." (170.txt)
  This_DT study_NN addresses_VVZ the_DT challenge_NN of_IN inclusion_NN in_IN mainstream_JJ schools_NNS of_IN learners_NNS with_IN developmental_JJ and_CC attention_NN deficits_NNS and_CC examines_VVZ the_DT potential_NN of_IN a_DT digital_JJ structuring_NN tool_NN ,_, MobilizeMe_NP ,_, to_TO scaffold_VV this_DT process_NN ,_, including_VVG the_DT impact_NN and_CC implications_NNS associated_VVN with_IN the_DT implementation_NN ._SENT
- "Taking into account teachers' concerns from these studies , as well as drawing from concrete examples from a newly designed tertiary course in French studies in Australia , **this paper considers** what a politically engaged pedagogy involves from a teacher's perspective in a tertiary context ." (340.txt)
  Taking_VVG into_IN account_NN teachers'_NN concerns_NNS from_IN these_DT studies_NNS ,_, as_RB well_RB as_IN drawing_VVG from_IN concrete_JJ examples_NNS from_IN a_DT newly_RB designed_VVN tertiary_JJ course_NN in_IN French_JJ studies_NNS in_IN Australia_NP ,_, this_DT paper_NN considers_VVZ what_WP a_DT politically_RB engaged_VVN pedagogy_NN involves_VVZ from_IN a_DT teacher's_JJ perspective_NN in_IN a_DT tertiary_JJ context_NN ._SENT
- "**In this study** , **allelopathic interactions** between Skeletonema costatum and Alexandrium minutum **were investigated** by bi - algal culture , filtrate culture in f / 2 medium ." (433.txt)
  In_IN this_DT study_NN ,_, allelopathic_JJ interactions_NNS between_IN Skeletonema_NP costatum_NP and_CC Alexandrium_NP minutum_NN were_VBD investigated_VVN by_IN bi_NNS -_: algal_JJ culture_NN ,_, filtrate_NN culture_NN in_IN f_NN /_SYM 2_CD medium_NN ._SENT

4.8.2 **Investigating**

The verb *investigate* can be defined as follows: "to observe or study by close examination and systematic inquiry" (Merriam-Webster Dictionary) or "to examine something carefully" (Cambridge Dictionary). Based on these definitions, I propose the following definition for the F2 *Investigating* function: it relates to the systematic process of inquiry aimed at discovering information about a particular subject. This function can be associated with activities such as investigating specific hypotheses, assessing the impact of various factors, identifying gaps in knowledge, conducting individual case studies, analyzing data sets to uncover trends or patterns, or examining theoretical frameworks.

The methodological approach to labeling the F2 function involved examining how F and M authors describe their research processes, methodologies, and investigative procedures. This entailed searching for technical terms specific to the field of study and evaluating how these terms are defined and explained within the text. Additionally, reflections on the research process, including justifications for methodological choices and discussions of potential limitations or challenges, were noted. Further elements examined included detailed descriptions of research methods and procedures, the chronological or logical order in which the investigative process was presented, and language that indicated steps or stages in the research.

Sections 4.5.1, 4.5.2, 4.6.1 and 4.6.2 presented the vocabulary which was taken into

consideration while tagging the F2 *Investigating* function. In the second function, the following nouns and verbs can be observed in the excerpt: *scientists, users, quality control, data processing, data quality, courses, trend, development, debate, success rates, students, games, research, certain expression, case studies, language teachers, political engagement, blooms, austral season, summer; be, is, ensure, are based, are reported, have been conducted, delivering, does, ignores, has suggested, gauge, have shown, have, have been reposted, was,* and *is found.* The example sentences from the corpus, realizing the F2 *Investigating* function, are presented in the excerpt below.

Class 2 (investigating)

- "Relationships between **data scientists** and **business users can** often **be** very transactional in nature (i . e . give us some data and we'll give you a solution) ." (36.txt)
  Relationships_NNS between_IN data_NN scientists_NNS and_CC business_NN users_NNS can_MD often_RB be_VB very_RB transactional_JJ in_IN nature_NN (_(i_NP ._SENT e_LS ._SENT give_VV us_PP some_DT data_NN and_CC we'll_NN give_VVP you_PP a_DT solution_NN)_) ._SENT
- "While automated **quality control is a key component** of modern **data processing** workflows , visual review by a trained eye can further **ensure data quality** ." (54.txt)
  While_IN automated_JJ quality_NN control_NN is_VBZ a_DT key_JJ component_NN of_IN modern_JJ data_NNS processing_VVG workflows_NNS ,_, visual_JJ review_NN by_IN a_DT trained_JJ eye_NN can_MD further_RBR ensure_VV data_NNS quality_NN ._SENT
- "Massive Open **Online Courses** (MOOCs) , which **are based** in approach that offers online courses to great masses in an open and free manner , **are a** recent **trend in** distance **education** and are still under debate ." (149.txt)
  Massive_JJ Open_NP Online_NP Courses_NP (_(MOOCs_NP)_) ,_, which_WDT are_VBP based_VVN in_IN approach_NN that_WDT offers_VVZ online_JJ courses_NNS to_TO great_JJ masses_NNS in_IN an_DT open_JJ and_CC free_JJ manner_NN ,_, are_VBP a_DT recent_JJ trend_NN in_IN distance_NN education_NN and_CC are_VBP still_RB under_IN debate_NN ._SENT "_NN
- "Student **success rates are** widely **reported to be** lower for part - time than full - time students , and lower for OEDL than for part - time students as a whole ." (133.txt)
  Student_NN success_NN rates_NNS are_VBP widely_RB reported_VVN to_TO be_VB lower_JJR for_IN part_NN -_: time_NN than_IN full_JJ -_: time_NN students_NNS ,_, and_CC lower_JJR for_IN OEDL_NP than_IN for_IN part_NN -_: time_NN students_NNS as_IN a_DT whole_NN ._SENT
- "As a result , a very limited number of cost effectiveness **analyses have been conducted** ." (256.txt)
  As_IN a_DT result_NN ,_, a_DT very_RB limited_JJ number_NN of_IN cost_NN effectiveness_NN analyses_NNS have_VHP been_VBN conducted_VVN ._SENT
- "**Treating games as simple vehicles** for '**delivering a message**' **does a disservice** to play and **ignores the role of design** in their creation ." (272.txt)
  Treating_VVG games_NNS as_IN simple_JJ vehicles_NNS for_IN 'delivering_VVG a_DT message'_NN does_VVZ a_DT disservice_NN to_TO play_VV and_CC ignores_VVZ the_DT role_NN of_IN design_NN in_IN their_PP$ creation_NN ._SENT
- "Previous **research has suggested that** Finland - Swedish is overall more formal than Sweden Swedish (e . g . Norrby et al 2015) , but **it is difficult to gauge how** formal a **certain expression is considered** across **speech communities** ." (332.txt)
  Previous_JJ research_NN has_VHZ suggested_VVN that_DT Finland_NP -_: Swedish_NP is_VBZ overall_RB more_RBR formal_JJ than_IN Sweden_NP Swedish_NP (_(e_NP ._SENT g_NN ._SENT Norrby_NP et_FW al_NP 2015_CD)_) ,_, but_CC it_PP is_VBZ difficult_JJ to_TO gauge_VV how_WRB formal_JJ a_DT certain_JJ expression_NN is_VBZ considered_VVN across_IN speech_NN communities_NNS
- "However , recent **case studies** (e . g . Sercu 2006 , Johnstone Young & Sachdev 2011 , Diaz 2013) **have shown that language teachers have mitigated views** on the role **political engagement ought to have** in intercultural **language teaching and learning** ." (340.txt)

However_RB ,_, recent_JJ case_NN studies_NNS (_(e_LS ._SENT g_NN ._SENT Sercu_NP 2006_CD ,_, Johnstone_NP Young_NP &_CC Sachdev_NP 2011_CD ,_, Diaz_NP 2013_CD)_) have_VHP shown_VVN that_IN/that language_NN teachers_NNS have_VHP mitigated_VVN views_NNS on_IN the_DT role_NN political_JJ engagement_NN ought_MD to_TO have_VH in_IN intercultural_JJ language_NN teaching_NN and_CC learning_NN ._SENT

- "Dinophysis **blooms have been reported** associated **to warm austral season** since 1992 in Uruguay , but the **summer** 2015 **was exceptional for** the **development** of a Dinophysis of the acuminata complextoxic bloom the bloom initiated during a dry period where we noted the increase of salinity due the major incidence of warm waters coming from Brazil fluxing from North to South along the Uruguayan coast ." (439.txt)
  Dinophysis_NN blooms_NNS have_VHP been_VBN reported_VVN associated_VVN to_TO warm_VV austral_JJ season_NN since_IN 1992_CD in_IN Uruguay_NP ,_, but_CC the_DT summer_NN 2015_CD was_VBD exceptional_JJ for_IN the_DT development_NN of_IN a_DT Dinophysis_NN of_IN the_DT acuminata_NP complextoxic_JJ bloom_NN the_DT bloom_NN initiated_VVN during_IN a_DT dry_JJ period_NN where_WRB we_PP noted_VVD the_DT increase_NN of_IN salinity_NN due_RB the_DT major_JJ incidence_NN of_IN warm_JJ waters_NNS coming_VVG from_IN Brazil_NP fluxing_VVG from_IN North_NP to_TO South_NP along_IN the_DT Uruguayan_JJ coast_NN ._SENT
- "The dinoflagellate genus Coolia Meunier **is an important** epi - benthic **organism** that **is commonly found in association with** other dinoflagellates known to cause ciguatera ." (462.txt)
  The_DT dinoflagellate_NN genus_NN Coolia_NP Meunier_NP is_VBZ an_DT important_JJ epi_NN -_: benthic_JJ organism_NN that_WDT is_VBZ commonly_RB found_VVN in_IN association_NN with_IN other_JJ dinoflagellates_NNS known_VVN to_TO cause_VV ciguatera_NN ._SENT

4.8.3 **Analyzing**

The following three functions were treated as a group because they are constituents of the analyzing function, as presented in Chapter One (see Table 7). The verb *analyze* is defined as "to study something in a systematic and careful way" (Cambridge Dictionary), or "to study or determine the nature and relationship of the parts of (something) by analysis" (Merriam-Webster Dictionary). The F3 *Analyzing* function, in all of its parts (F3.1, F3.2, and F3.3), encompasses vocabulary linked with methodology. The most frequently used nouns and verbs can be found in Sections 4.5.1, 4.5.2, 4.6.1 and 4.6.2.

The methodological aspect of labeling the F3 functions involved examining the linguistic structures and discourse strategies used to break down and interpret information. In addition to searching for specific vocabulary, elements of sentence structure, such as cause-and-effect sentences, conditional statements, and detailed descriptions, were analyzed. Elements such as how F and M authors present arguments, draw conclusions from existing data, and highlight key findings from other works were also observed. Furthermore, the types of evidence used to support analyses, including references to studies, quotations, and examples, were considered in the process of labeling corpus data.

4.8.3.1 Presenting methodology

In order to ensure clarity and consistency, it is important to define key terms used in this section. I propose the following definition: the F3.1 *Presenting methodology* function can be

defined as the systematic and detailed description of the research methods and procedures used in a study to ensure transparency and credibility. The F3.1 *Presenting methodology* function encompasses elements such as research design, data collection methods, data analysis techniques, and the presentation of ethical considerations by authors.

The methodological aspect of labeling the F3.1 *Presenting methodology* function involved examining how F and M authors described and justified the research methods and procedures employed in their studies. This included analyzing the lexical choices and syntactic structures used to explain the rationale behind methodological decisions, the step-by-step processes followed, and the tools or techniques utilized. Specific elements examined included methodological descriptions with detailed terminology and clear procedural steps, comparisons of methodologies with previous studies or alternative approaches, and highlighting similarities, differences, and improvements.

In addition to the vocabulary presented in sections 4.5.1, 4.5.2, 4.6.1, and 4.6.2, the following nouns and verbs were identified in the excerpt below: *dashboards, datasets, profiles, forecasts, model based method, models, patients, questions, technologies, ethnography, inquiry, case study, functions, working memory, measurement, assessments, population level, data, speech act, involvement, alignment, results, performance, relationship, light intensity; developed, propose, based, detect, compute, obtain, included, used, were assessed, has been coded, affect, are compared, were conducted, was compared, was formulated.* The following excerpt presents sentences labeled with the F3.1 *Presenting methodology* function tag.

Class 31 (presenting methodology)

- "Using Shiny , **we developed data quality control dashboards** to facilitate analyst review of **ambient air quality** and **meteorological datasets** such as criteria **pollutant concentrations** , **wind profiles** , and **weather forecasts** ." (54.txt)
  Using_VVG Shiny_NP ,_, we_PP developed_VVD data_NNS quality_NN control_NN dashboards_NNS to_TO facilitate_VV analyst_NN review_NN of_IN ambient_JJ air_NN quality_NN and_CC meteorological_JJ datasets_NNS such_JJ as_IN criteria_NNS pollutant_NN concentrations_NNS ,_, wind_NN profiles_NNS ,_, and_CC weather_NN forecasts_NNS ._SENT
- **"We propose model - based** random forests as a **method** to **detect similarities** between patients with respect to their treatment effect and on this basis **compute personalized models** for new **patients** to **obtain their** individual **treatment** effect ." (78.txt)
  We_PP propose_VVP model_NN -_: based_VVN random_JJ forests_NNS as_IN a_DT method_NN to_TO detect_VV similarities_NNS between_IN patients_NNS with_IN respect_NN to_TO their_PP$ treatment_NN effect_NNS and_CC on_IN this_DT basis_NN compute_VV personalized_JJ models_NNS for_IN new_JJ patients_NNS to_TO obtain_VV their_PP$ individual_JJ treatment_NN effect_NNS ._SENT
- "These **included questions** regarding the type of **technologies used** for learning , frequency of usage , questions concerning the scope and genre used for reading and writing while utilizing different technologies , usage of assistive technology , and questions regarding technological preferences ." (113.txt)
  These_DT included_VVN questions_NNS regarding_VVG the_DT type_NN of_IN technologies_NNS used_VVN for_IN learning_NN ,_, frequency_NN of_IN usage_NN ,_, questions_NNS

concerning_VVG the_DT scope_NN and_CC genre_NN used_VVN for_IN reading_VVG and_CC writing_VVG while_IN utilizing_VVG different_JJ technologies_NNS ,_, usage_NN of_IN assistive_JJ technology_NN ,_, and_CC questions_NNS regarding_VVG technological_JJ preferences_NNS ._SENT

- "The methodology used is a combination of four different **qualitative methods** : grounded theory , **digital ethnography** , **narrative inquiry** and **case study** ." (169.txt)
  The_DT methodology_NN used_VVN is_VBZ a_DT combination_NN of_IN four_CD different_JJ qualitative_JJ methods_NNS :_: grounded_VVN theory_NN ,_, digital_JJ ethnography_NN ,_, narrative_JJ inquiry_NN and_CC case_NN study_NN ._SENT
- "Three executive **functions** (inhibition , **working memory** and shifting) **were assessed using** the Early Years Toolbox ." (253.txt)
  Three_CD executive_JJ functions_NNS (_(inhibition_NN ,_, working_VVG memory_NN and_CC shifting_JJ)_) were_VBD assessed_VVN using_VVG the_DT Early_NP Years_NPS Toolbox_NP ._SENT
- "METHODS: Review of **the diversity** of **physical activity measurement** **methods used** in **epidemiology**, **in population level** **assessment** of physical activity**, ranging from** self - report questionnaires through **to** objective **assessments of populations using accelerometers** ." (282.txt)
  METHODS_NNS :_: Review_VV of_IN the_DT diversity_NN of_IN physical_JJ activity_NN measurement_NN methods_NNS used_VVN in_IN epidemiology_NN ,_, in_IN population_NN level_NN assessment_NN of_IN physical_JJ activity_NN ,_, ranging_VVG from_IN self_NN -_: report_NN questionnaires_NNS through_IN to_TO objective_JJ assessments_NNS of_IN populations_NNS using_VVG accelerometers_NNS ._SENT
- "The **data has been coded** according to **speech act** , clusivity , and **stance** , using Kiesling's (2004) three dimensions of stance : **affect** , **involvement** and **alignment** , and the **results are compared** with existing **work on the mitigation** of directives in **leadership discourse** (e . g . Takano 2005 ; Vine 2009) ." (328.txt)
  The_DT data_NN has_VHZ been_VBN coded_VVN according_VVG to_TO speech_NN act_NN ,_, clusivity_NN ,_, and_CC stance_NN ,_, using_VVG Kiesling's_NP (_(2004_CD)_) three_CD dimensions_NNS of_IN stance_NN :_: affect_VV ,_, involvement_NN and_CC alignment_NN ,_, and_CC the_DT results_NNS are_VBP compared_VVN with_IN existing_VVG work_NN on_IN the_DT mitigation_NN of_IN directives_NNS in_IN leadership_NN discourse_NN (_(e_LS ._SENT g_NN ._SENT Takano_NP 2005_CD ;_: Vine_NP 2009_CD)_) ._SENT
- "Semi - structured **interviews were conducted** with ten Chinese - English **bilingual families** ."(337.txt)
  Semi_NN -_: structured_VVN interviews_NNS were_VBD conducted_VVN with_IN ten_CD Chinese_NN -_: English_JJ bilingual_JJ families_NNS ._SENT
- "The **performance of four** commercially available **PST test kits** , Abraxis™ , Europroxima™ , Scotia™ and Neogen™ , **was compared** on contaminated mussels and oysters ." (463.txt)
  The_DT performance_NN of_IN four_CD commercially_RB available_JJ PST_NP test_NN kits_NNS ,_, Abraxis™_NP ,_, Europroxima™_NP ,_, Scotia™_NP and_CC Neogen™_NP ,_, was_VBD compared_VVN on_IN contaminated_JJ mussels_NNS and_CC oysters_NNS ._SENT
- "The **relationship** between the **observed light intensity** and **depth** **was formulated using** Beer's Law ." (484.txt)
  The_DT relationship_NN between_IN the_DT observed_JJ light_JJ intensity_NN and_CC depth_NN was_VBD formulated_VVN using_VVG Beer's_NP Law_NP ._SENT

4.8.3.2 Design the experiment

To ensure clarity and precision, it is important to define key concepts used in this section. I propose the following definition: the F3.2 *Design the experiment* function refers to the process of preparing, planning, and structuring an experiment according to specific hypotheses and aims to test them. This function involves several steps, such as describing and defining variables, formulating hypotheses, designing experimental conditions (i.e., how to manipulate the variables), and determining measures for the outcomes of a study.

The methodological aspect of labeling the F3.2 *Design the experiment* function

involved examining how F and M authors outline their experimental plans and procedures, including formulating hypotheses, selecting variables, and establishing experimental controls. The elements examined included the language used to articulate research hypotheses, the terminology and structures used to define variables, clear and specific descriptions of experimental design, and detailed descriptions of experimental procedures. This analysis also considered the use of step-by-step instructions, chronological ordering, and operational definitions.

The most frequently occurring nouns and verbs considered while labeling the F3.2 *Design the experiment* function were presented in sections 4.5.1, 4.5.2, 4.6.1, and 4.6.2. Additionally, the following nouns and verbs were identified in the excerpt below: *article, sections, requirements, organization, situation, barriers, definition, research action, university teachers, agents, education, construction, self-assessment, methodology, participants, control group, exposure, data, focus groups, encounters, locations, perspective, choices, activities, strain, objectives, variables; is divided, use, developing, can be overcome, present, discuss, have developed, aimed to, explore, is described, were randomized, had, were, video-recorded, focuses on, can pervade, was grown, tested,* and *sampled.*

Class 32 (design the experiment)

- **"The article is divided** into theoretical **sections** and applied sections ." (62.txt)
- "**It will** then **use three projects** to show how , by understanding the **requirements of the organisation**, and **developing situation** - specific roll - out **strategies** , these **barriers** to entry **can be overcome** ." (66.txt)
- "**We present and discuss the definition** of Open Educator that **we have developed** in the frame of the Open Educators Factory project , a **research action aimed to explore** how to transform **university teachers** from **agents of resistance** into **agents of change** for openness in **education** ." (140.txt)
- "In a **second section** , the **construction of a contextualized self - assessment methodology** for **adult education centres is described** ."(182.txt)
- **"Participants were randomized** to the HEB group (n=40) , which included a weekly blog post over a six - month period , or to a **control group** (n=40) **that had no exposure to** the HEB ." (223.txt)
- **"Data** from the interviews **and focus groups were subject to thematic analysis** ." (299.txt)
- **"The data consist of 260 service encounters** drawn from a corpus of 1300 such **encounters** (nearly 50 hours) , **video - recorded in various locations** in Sweden and Finland in 2013–2015 ." (332.txt)
- **"It focuses in** particular on the meaning of the 'political ' from an **intercultural perspective** , what it means in terms of teachers' knowledge and how **it can pervade a teacher's choices** regarding the selection of class content , **activities** and **assessment** ." (340.txt)
- "Each algal **strain was grown** from 10 . 000 cells ml - 1 to 110 . 000 cells ml - 1 before use (3 replicates) **and tested** on juvenile rainbow trout (LT 50) and gill cells (LC 50) in true replicates ." (435.txt)
- **"With this objective we sampled the lagoon** in three sites (limnic , central and marine influence) during one year evaluating phytoplankton and **environmental variables** ." (447.txt)

4.8.3.3 Simulating situation/developing model

Before delving into the specifics of the F3.3 *Simulating situation/developing model* function, it is essential to establish a clear definition of this key concept. I propose the following definition: The *Simulating situation/developing model* function refers to creating a

representation of a concept or process to predict its behavior and facilitate a better understanding prior to analysis. The model serves as a simplified representation of reality and is employed for research examination. The development of such a model may include several steps: defining the problem before modeling, collecting relevant data, developing a conceptual framework, formulating the model in physical or computational form, validating and testing the model, simulating scenarios to make predictions, and updating and refining the model.

The F3.3 *Simulating situation/developing model* function encompasses elements such as model creation (e.g., conceptual, computational, or mathematical), validation and testing of the model, scenario analysis, identification of the impacts of various factors, assessment of the risks and benefits of different strategies, and observation of how changes in the model affect simulation outcomes.

The methodological aspect of labeling the F3.3 *Simulating situation/developing model* function involved examining how F and M authors describe the creation and validation of simulations or theoretical models. This included analyzing the language used to detail the development of models, encompassing the theoretical foundations, assumptions made, and components included. Additionally, it involved identifying explanations of the processes used to simulate real-world scenarios or phenomena. Attention was also given to the use of technical terms, procedural verbs, and detailed descriptions of the simulation environment.

The vocabulary presented in sections 4.5.1, 4.5.2, 4.6.1 and 4.6.2 highlights the most frequently used nouns and verbs in the corpus. The third type of the analyzing function, F3.3 *Simulating situation/developing model* is indicated by the following nouns and verbs, among others (refer to the excerpt below): *framework, amount, server, interfaces, toolboxes, machine learning, volunteer training, experience, self-development, dimensions, definition, paths, measurements, exercise, changes, variables, participants, control group, study, utterance, learner, students, feedback, genre, principles, practice, studies, analyses, genes; address, have developed, allows, specify, select, supports, developing, added, introduce, included, provides, will be taken, examine, might affect, were exposed, remained, asked, have been produced, were, improve, focused, analyzing, using, are, complement.*

The third type of the analyzing function is exemplified by the following instances, as presented in the excerpt below.

Class 33 (simulate situation or develop model)

- **“To address this** , **we have developed a framework** we call adaptive gPCA , **which allows the user to specify** the **amount** of weight given to the tree and **which will automatically select** an amount of weight to give to the tree .” (65.txt.)
- “The OpenML **server currently supports** client **interfaces** for **Java** , **Python** , . **NET** and **R** as well as specific interfaces for the **WEKA** , **MOA** , **RapidMiner** , scikit - learn and mlr **toolboxes**

for machine learning . " (72.txt)

- **"In developing the volunteer training** for Vancouver 2010 **our team added** eLearning – reticulius , or networked (learning) – to the volunteer Olympic **experience** ." (123.txt)
- "Subsequently , **we introduce a self - development framework for teachers** , which takes into account all the **dimensions** of openness **included in the definition** and **which provides** teachers with capacity building **paths** along each dimension ." (140.txt)
- **"The same measurements will be taken** for the cohort **a year after** the initial meas**urements were taken to examine** whether rope skipping or **exercise** behaviors **might affect changes** in these **variables** ." (220.txt)
- "Thirty **participants were exposed to an eight week intervention** twice a week for 60 minutes while the **control group** (n=30) **remained inactive** ." (251.txt)
- "Unlike Hashimoto et al . ' s **study** , **we asked the participants** to **make judgments** on the likelihood of an L2 English **utterance** to **have been produced** by an L2 English **learner** whose L1 is Japanese ." (374.txt)
- **"Students were** then provided with **feedback to improve** and **accelerate** their learning about the **genre** , underpinned by Nicol and Macfarlane - Dick's (2006) seven **principles** of good feedback **practice** ." (388.txt)
- **"Further studies would be focused** on **analyzing** allelochemicals which **could inhibit** the growth of A . minutum in the filtrate culture of S . costatum ." (433.txt)
- **"Molecular analyses using ribossomal genes are underway** to **complement the morphological study** ." (441.txt)

4.8.4 Reporting results

It is widely acknowledged that several crucial aspects must be considered when reporting results. Firstly, the study's claims must be validated by providing evidence that supports the research questions or hypotheses. Secondly, transparency is essential, necessitating the clear presentation of findings to enable other researchers to replicate the study. Thirdly, contextualization is important, placing one's findings within the broader context of existing research. Finally, objectivity is paramount, presenting results without bias or interpretation to lay a proper foundation for the discussion section, where the findings are interpreted.

Based on these assumptions, I propose the following definition: The F4 *Reporting results* function can be defined as the process of presenting the findings or outcomes of any form of research, specifically what was discovered, observed, and/or concluded. This function can involve the use of visuals (tables, figures) to illustrate the data, interpreting statistical analysis results, and transparently describing the potential limitations of a study.

The methodological aspect of labeling the F4 *Reporting results* involved a qualitative examination of how F and M authors present their research findings. This includes analyzing the linguistic and rhetorical strategies used to convey data and interpret results. Specific phrases that highlight findings, such as *the results indicate*, *data shows*, and *findings reveal*, were observed. Moreover, attention was given to how F and M authors interpret their findings within the broader context of their research questions and hypotheses, if present in the abstract. Additional procedures could include observing the use of visual aids such as tables, charts, and graphs, and how these are integrated into the narrative.

The following nouns and verbs, present in the corpus excerpt, were identified in the sentences exemplifying the F4 *Reporting results* function: *results, analysis, feasibility, topics, models, patients, research, conclusions, recommendations, systems, framework, conditions, education, modes of teaching, research model, impact, attitudes, advantages, multiculruralism, cholesterol, genders, groups, effect, gender, women, families, bilingualism, language maintenance, loss, findings, students, modules, course disciplines, view, growth, effect, proportion of filtrate, irradiance, strain; extracting, suggest, benefit, resulted, can be tailored, support, learning, proposed, have, decreased, found, is moderated, was found, desired, promote, indicate, valued, expressed, recommend, demonstrated, excerted, was related,* and *was.*

The most frequently occurring nouns and verbs used in all of the functions were presented in sections 4.5.1, 4.5.2, 4.6.1, and 4.6.2. The fourth function is exemplified by the following instances presented in the excerpt below.

Class 4 (reporting results)

- **"The results of the empirical analysis show** the **feasibility in extracting the** main **topics** from the considered in corpus ." (43.txt)
- "**The personalized models suggest that some patients benefit** more **from** the drug than others ." (78.txt)
- "**The research resulted in conclusions and recommendations** on how these **systems** – the **framework conditions** for **higher education** – **can** best **be tailored to support new modes of teaching** and **learning** . " (135.txt)
- "**The proposed research model** Amongst ICT coordinators participating in the OCL programs , attitudes regarding their student's openness to **multiculturalism have a significant impact** on the **attitudes** towards the **advantages** and the **disadvantages** of OCL (β = . 54*** , β = - . 40*** respectively) ." (147.txt)
- "Results : **Mean cholesterol decreased** during 1979 - 2008 in both **genders** and all age **groups** ." (207.txt)
- "**We further found that** the mediation **effect** of cake **is moderated** by **gender** , with stronger effects for the **women** ." (277.txt)
- "**It was found that** all the **families desired to promote bilingualism** in their families for Chinese **language maintenance** and to prevent Chinese language **loss** ." (337.txt)
- "**Findings indicate that** those **students** who were engaged with the **online** REACH **modules** attached to their **course disciplines** passed and **they** generally **valued the support** and **expressed the view** that **they would recommend other students** to participate in the program although the overall engagement **level of** at - risk **students** was lower than those with higher grades ." (356.txt)
- "**Results demonstrated that** A . minutum **exerted** no allelopathic activityon S . costatum . Thecell - free filtrate of the S . costatum culture in later exponential growth phase **inhibited the growth** of A . minutum significantly , and the inhibitory **effect was related to the proportions of** S . costatum **filtrate** in culture ." (433.txt)
- "**Irradiance of the maximum strain growth** (L m) **was 115 µmol photons** m −2 s −1 ." (484.txt)

4.8.5 **Drawing conclusions**

The dictionary definition of *draw a conclusion* is "to make a judgement" (Merriam-Webster) or "consider the facts of a situation and make a decision about what is true, correct, likely to happen" (Cambridge Dictionary). I propose the following definition: The F5 *Drawing*

conclusions function involves analyzing the available evidence or information to make a decision or judgement based on the analysis conducted. The elements of drawing conclusions can include synthesizing information from various sources, analyzing data, or interpreting results to form a judgement that is coherent and logical. Example sentences realizing this function should be clear, precise and consistent, and imply their contribution to the field of study.

The methodological aspect of labeling the F5 function involved a qualitative analysis of how female (F) and male (M) authors synthesize their findings and articulate the implications of their research. Key elements to search for included summary statements that encapsulate the main findings, such as *in summary*, *to conclude*, or *overall, the study demonstrates*. It was also essential to observe how F and M authors discuss the broader implications of their research for theory, practice, or future studies. Furthermore, it was necessary to identify how F and M authors acknowledge the limitations of their study and provide recommendations for future research. Finally, reflective commentary on the research process and its outcomes was also sought.

The F5 *Drawing conclusions* function may be correlated with the following nouns and verbs identified in the corpus (refer to excerpt below): *packages, tasks*, *reproducibility, results, drudge work, collaboration, visibility, categorization, aspect, strategies, process, learning, platforms, implications, tuition, delivery, module material, staff development, tutors, trend, activity, conclusions, development, reason, programs, opportunities, teaching, practices, needs, variation, factors, relationship, venue, theatre show, age, religion, knowledge, conditions, bloom, risk periods, species; employed, have proved, ensuring, automates, facilitates, offers, analyzing, learining, can be used, compare, have implications, occurs, was, increased, is, suggest, can be enriched, respond, is explained, associated,* and *suggest.*

The most frequently occurring nouns and verbs used across all functions were presented in sections 4.5.1, 4.5.2, 4.6.1 and 4.6.2. Additional vocabulary correlated with F5 *Drawing conclusions* may be found in the excerpt below. The F5 *Drawing conclusions* function label is marked in the following excerpt:

Class 5 (drawing conclusions)

- **"The R employed packages have proved suitable** and comprehensive **for the** required **tasks** ." (43.txt)
- "**Beyond ensuring reproducibility of results** , **it automates much of the drudge work** , speeds up research , **facilitates collaboration** and **increases user's visibility** online ." (72.txt)
- "**My categorization** by **learning strategies** **offers a way of analysing this specific aspect** of learning in immersive environments : it can be used to compare the learning process about PLACE - Hampi with

the immersive learning of other **platforms** ." (169.txt)

- "From the perspective of a Higher Education (HE) institute , **this may have implications for the tuition delivery strategy** used to deliver the **module material** , as well as **how staff development occurs for the tutors** that deliver the material ." (190.txt)
- "Between 2001 - 10 , **there was a trend for increased sufficient muscle - strengthening activity** participation ." (238.txt)
- "**Conclusions** **The participatory quality development is** supposed to pave the way for the **permanent establishment of the certification process** ." (294.txt)
- "Instead **there is reason to suggest that** contemporary **bilingual programs can be enriched** with **opportunities** for translanguaging **teaching** and **learning practices** in ways that **respond to changing needs** of contemporary multilingualism ." (330.txt)
- "Instead , **the variation is better explained** as the outcome of several interacting **contextual factors** , such as perceived **interpersonal relationships** , the overall **type of venue** and **theatre show** / event , as well as **macro - level factors** such as **age** and **region** / nation ." (332.txt)
- "**The knowledge of the environmental conditions associated to toxic blooms** of Dinophysis of the acuminata complex in this coastal area **would be indicative of future risk periods** of DSP events ." (439.txt)
- **"These results suggest that Gambierdiscus cf . silvae is a dominant species** below 15 m in Japanese coastal waters ." (484.txt)

4.8.6 **Reviewing**

The verb *review* can be defined as "to give a critical evaluation of" (Merriam-Webster Dictionary) or "to think about something again, in order to make changes to it or to make a decision about it" (Cambridge Dictionary). In an academic setting *reviewing* can be defined, following the Cambridge Dictionary, as "to describe the most important facts in a piece of academic research (= detailed study of a subject) that you have read, and give your opinion of the research". Reviewing may also refer to actions within academic communication, such as summarizing existing research on a particular topic to provide a comprehensive literature review, assessing the quality and relevance of manuscripts submitted to academic journals, evaluating academic work by other experts in peer review, and evaluating the significance and methodology of proposed research projects. Thus, I would define the F6 *Reviewing* function, in brief, as the process of critically evaluating scholarly work.

The methodological approach to labeling the F6 *Reviewing* function entailed a qualitative analysis of how female (F) and male (M) authors evaluate, summarize, and synthesize existing literature, theories, and research findings. This analysis focused on the use of evaluative language and the integration of previous research. Key elements considered included: Evaluating how F and M authors summarize key points from existing literature, using phrases such as *according to*, *previous studies have shown*, or *as outlined by.* Observing the use of language that evaluates the strengths and weaknesses of existing research. Examining how F and M authors integrate various sources to provide a comprehensive overview, highlighting connections and gaps in the literature.

The adjectives and adverbs that exemplify the F6 *Reviewing* function within sentences are presented below. They are also highlighted in the excerpt: *extremely easy, single, dimensional, classic, the same, several, intentional, likely, substantial, compelling, more enjoyable, important, necessary, such, pre-recorded, new, pragmatic, temporally, also, heavier, more, diverse, traditional, institutional, previous, better, only, physically, currently, suitable, too small,* and *confidently*.

Examples of the F6 *Reviewing* function labeled in the corpus are presented below in the form of excerpts. Additionally, the most frequently used adjectives and adverbs in the corpus were presented in sections 4.5.3., 4.5.4., 4.6.3., and 4.6.4.

Class 6 (reviewing)

- "The shinyjs package makes all of these **extremely easy** by calling a **single** function ." (35.txt)
- "RosettaHUB - Sheets combine the flexibility of the bi - **dimensional** data representation model of **classic** spreadsheets with the power of R , Python , Spark and SQL ." (67.txt)
- "The exercise should be done by the course team and support staff before the course starts , in order to anticipate opportunities for improvement and at **the same** time establish expectations , taking into account **several** aspects (e . g . the number of learners , their needs and the specificities of the course) ." (105.txt)
- "It is **likely** that the cholesterol decrease has contributed to the observed **substantial** decline in cardiovascular disease in this population"(207.txt)
- "Interventions for helping pre - **intentional** HCS form an exercise intention may wish to focus on detailing the **compelling** benefits of strength exercise and making it **more enjoyable** (attitude) , engaging support from **important** others (norms) , and providing the **necessary** access and exercise skills (perceived control) ." (300.txt)
- "**Such** pre - recorded online lectures create **new pragmatic** challenges for lecturers , because the MOOC environment asks them not **only** to engage with audiences from which they are **physically** and **temporally** disconnected , but **also** to bear **heavier** burdens in engaging with students from **more diverse** backgrounds and retaining these students throughout the course without the **traditional institutional** incentives of grades and course progression ." (326.txt)
- "In a **previous** presentation (Hashimoto , Takeyama , & Yamato , 2015) , we reported that reaction times — measuring processing difficulty to understand accented speeches — can be a **better** predictor of the comprehension performance of the accented speeches than a questionnaire - based measure (i . e . , comprehensibility) ."(374.txt)
- "**Currently** all biotoxin - producing phytoplankton species are monitored using light microscopy , however this technique is not **suitable** for A . spinosum as cells are **too small** to **confidently** identify ."(411.txt)

4.8.7 **Evaluating**

The dictionary definitions for the verb *evaluate* include: "to determine the significance, worth, or condition of usually by careful appraisal and study" (Merriam-Webster Dictionary), and "to judge, calculate the quality, importance, amount, or value of something" (Cambridge Dictionary).

I would define the F7 *Evaluating* function as the process of assessing the quality, value, or significance of a particular element based on specific criteria. This function encompasses various activities within academic communication, such as examining the methods, data, and conclusions of research to determine the validity of a study, assessing student performance in relation to learning objectives, and evaluating academic programs to

ascertain their effectiveness. These evaluative actions are essential for maintaining academic standards and are often supported by quantitative and qualitative methods.

The methodological aspect of labeling the F7 *Evaluating* function involved examining how F and M authors critically assess data, methods, theories, or findings within their work. Key aspects included identifying language that conveys judgment or assessment, such as *effective*, *valid*, *reliable*, *insufficient*, or *flawed*. Additionally, it involved observing how F and M authors compare their findings with benchmarks, standards, or previous research, using comparative phrases such as *more effective than*, *less reliable than*, or *comparable to*. Furthermore, it included evaluating the justifications provided for assessments, such as the use of data, statistical evidence, and theoretical support.

In addition to the most frequently used adjectives and adverbs presented in sections 4.5.3, 4.5.4, 4.6.3, and 4.6.4, which, according to Martin (1976), indicate the use of the evaluating function in academic texts, additional adverbs and adjectives used in sentences that realize this function are: *right, straightforward, suspect, invalid, different, significant, transformative, Laureate, acceptable, challenging, comparable, multilingual, suitable, genetic, basic, molecular, new, higher, often, systematically, interactively, naturally, profoundly,* and *usually*. Examples of the F7 Evaluating function identified in the corpus are presented below in the form of excerpts:

Class 7 (evaluating)

- "These feedback loops help ensure that data scientists are answering the **right** questions and that business users are given the opportunity to invest themselves in the analysis , which **often** expedites the execution and adoption of the data science work." (36.txt)
- "These dashboards allow analysts to **systematically** visualize and validate data using a **straightforward** user interface , and **interactively** mark data points that are **suspect** or **invalid** ." (54.txt)
- "From Jordan , we learn how to blend **different** learning resources and how teachers and students roles are **naturally** reconceived in a scenario where they have the opportunity to master their actions and **profoundly** develop their accountability ." (188.txt)
- "This represents a **significant transformative** mandate for all **Laureate** institutions ." (141.txt)
- "Sensor validity and data transmission reliability were **acceptable** ." (239.txt)
- "However in Australia , no national land use dataset exists , making it **challenging** to calculate nationally **comparable** walkability indices ." (249.txt)
- "Singapore , being a **multilingual** society , provides a **suitable** context for this study as its Bilingual policy ensures that English language and the mother tongue language are taught in schools ." (337.txt)
- "Bias due to extraction efficiency and quantification accuracy of **genetic** material is usually overlooked , although both steps are **basic** for any **molecular** analysis and the development of **new molecular** - based methods ." (440.txt)
- "Although blooms of 7 . 8 x 10 4 cells L - 1 were registered in 2009 and a **higher** density bloom of 1 . 73 x 10 6 cells L - 1 in 2013 , it was never before associated to bioluminescence ." (448.txt)

4.8.8 **Organizing order of information**

The elements that organize the order of information in an academic setting can be described as structuring academic writing to enhance readability and logical flow. This organizational

framework may involve various methods of arranging information, depending on the nature of the content and the objectives of the writing. Such methods include using chronological order, dividing information by related ideas, employing an order of importance, describing a spatial or sequential order, and using comparison and contrast, cause and effect, or problem and solution approaches.

The dictionary definitions for the verb *organize* include: "to form into a coherent unity or functioning whole" (Merriam-Webster Dictionary) and "to do or arrange something according to a particular system" (Cambridge Dictionary). With this information in mind, the following definition of the F9 function can be formulated: the *organizing order of information* function, which involves the choice and use of methods helping to structure academic writing in order to enhance cohesion and coherence.

The methodological aspect of researching the F9 *Organizing order of information* function involved a qualitative analysis of how F and M authors structure and sequence their content to enhance clarity, coherence, and logical progression in academic texts. Key elements to consider included the overall organization of the text, such as the use of headings, subheadings, and paragraph divisions to guide the reader through the content. Additionally, it involved observing linguistic markers that indicate the sequence of information, such as *firstly*, *next*, *then*, and *finally*. It also included examining the use of cohesive devices, such as conjunctions, pronouns, and repetition of key terms, used to link ideas and maintain continuity, and observe punctuation.

According to Martin (1976), nouns, verbs, and two-word verbs signal the use of organizing order of information function in academic texts. The following nouns and verbs identified in sentences realizing the F9 *Organizing order of information* function can be found in the corpus: *reflectance to, control of, research of, development of, implementation of, monitoring of, evaluation of, based on, refers to, timing of, frequency of, changes to, Interventions include, for example, perceived as, can be summarised as, engagement of, reduction of, changed rate by, as well as*.

The excerpt is presented below. Additionally, the high-frequency nouns and verbs were discussed in sections 4.5.1, 4.5.2, 4.6.1, and 4.6.2.

Class 9 (organising_order_of_information)

- "Model **reflectance to describe** a model's structure , parameters , and derived quantities (useful for meta - programming , e . g . , automatic generation of shiny applications) ." (73.txt)"
- "Furthermore , while some visual - ization tools **provide** graphical user interfaces , many humanities researchers desire more interactive and user - friendly **control of** their data ."(95.txt)
- "**Those four phases are :** (a) Identification and **research of** the problem , (b) **Development of** the methodology of strategic DM , (c) Implementation and **monitoring of** strategic decision and (d)

Evaluation of the effects of strategic decision ."(132.txt)

- "And finally , I wanted to create a model which could predict students' expected learning performance **based on** their online learning characteristics ." (163.txt)
- "Eating architecture **refers to** the size , timing and frequency of eating occasions and **includes** : the energy content of eating occasions , the number of eating occasions per day , time intervals between eating occasions , timing of the first and last eating occasions and total eating period ." (229.txt)
- **"Interventions tested include** 1) **changes to** product positioning (**increasing** the proportion of healthy " " green " " products compared to unhealthy " " red " " products at end - of - aisle and island bin displays) , 2) adding " " Health Star " " rating shelf labels to all 4 . 5 and 5 star rated products store wide and posters promoitng fresh fruit and vegetables , and 3) in - trolley and floor based signage to promote healthy purchasing decisions ." (234.txt)"
- "**For example** , goddag **is viewed** as a formal greeting in Sweden Swedish , but **may not be perceived as** equally formal by speakers of Finland Swedish ." (332.txt)
- "These dimensions **can be summarised as** : (1) second language learning ; (2) cross - cultural understanding and multiculturalism ; (3) **development of** intergenerational empathy ; and (4) **engagement of** older people in community life ." (360.txt)
- "Many **proposed ITCs involve** reduction of fluxes , **including** : 1) water **taken into** gill channels ; 2) O 2 content of this water ; 3) O 2 **taken up** at gill membranes and **passed into** the blood ; 3) fluxes of ions (e . g . K + , Na + , Cl -) , relative to osmoregulation and cell metabolism ; 4) changed O 2 utilization rate by the whole fish due to toxic **action , as well as stress** ." (434.txt)

4.8.9 Reinforcing research process

The dictionary entries for *reinforce* include the following definitions: "to strengthen by additional assistance, material, or support: make stronger or more pronounced" (Merriam-Webster Dictionary) and "If something reinforces an idea or opinion, it provides more proof or support for it and makes it seem true" (Cambridge Dictionary). In an academic context, the verb *reinforce* can be applied to several elements, such as supporting arguments by citing sources of evidence, strengthening understanding by reinforcing key concepts through diverse teaching methods or additional examples and exercises, enhancing skills through programs or activities designed to reinforce students' abilities, and building knowledge by reinforcing learning with supplementary materials.

Therefore, the F10 reinforcing research process function involves understanding and utilizing academic vocabulary and specific terminology pertinent to one's field of study. This includes evaluating sources, identifying biases, and synthesizing information to articulate the critical points of a discourse while maintaining the integrity of an academic text. The reinforcing research process function, in reference to the research process of a written study, should enhance the precision, clarity, and credibility of the work. The use of precise language aids in clearly articulating critical insights. Overall, these reinforcements aim to achieve more effective communication within the academic setting.

The high-frequency nouns and verbs discussed in sections 4.5.1, 4.5.2, 4.6.1, and 4.6.2 were considered while labeling the F10 reinforcing research process function. Additionally, nouns and verbs identified in the sentences realizing the F10 function are presented in the excerpt below:

Class 10 (reinforcing_research_process_vocabulary)
"The iterative **development** of these **Shiny applications** also **works** well **within** the agile **framework** that **is becoming** common for **data science projects**." (36.txt)
"The interactive **graphing capabilities provided by** the **R Shiny package presents** an **opportunity** to **explore** and **interact with data** in practical and user - friendly **ways** that are relatively simple to **implement** and flexible to the **needs** of particular **datasets** ." (54.txt)
"And additionally – how should this **be reflected in supporting** technologies of **education**? " (129.txt)
"In **order** to **achieve** this **goal** we **reviewed** 40 **research papers** ." (132.txt)
"Annual **count absenteeism data were collected** from the Human Resources Department of each **workplace** ." (242.txt)
"Additional **research needs to be conducted using** a larger **sample size** ." (251.txt)
"But we **need to ask** : intelligibility for whom?" (333.txt)
"**The matched - guise technique is** a widely indirect **technique used to study language attitudes** that **has been adopted** by many **researchers** in the Catalan - speaking area ." (338.txt)
"However a single **data - file may contain** 20000 **features including** many random false **features** . " (429.txt)
"**Participation in** the ongoing **survey is** highly **desired** and strongly **encouraged** ." (480.txt)

4.9 Academic Language Functions use

The following tables and figures present the quantitative use of academic language functions in abstracts within the researched corpus regarding the parameters of gender, speaker origin, and fields of study. Figures refer either to gender: Females (blue) and Males (red); or to academic fields: Computational Statistics (blue), Humanities (red), and Biology and Health (green). The academic function numbers (F1-F10) presented in Tables 54 to 67 correspond to the data outlined in Table 7. The method utilized for this part of the dissertation is described in Section 3.5. It is important to note that each sentence corresponds to one function used, as it was set as the unit of coding. The .txt files from AntMover, with the function labels added to the sentences, were processed using ProtAnt, employing the prepared key list of academic language functions. ProtAnt summarized the data in quantitative form, which was subsequently copied into MS Excel to prepare the tables for this section.

Table 54 Academic Language Functions Quantitative Use Total in Academic Fields

Function no.	Academic Field						Total
	Computational Statistics		Humanities		Biology and Health		
	quantity	%	quantity	%	quantity	%	
F1	42	13.46%	143	45.83%	127	40.71%	312
F2	88	21.15%	165	39.66%	163	39.18%	416
F3.1	56	18.54%	101	33.44%	145	48.01%	302
F3.2	65	16.62%	154	39.39%	172	43.99%	391
F3.3	38	25.50%	77	51.68%	34	22.82%	149
F4	12	5.24%	66	28.82%	151	65.94%	229
F5	47	14.20%	141	42.60%	143	43.20%	331
F6	15	60.00%	8	32.00%	2	8.00%	25
F7	49	16.61%	124	42.03%	122	41.36%	295

F9	9	10.84%	41	49.40%	33	39.76%	83
F10	28	45.90%	18	29.51%	15	24.59%	61

(F1) Formulating
(F2) Investigating
(F3) Analyzing
  (F3.1) Presenting methodology
  (F3.2) Design the experiment
  (F3.3) Simulate the situation/develop a model
(F4) Reporting results
(F5) Drawing conclusions
(F6) Reviewing
(F7) Evaluating
(F8) Criticizing
(F9) Organizing order of information
(F10) Reinforcing research process

Table 54 and Figure 25 present the number of functions used throughout the corpus across all academic fields. The three most frequently used functions in the CS field are: F2 *Investigating* with 88 examples (21.15%), F3.2 *Design the experiment* with 65 examples (16.62%), F3.1 *Presenting methodology* used in 56 texts (18.54%). In the H field, the most frequent functions are: F2 *Investigating* with 165 examples (39.66%), F3.2 *Design the experiment* with 154 examples (39,39%), F1 *Formulating* with 143 examples (45.83%), and F5 *Drawing conclusions* with 141 examples (42.60%). The BH list shows the following most frequent functions: F3.2 *Design the experiment* with 172 examples (43.99%), F2 *Investigating* with 163 examples (39.18), F4 *Reporting results* with 151 examples (65.94%), F3.1 *Presenting methodology* with 145 examples (48.01%), and F5 *Drawing conclusions* with 143 examples (43.20%).

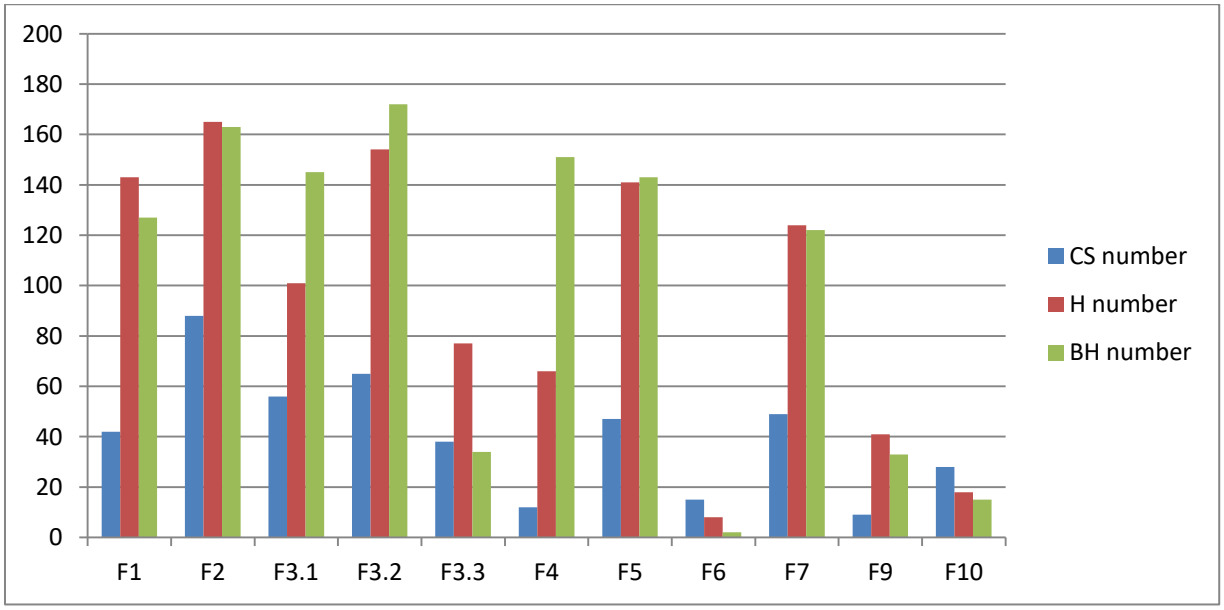

Figure 25 The number of academic language functions used in the three fields under analysis

The most frequently used functions across all fields are F1 *Formulating,* used in 416 texts, F3.2 *Design the experiment,* used in 391 texts and F2 *Investigating,* used in 389 texts. Conversely, the least used function in all academic fields is the F8 *Criticizing* function, which is not found at all. The second least frequent is the F6 *Reviewing* function, which appears in only 25 texts.

Table 55 Number of Academic Language Functions used in abstracts

Academic Field	Number of Functions	G	%	F	%	M	%	Total
Computational Statistics	1 function			1	0.21%			1
	2 functions			1	0.21%	10	2.08%	11
	3 functions	4	0.83%	2	0.42%	6	1.25%	12
	4 functions	13	2.71%	7	1.46%	11	2.29%	31
	5 functions	9	1.88%	4	0.83%	4	0.83%	17
	6 functions	8	1.67%	3	0.63%	7	1.46%	18
	7 functions	4	0.83%	1	0.21%	2	0.42%	7
	8 functions	2	0.42%	1	0.21%			3
Humanities	2 functions			1	0.21%	1	0.21%	2
	3 functions	1	0.21%	3	0.63%	8	1.67%	12
	4 functions	4	0.83%	13	2.71%	10	2.08%	27
	5 functions	11	2.29%	13	2.71%	8	1.67%	32
	6 functions	13	2.71%	14	2.92%	18	3.75%	45
	7 functions	18	3.75%	11	2.29%	11	2.29%	40
	8 functions	11	2.29%	3	0.63%	4	0.83%	18
	9 functions	2	0.42%	2	0.42%			4
Biology and Health	2 functions	2	0.42%	1	0.21%	2	0.42%	5
	3 functions	5	1.04%	1	0.21%	3	0.63%	9
	4 functions	15	3.13%	8	1.67%	13	2.71%	36
	5 functions	24	5.00%	13	2.71%	4	0.83%	41
	6 functions	35	7.29%	18	3.75%	9	1.88%	62
	7 functions	11	2.29%	13	2.71%	7	1.46%	31
	8 functions	6	1.25%	4	0.83%	2	0.42%	12
	9 functions	2	0.42%	2	0.42%			4

(G) multi-authored abstracts
(F) female abstracts
(M) male abstracts

Table 55 and Figure 26 present the quantitative distribution of functions used in abstracts across academic fields and gender divisions. As the table above shows, F authors in the CS field use between three and six functions in an abstract. The most common scheme, with the highest number of abstracts using the same number of functions by F authors, involves the use of four functions (seven texts, 1.46%). M authors in CS vary the use of functions in their texts between two and seven functions; they most frequently use four functions (11 texts, 2.29%) and two functions (10 texts, 2.08%).

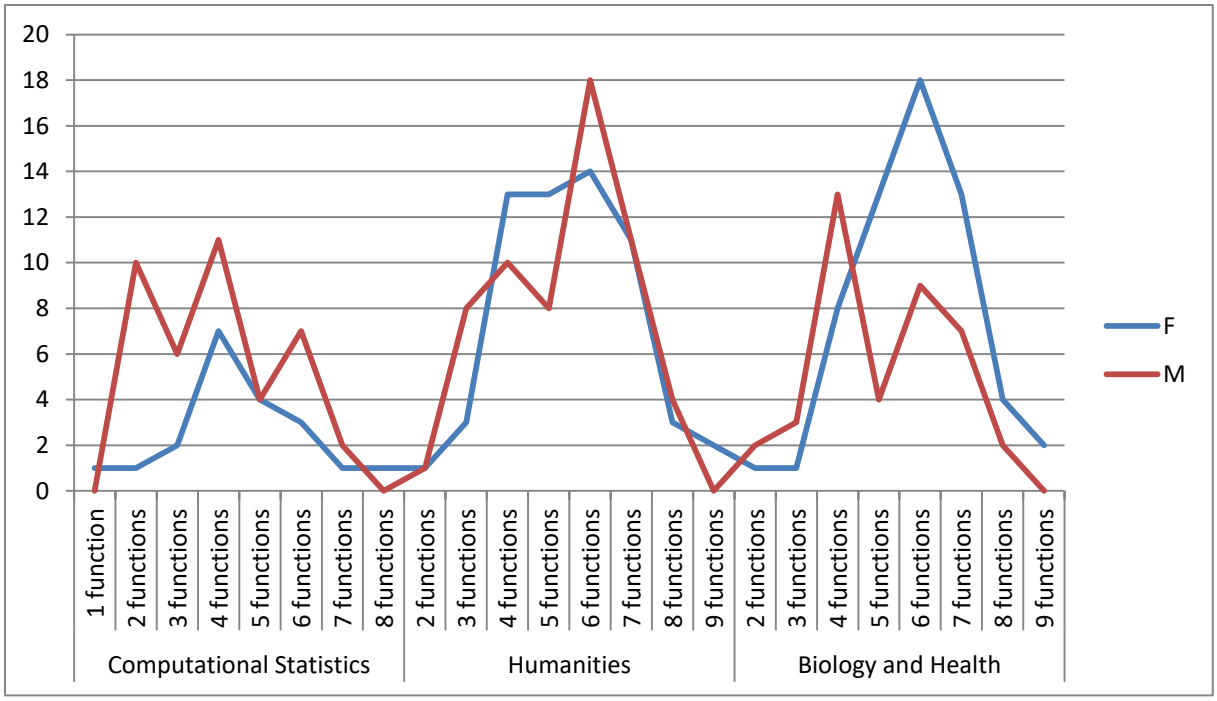

Figure 26 The number of academic language functions used in the three fields under analysis in reference to gender

Gender variation in the number of functions in the H field shows that F authors use between four and seven functions, while M authors tend to use between three and seven functions. Both female and male authors most frequently use six functions: 14 texts (2.92%) in the first group and 18 texts (3.75%) in the second group. In the third academic field, BH, as illustrated in Figure 26, both female and male authors use between four and seven functions. F authors most frequently use six functions, represented in 18 abstracts (3.75%). In comparison, M authors most frequently use four functions, represented in 13 abstracts (2.71%).

Pearson's Chi-Squared tests were conducted for all data shown in Table 55 and then repeated for the three most frequently used patterns in each academic field, as the data, when taken as a whole, showed an insufficient number of examples (i.e., less than 5), which might influence the outcome of the chi approximation. The conducted tests indicate that the number of functions used is not influenced by gender when considering the data presented in Table 55. The outcomes of Pearson's Chi-Squared test are as follows: for the CS field, X-squared = 8.4591, df = 7, p-value = 0.2939; for the H field, X-squared = 6.4974, df = 7, p-value = 0.483; and for the BH field, X-squared = 11.203, df = 7, p-value = 0.13.

In another attempt, the three most frequent patterns (i.e., patterns with four, five, and six functions) were taken into consideration. Interestingly, this second attempt indicates that in the BH field, the number of functions used might be related to the gender of the author. In the CS field, the Pearson's Chi-Squared test for significance yields X-squared = 0.74805, df = 2, p-value = 0.688. In the H field, the outcome is X-squared = 1.8765, df = 2, p-value =

0.3913. Lastly, in the BH field, X-squared = 6.62, df = 2, p-value = 0.03652. The last chi approximation shows a p-value below 0.05, providing statistical evidence of the influence of gender on the number of functions used in abstracts in this field, but only under the given circumstances, as the p-value for the whole field equaled 0.13. Nevertheless, both outcomes of the chi approximation for the BH field can be used as a basis for further study on abstracts.

Table 56 Academic Language Functions quantitative use in reference to gender and academic field

Function/Academic Field	Computational Statistics		Humanities		Biology and Health		Total
	quantity	%	quantity	%	quantity	%	
F1	42	13.46%	143	45.83%	127	40.71%	**312**
F1 G	19	6.09%	45	14.42%	60	19.23%	124
F1 F	6	1.92%	48	15.38%	48	15.38%	102
F1 M	17	5.45%	49	15.71%	19	6.09%	85
F2	88	21.15%	165	39.66%	136	32.69%	**416**
F2 G	38	9.13%	58	13.94%	85	20.43%	181
F2 F	17	4.09%	54	12.98%	43	10.34%	114
F2 M	33	7.93%	53	12.74%	35	8.41%	121
F31	56	18.54%	101	33.44%	145	48.01%	**302**
F3.1 G	26	8.61%	36	11.92%	74	24.50%	136
F3.1 F	10	3.31%	32	10.60%	46	15.23%	88
F3.1 M	20	6.62%	33	10.93%	25	8.28%	78
F32	65	16.62%	154	39.39%	172	43.99%	**391**
F3.2 G	28	7.16%	55	14.07%	85	21.74%	168
F3.2 F	13	3.32%	54	13.81%	54	13.81%	121
F3.2 M	24	6.14%	45	11.51%	33	8.44%	102
F33	38	25.50%	77	51.68%	34	22.82%	**149**
F3.3 G	20	13.42%	35	23.49%	19	12.75%	74
F3.3 F	9	6.04%	19	12.75%	11	7.38%	39
F3.3 M	9	6.04%	23	15.44%	4	2.68%	36
F4	12	5.24%	66	28.82%	151	65.94%	**229**
F4 G	5	2.18%	28	12.23%	72	31.44%	105
F4 F	4	1.75%	19	8.30%	53	23.14%	76
F4 M	3	1.31%	19	8.30%	26	11.35%	48
F5	47	14.20%	141	42.60%	143	43.20%	**331**
F5 G	25	7.55%	55	16.62%	69	20.85%	149
F5 F	9	2.72%	44	13.29%	46	13.90%	99
F5 M	13	3.93%	42	12.69%	28	8.46%	83
F6	15	60.00%	8	32.00%	2	8.00%	**25**
F6 G	4	16.00%	4	16.00%	0	0.00%	8
F6 F	4	16.00%	2	8.00%	0	0.00%	6
F6M	7	28.00%	2	8.00%	2	8.00%	11
F7	49	16.61%	124	42.03%	122	41.36%	**295**
F7 G	18	6.10%	40	13.56%	65	22.03%	123
F7 F	10	3.39%	43	14.58%	35	11.86%	88
F7 M	21	7.12%	41	13.90%	22	7.46%	84
F9	9	10.84%	41	49.40%	33	39.76%	**83**
F9 G	5	6.02%	20	24.10%	16	19.28%	41

F9 F	1	1.20%	11	13.25%	10	12.05%	22
F9 M	3	3.61%	10	12.05%	7	8.43%	20
F10	28	45.90%	18	29.51%	15	24.59%	**61**
F10 G	10	16.39%	7	11.48%	7	11.48%	24
F10 F	5	8.20%	5	8.20%	5	8.20%	15
F10 M	13	21.31%	6	9.84%	3	4.92%	22
FS Suggesting	8	12.70%	28	44.44%	27	42.86%	**63**
FS G	1	1.59%	9	14.29%	11	17.46%	21
FS F	6	9.52%	9	14.29%	9	14.29%	24
FS M	1	1.59%	10	15.87%	7	11.11%	18

Table 56 and Figure 27 present the academic language functions divided according to fields and gender. The functions used more frequently by the F authors than the M authors are: F1 *Formulating*, F3.2 *Design the experiment*, F4 *Reporting results*, F5 *Drawing conclusions*, and FS *Suggesting*. In contrast, male authors predominantly use: F2 *Investigating*, F6 *Reviewing*, and F10 Re*inforcing the research process* functions. Other functions are almost equally distributed between the two genders. The F3.3 *Simulating situations/developing models* function is present in 39 texts by the F authors and 36 by the M authors. The F7 *Evaluating* functions are found in 88 abstracts by the F authors and 84 abstracts by the M authors. The F3.1 *Presenting methodology* functions appear in 88 abstracts by the F authors and 78 abstracts by the M authors. Lastly, the F9 *Organizing the order of information* function is found in 42 abstracts, with 22 by the F authors and 20 by the M authors.

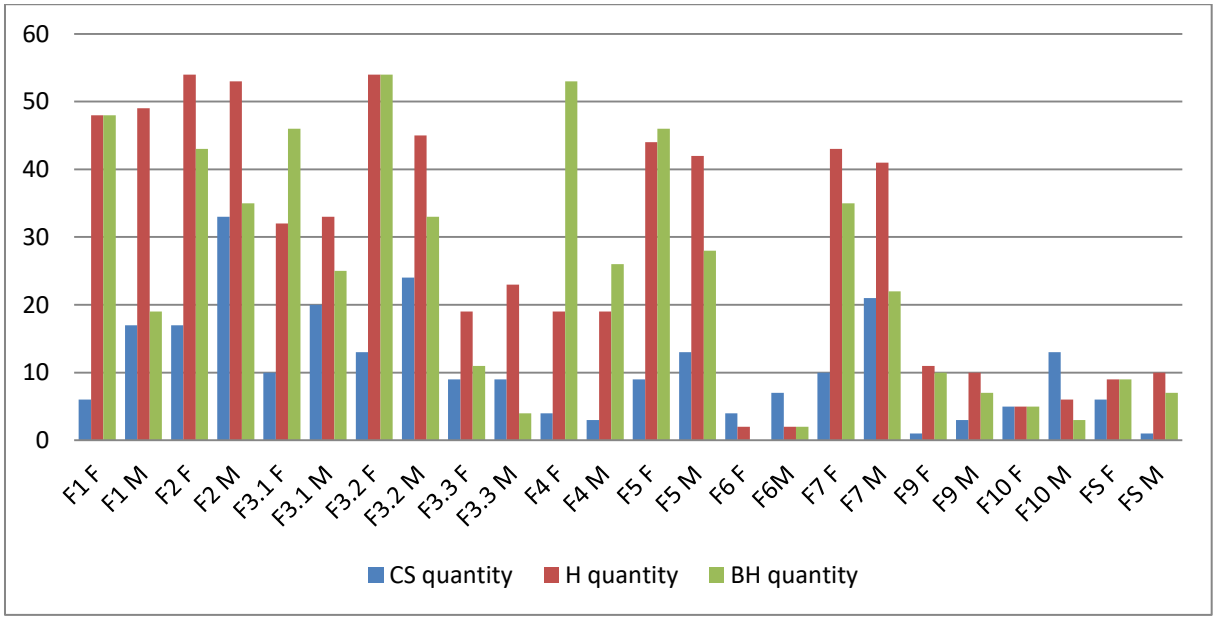

Figure 27 The number of academic language functions used in the three fields under analysis in reference to function number and gender

As illustrated in Figure 27, the distribution of functions within the Humanities field is relatively balanced between male and female authors, with minor variations of one or two

points. The only significant exception is the F3.2 function, where female authors exhibit a higher frequency (54 instances) compared to male authors (45 instances). Conversely, for the F3.3 function, male authors demonstrate a higher frequency (23 instances) than female authors (19 instances).

No instances of function F8 *Criticizing* were identified in the research corpus. As explained in Section 3.5, the FS *Suggesting* function has replaced the F8 *Criticizing* function. The most notable gender disparity in the use of FS *Suggesting* occurs in the CS field, with six uses by female authors and only one by male authors. In the H field, male authors employ FS *Suggesting* more frequently (15.87%) than female authors (14.29%). Conversely, in the BH field, female authors exhibit a higher frequency of FS *Suggesting* use (14.29%) compared to male authors (11.11%). However, a statistical analysis revealed no significant differences in the use of the FS *Suggesting* function across gender and academic field (X-Squared = 3.0798, df = 2, p-value = 0.2144), as the p-value exceeds the 0.05 threshold.

4.9.1 Vocabulary of Research Process Functions

Table 57 Formulating function quantitative use in reference to academic field, gender and speaker origin

Academic field	speaker origin	G	%	F	%	M	%	Total	%	Pearson's Chi-squared test
Computational Statistics	native	12	3.69%	4	1.23%	11	3.38%	27	8.31%	X-squared = 0.0075163, df = 1, p-value = 0.9309
	non-native	7	2.15%	2	0.62%	6	1.85%	15	4.62%	
Humanities	native	15	4.62%	24	7.38%	22	6.77%	61	18.77%	X-squared = 0.089871, df = 1, p-value = 0.7643
	non-native	30	9.23%	24	7.38%	27	8.31%	81	24.92%	
Biology and Health	native	35	10.77%	28	8.62%	12	3.69%	75	23.08%	X-squared = 0.0074995, df = 1, p-value = 0.931
	non-native	25	7.69%	20	6.15%	7	2.15%	52	16.00%	
				X-squared = 9.4131, df = 5, p-value = 0.09368						
Total	native	62	19.87%	56	18.0%	45	13.85%	163	52.24%	X-squared = 0.32636, df = 2, p-value = 0.8494
	non-native	62	19.87%	46	14.79%	40	12.31%	148	47.58%	

Table 57 and Figure 28 present the quantitative use of the F1 *Formulating* function within the corpus. The total number of native and non-native abstracts in the research disciplines is almost equal for both groups, with 163 (52.24%) examples for native (N) writers and 148 (47.58%) examples for non-native (NN) writers. The F1 *Formulating* function was utilized more frequently by the F authors, with 62 instances (19.08%, native) and 54 instances (16.62%, non-native), compared to the M authors, who used it 45 times (13.82%, native) and 40 times (12.31%, non-native). Furthermore, there is a slightly higher quantitative use of the F1 *Formulating* function by native writers overall, with 56 instances for the NF authors (18%)

and 45 cases for the NM authors (13.85%). The Pearson's Chi-Squared tests for significance in all fields show p-values of 0.930, 0.764, and 0.931. These outcomes cannot be considered significant in this research as they do not meet the established p-value threshold of less than 0.05.

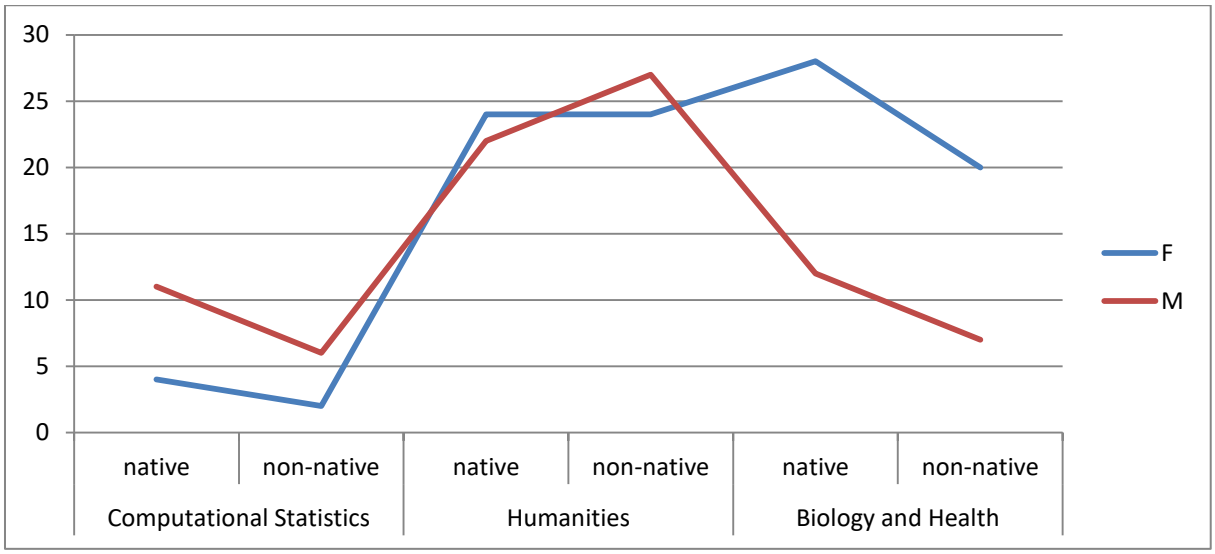

Figure 28 Formulating function: quantitative use by gender and speaker origin

Subsequently, there is a nearly threefold difference in the quantitative use of the F1 *Formulating* function between female and male NN writers in the fields of Biology and Health, with 20 instances (6.15%) compared to seven instances (2.15%). Moreover, there is a more than twofold difference between the N writers of different genders in this domain, with the F authors exhibiting 28 instances (8.62%) and M authors demonstrating 12 instances (3.69%). A similar trend is observed for the N writers' gender in the H field, albeit with a smaller discrepancy: the F authors used the F1 *Formulating* function in 24 texts (7.38%) compared to 22 texts (6.77%) by male authors. Conversely, an opposite pattern emerges in the CS field. Here, the M authors employed the function in 11 texts (3.38%), while the F authors in 4 texts (1.23%).

The Pearson Chi-Squared test for significance in differences between the fields of Humanities (H) and Biology and Health (BH) indicates that it cannot be deemed valid for this research due to a p-value of 0.09. Nonetheless, this finding may serve as a suitable foundation for future investigations of abstracts.

Table 58 Investigating function quantitative use in reference to academic field, gender and speaker origin

Academic field	speaker origin	G	%	F	%	M	%	Total	%	Pearson's Chi-squared test
Computational Statistics	native	18	4.33%	8	1.92%	15	3.61%	41	9.86%	X-squared = 0.011625, df = 1, p-value = 0.9141
	non-native	20	4.81%	9	2.16%	18	4.33%	47	11.30%	

Humanities	native	20	4.81%	26	6.25%	27	6.49%	73	17.55%	X-squared = 0.0091735, df = 1, p-value = 0.9237
	non-native	38	9.13%	28	6.73%	26	6.25%	92	22.12%	
Biology and Health	native	54	12.98%	20	4.81%	16	3.85%	90	21.63%	X-squared = 0.0049359, df = 1, p-value = 0.944
	non-native	31	7.45%	23	5.53%	19	4.57%	73	17.55%	
					X-squared = 5.8454, df = 5, p-value = 0.3215					
Total	native	92	22.12%	54	12.98%	58	13.94%	204	49.04%	X-squared = 0.41843, df = 2, p-value = 0.8112
	non-native	89	21.39%	60	14.42%	63	15.14%	212	50.96%	

Table 58 and Figure 29 display the data concerning the use of the F2 *Investigating* function, compared to the combined parameters of gender and speaker origin in all academic fields. There are more instances where the function is utilized by the NN authors in total, amounting to 212 examples (50.96%). In contrast, within the field of BH, this function is employed in 90 texts by the N authors (21.63%), compared to 73 texts by the NN authors (17.55%). In CS field, gender differences reveal fewer instances of the F2 *Investigating* function in abstracts by the F authors, with 8 occurrences among the N writers (1.92%) and 9 among the NN writers (2.16%). By comparison, the M authors use this function nearly twice as frequently, with 15 occurrences among the N writers (3.61%) and 18 among the NN writers (4.33%). The Pearson's Chi-Squared test for significance across all academic fields indicates no significant differences between speaker origin and the use of the F2 *Investigating* function, as the p-values are notably high: 0.91 in CS, 0.92 in H, and 0.94 in BH. Consequently, these results are not statistically significant.

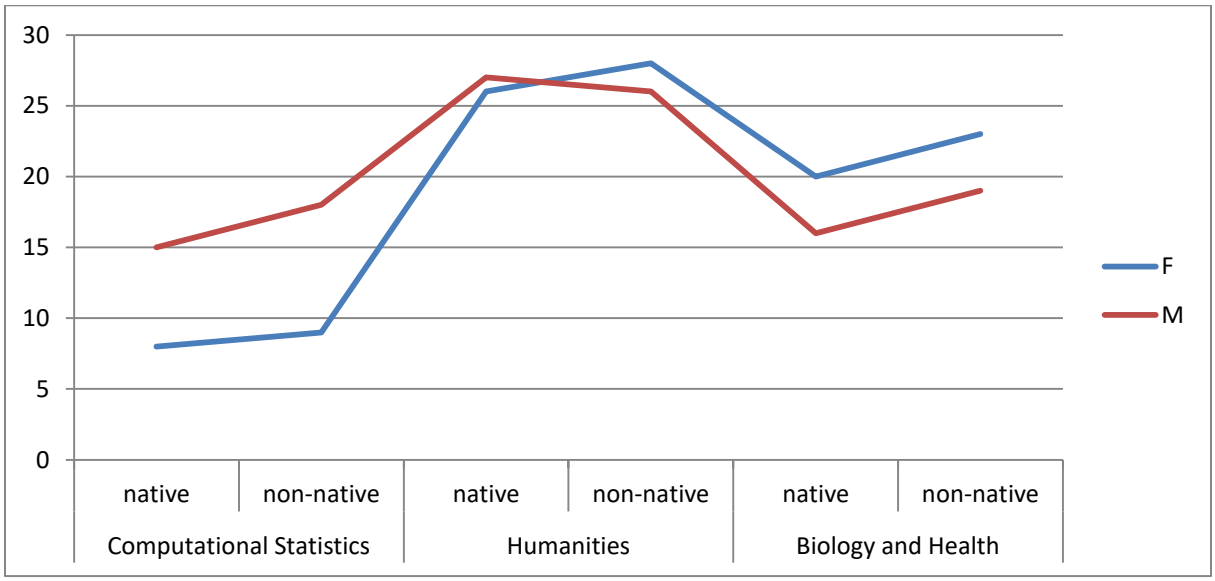

Figure 29 Investigating function: quantitative use by gender and speaker origin

A slight shift towards more frequent use of the function by the NN authors can be observed in the fields of CS and BH. In non-native female (NNF) abstracts, 23 instances of the F2

Investigating function were identified (5.53%). Additionally, in the M abstracts, 19 instances of F2 *Investigating* usage were observed (4.57%). Conversely, a minor but opposite trend is noted in the H field among the M writers, with 27 instances found in the N abstracts (6.49%) and 26 instances in the NN abstracts (6.25%). The Pearson's Chi-Squared test for significance in each field yields the following p-values: 0.91 for Computational Statistics, 0.94 for Biology and Health, and 0.92 for Humanities. The significance test for differences concerning the gender of authors across all academic fields has a p-value of 0.32, which is not significant. Thus, it can be stated that there is no statistically proven link between gender, speaker origin, and the researched academic fields in the use of the *F2 Investigating* function.

Table 59 Presenting methodology function quantitative use in reference to academic field, gender and speaker origin

Academic field	speaker origin	G	%	F	%	M	%	Total	%	Pearson's Chi-squared test
Computational Statistics	native	15	4.97%	5	1.66%	8	2.65%	28	9.27%	X-squared = 0.016968, df = 1, p-value = 0.8964
	non-native	11	3.64%	5	1.66%	12	3.97%	28	9.27%	
Humanities	native	11	3.64%	16	5.30%	15	4.97%	42	13.91%	X-squared = 0.01403, df = 1, p-value = 0.9057
	non-native	25	8.28%	16	5.30%	18	5.96%	59	19.54%	
Biology and Health	native	45	14.90%	27	8.94%	17	5.63%	89	29.47%	X-squared = 0.26568, df = 1, p-value = 0.6062
	non-native	29	9.60%	19	6.29%	8	2.65%	56	18.54%	
				X-squared = 9.9123, df = 5, p-value = 0.07776						
Total	native	71	23.51%	48	15.89%	40	13.25%	159	52.65%	X-squared = 0.19613, df = 2, p-value = 0.9066
	non-native	65	21.52%	40	13.25%	38	12.58%	143	47.35%	

Table 59 and Figure 30 show the quantitative use of the F3.1 *Presenting methodology* function in corpus abstracts. Notably, 52.65% of the total originate from native authors' abstracts. In the research fields of CS and H, there appears to be a tendency towards more frequent use of the function by the M authors, with eight (2.65%), 12 (3.97%), and 18 (5.96%) examples, respectively. Conversely, in the BH field, the opposite trend is observed, with the F authors utilizing the F3.1 more frequently than the M authors. There are 27 (8.94%, NF) and 19 (6.29%, NNF) examples, compared to 17 (5.63%, NM) and 8 (2.65%, NNM) examples.

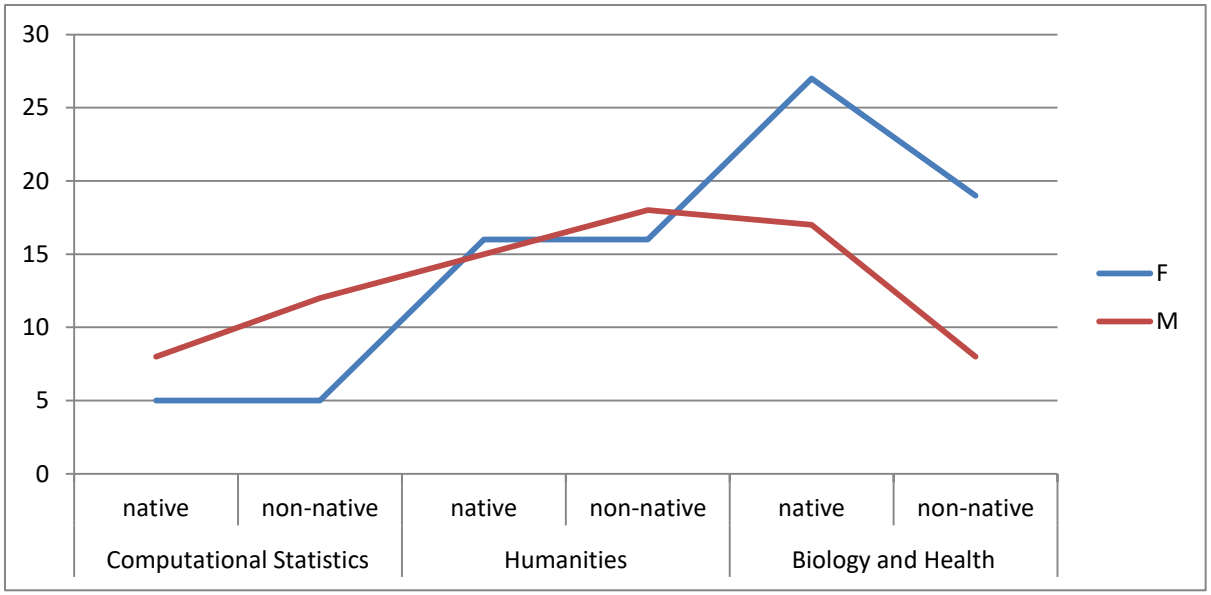

Figure 30 Presenting methodology: quantitative use by gender and speaker origin

Furthermore, F abstracts in H and CS seem not to show any significant connection with speaker origin, as the quantity of the function used in the NF and NNF groups is the same: five instances for CS (1.66%) and 16 for H (5.30%). The most significant difference between genders can be observed in the field of BH. In the NNM abstracts, there are eight instances (2.65%), compared to 19 instances (6.29%) in the NNF abstracts. The Pearson's Chi-squared tests show p-values of 0.89 in CS, 0.90 in H, and 0.60 in BH. The differences concerning the parameters of gender and speaker origin in relation to all academic fields have a p-value of 0.07, which is not considered significant for this research. However, it serves as a reasonable basis for further investigation.

Table 60 Design the experiment function quantitative use in reference to academic field, gender and speaker origin

Academic field	speaker origin	G	%	F	%	M	%	Total	%	Pearson's Chi-squared test
Computational Statistics	native	14	3.58%	6	1.53%	13	3.32%	33	8.44%	X-squared = 0.01465, df = 1, p-value = 0.9037
	non-native	14	3.58%	7	1.79%	11	2.81%	32	8.18%	
Humanities	native	17	4.35%	26	6.65%	22	5.63%	65	16.62%	X-squared = 0.0053922, df = 1, p-value = 0.9415
	non-native	38	9.72%	28	7.16%	23	5.88%	89	22.76%	
Biology and Health	native	49	12.53%	30	7.67%	18	4.60%	97	24.81%	X-squared = 0.0084499, df = 1, p-value = 0.9268
	non-native	36	9.21%	24	6.14%	15	3.84%	75	19.18%	
				X-squared = 7.8061, df = 5, p-value = 0.1672						
Total	native	80	20.46%	62	15.86%	53	13.55%	195	49.87%	X-squared = 0.60964, df = 2, p-value = 0.7373
	non-native	88	22.51%	59	15.09%	49	12.53%	196	50.13%	

Table 60 and Figure 31 illustrate the quantitative use of the F3.2 *Design the experiment* function across the researched academic disciplines. The total number of the N and NN abstracts employing this function is almost equal, with 195 N abstracts (49.87%) compared to 196 NN abstracts (50.13%). It can be observed that the N authors use the function more frequently in the BH field, with 97 examples (24.81%), while the NN authors predominantly use the function in the H field, with 89 texts (22.76%). In the field of CS, the M authors demonstrate a higher frequency of this function usage, with 13 instances among the N authors (3.32%) and 11 instances among the NN authors (2.81%). Conversely, F authors use the F3.2 *Design the experiment* function more frequently in the H and BH fields.

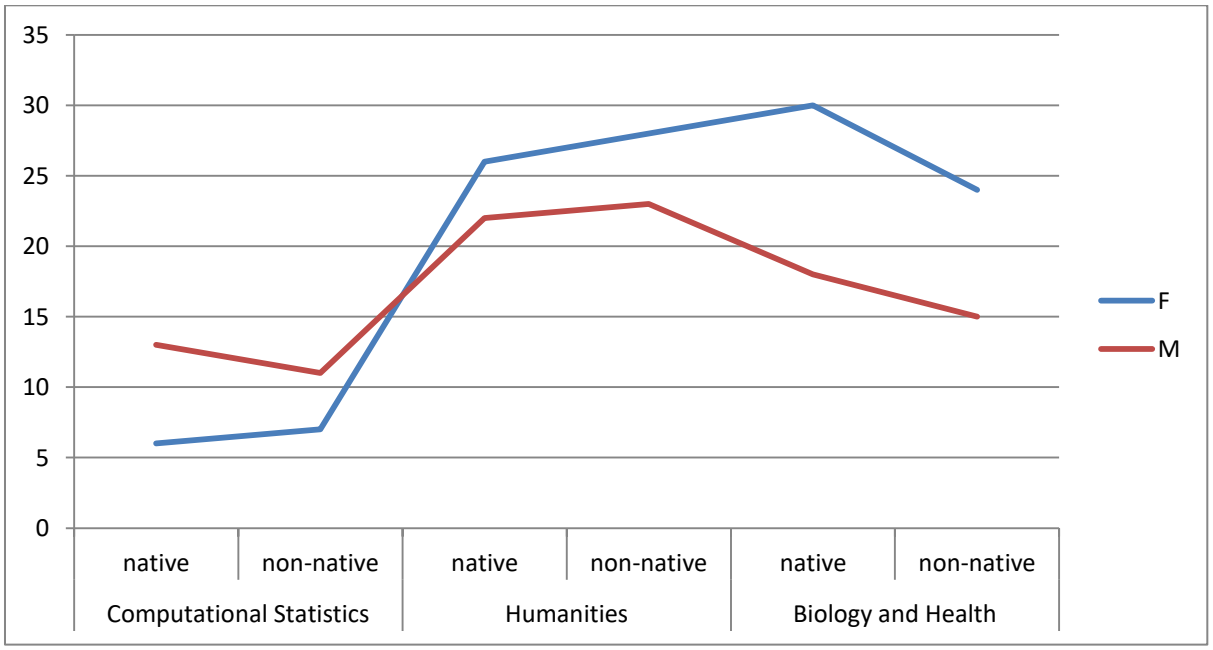

Figure 31 Design the experiment function: quantitative use by gender and speaker origin

In the H field, there are 26 instances (6.65%) in the NF group and 28 instances (7.16%) in the NNF group. Additionally, there are 22 instances (5.63%) for the NM writers and 23 instances (5.88%) for the NNM writers. In the field of BH, the quantitative differences show 30 instances (7.76%) for the NF authors and 24 instances (6.14%) for the NNF authors, as well as 18 instances (4.60%) for the NM authors and 15 (3.84%) for the NNM authors. The Pearson's Chi-squared tests for the significance of speaker origin influence on the use of F3.2 *Design the experiment* across all research fields yield p-values of 0.90 for CS, 0.94 for H, and 0.92 for BH. The p-value for total quantities is 0.73, indicating non-significant values. The quantitative differences between genders across all research fields result in a p-value of 0.16, which provides a solid foundation for further investigation, although it is not considered significant in the present study.

Table 61 Simulate situation/develop model function quantitative use in reference to academic field, gender and speaker origin

Academic field	speaker origin	G	%	F	%	M	%	Total	%	Pearson's Chi-squared test
Computational Statistics	native	11	7.38%	5	3.36%	3	2.01%	19	12.75%	X-squared = 0.225, df = 1, p-value = 0.6353
	non-native	9	6.04%	4	2.68%	6	4.03%	19	12.75%	
Humanities	native	6	4.03%	8	5.37%	11	7.38%	25	16.78%	X-squared = 0.0035189, df = 1, p-value = 0.9527
	non-native	29	19.46%	11	7.38%	12	8.05%	52	34.90%	
Biology and Health	native	15	10.07%	7	4.70%	0	0.00%	22	14.77%	X-squared = 2.5583, df = 1, p-value = 0.1097
	non-native	4	2.68%	4	2.68%	4	2.68%	12	8.05%	
				X-squared = 8.3105, df = 5, p-value = 0.1399						
Total	native	32	21.48%	20	13.42%	14	9.40%	66	44.30%	X-squared = 1.2312, df = 2, p-value = 0.5403
	non-native	42	28.19%	19	12.75%	22	14.77%	83	55.70%	

Table 61 and Figure 32 demonstrate the quantitative differences in the F3.3 *Simulate situation/develop model* function across the researched academic disciplines. A total of 83 examples (55.70%) of NN writers' texts use this function, compared to 66 examples (44.30%) of the N writers' texts. The F3.3 S*imulate situation/develop model* function is used most frequently in the H field, with 52 examples (34.90%) from the NN writers and 25 examples (16.78%) from native writers. Notably, in the H field, the function is utilized more frequently by male authors, with 11 instances (7.38%) among NM and 12 instances (8.05%) among NNM. Conversely, female authors' abstracts show 8 instances (5.37%) among NF and 11 ones (7.38%) among NNF. In the CS field, the differentiation between the N and NN abstracts employing the F3.3 S*imulate situation/develop model* function results in a combined total of 19 instances (12.75%). Furthermore, in the BH field, the N authors demonstrate a higher frequency of using the F3.3 S*imulate situation/develop model* function compared to the NN authors, with 22 instances (14.77%) versus 12 instances (8.05%), respectively.

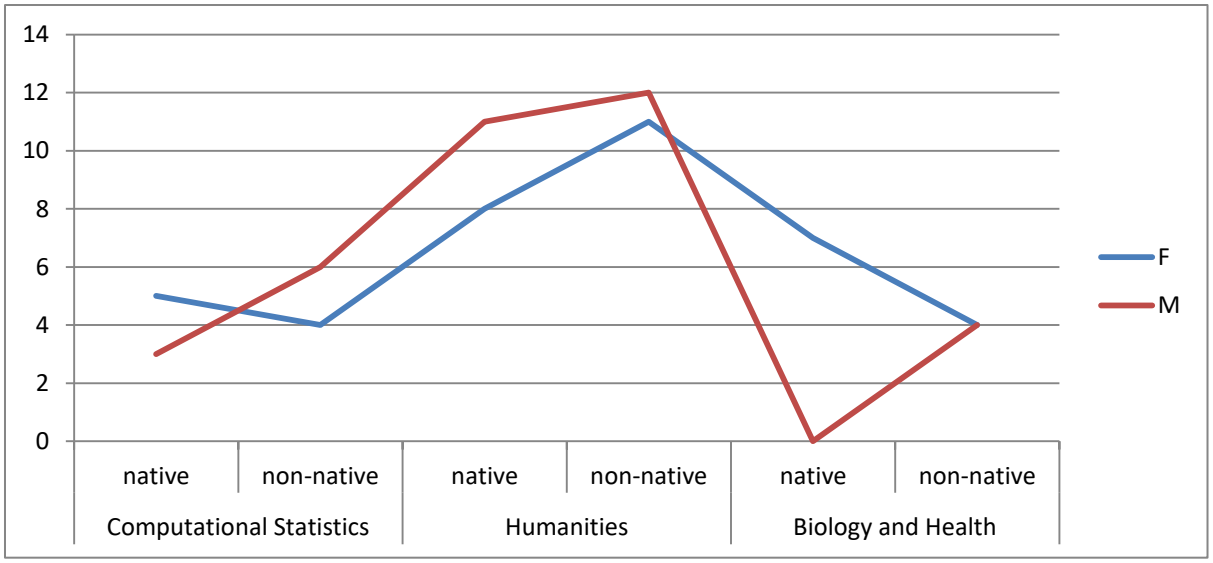

Figure 32 Simulate situation/develop model function: quantitative use by gender and speaker origin

When the gender parameter is considered, the BH field demonstrates the most significant quantitative disparity, with seven examples (4.70%) for the NF and none for the NM. Pearson's Chi-squared tests for significance in quantitative differences between speakers' origins across the fields yield scores of 0.63 for CS, 0.95 for H, and 0.10 for BH. However, it is crucial to note that these tests may be flawed, as some values presented in Table 61 and Figure 32 are less than 5. The differences in speaker origin, when tested for statistical significance, have a p-value of 0.54. In contrast, the gender differences within the research fields yield a p-value of 0.13, suggesting a potential avenue for further investigation. Nonetheless, none of the tests conducted in reference to F3.3 *Simulate situation/develop model* are found to be statistically significant in this study.

Table 62 Reporting results function quantitative use in reference to academic field, gender and speaker origin

Academic field	speaker origin	G	%	F	%	M	%	Total	%	Pearson's Chi-squared test
Computational Statistics	native	1	0.44%	0	0.00%	2	0.87%	3	1.31%	X-squared = 1.1812, df = 1, p-value = 0.2771
	non-native	4	1.75%	4	1.75%	1	0.44%	9	3.93%	
Humanities	native	12	5.24%	8	3.49%	6	2.62%	26	11.35%	X-squared = 0.1131, df = 1, p-value = 0.7366
	non-native	16	6.99%	11	4.80%	13	5.68%	40	17.47%	
Biology and Health	native	41	17.90%	26	11.35%	15	6.55%	82	35.81%	X-squared = 0.23257, df = 1, p-value = 0.6296
	non-native	31	13.54%	27	11.79%	11	4.80%	69	30.13%	
				X-squared = 8.0272, df = 5, p-value = 0.1547						
Total	native	54	23.58%	34	14.85%	23	10.04%	111	48.47%	X-squared = 0.79792, df = 2, p-value = 0.671
	non-native	51	22.27%	42	18.34%	25	10.92%	118	51.53%	

Table 62 and Figure 33 illustrate the quantitative differences in the use of the F4 *Reporting results* function, considering the combined parameters of gender and speaker origin across all academic fields. The most significant difference (6.12%) between the N and NN writers is observed in the H field, with 26 instances (11.35%) for N and 40 instances (17.47%) for NN authors. In the BH field, the total number of instances for the N and NN writers are 82 (35.81%) and 69 (30.13%), respectively. The F abstracts in the BH field exhibit a higher frequency of utilizing the F4 *Reporting results* function in both the N and NN groups, with 26 (11.35%) and 27 (11.79%) instances, respectively. In contrast, the M abstracts in the BH field present 15 instances (6.55%) for NM and 11 (4.80%) for NNM. Furthermore, similar gender-based quantitative differences are noted in the CS field for the NN group, where F authors employ the F4 *Reporting results* function more frequently, with 4 instances (1.75%) compared to 1 instance (0.44%) for M authors.

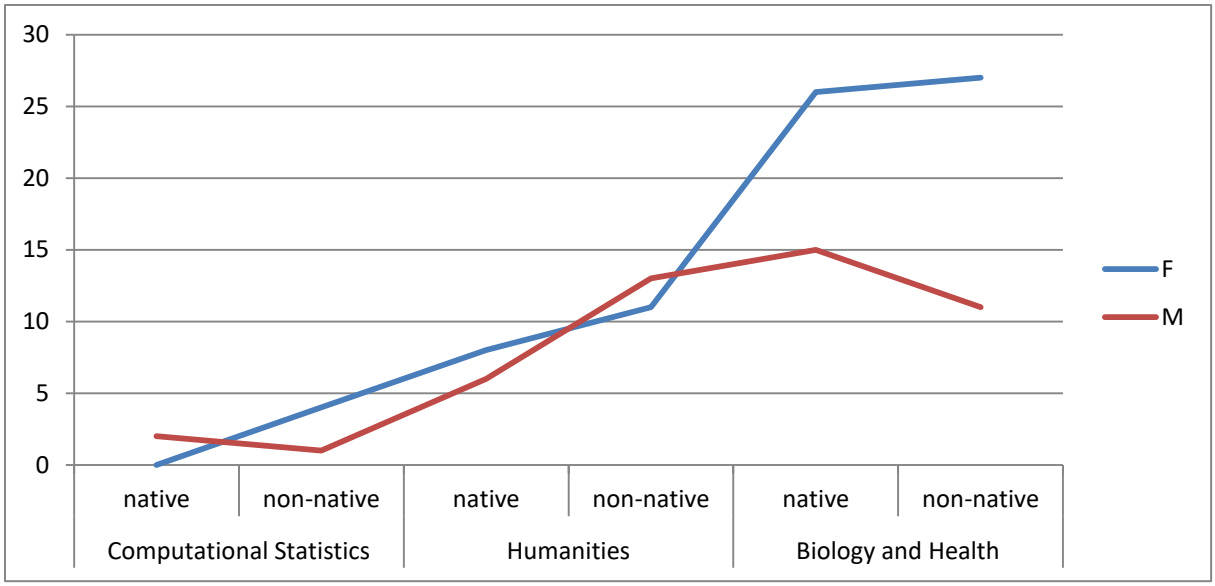

Figure 33 Reporting results function: quantitative use by gender and speaker origin

Overall, there are 34 instances (14.85% NF) and 42 instances (18.34% NNF) of the F abstracts utilizing the F4 *Reporting results* function. Conversely, there are 23 instances (10.04% NM) and 25 instances (10.92% NNM) of the M abstracts employing this function. Pearson's Chi-square tests for significance across different academic fields yield the following p-values: 0.27 for CS, 0.73 for H, 0.62 for BH, and 0.67 for the total number of examples. Unfortunately, none of these tests are regarded as significant within this study. The test for significance with respect to gender and speaker origin parameters across the fields yields a p-value of 0.15. Still it may be inaccurate due to insufficient data, i.e., values lower

than 5. Nevertheless, further investigation into the trend of more frequent use of the F4 *Reporting results* function by F authors may yield significant outcomes.

Table 63 Drawing Conclusions function quantitative use in reference to academic field, gender and speaker origin

Academic field	speaker origin	G	%	F	%	M	%	Total	%	Pearson's Chi-squared test
Computational Statistics	native	10	3.02%	4	1.21%	6	1.81%	20	6.04%	X-squared = 0.0062678, df = 1, p-value = 0.9369
	non-native	15	4.53%	5	1.51%	7	2.11%	27	8.16%	
Humanities	native	15	4.53%	21	6.34%	20	6.04%	56	16.92%	X-squared = 0.00010089, df = 1, p-value = 0.992
	non-native	40	12.08%	23	6.95%	22	6.65%	85	25.68%	
Biology and Health	native	40	12.08%	27	8.16%	18	5.44%	85	25.68%	X-squared = 0.053931, df = 1, p-value = 0.8164
	non-native	29	8.76%	19	5.74%	10	3.02%	58	17.52%	
				X-squared = 3.9973, df = 5, p-value = 0.5498						
Total	native	65	19.64%	52	15.71%	44	13.29%	161	48.64%	X-squared = 2.7339, df = 2, p-value = 0.2549
	non-native	84	25.38%	47	14.20%	39	11.78%	170	51.36%	

Table 63 and Figure 34 present the quantitative data for the F5 *Drawing conclusions* function within the corpus. In the fields of CS and H, NN writers employ this function more frequently, with 27 instances (8.16%) in CS and 85 instances (25.68%) in H. In contrast, the N writers utilize the F5 *Drawing conclusions* function in 20 instances (6.04%) in CS abstracts and 56 instances (16.92%) in H abstracts. In the field of BH, N writers lead with 85 instances (25.68%), while the NN writers utilize the F5 function in 58 instances (17.52%). The most significant quantitative difference in the use of the F5 *Drawing conclusions* function, concerning the gender parameter, is observed in the BH field, where F authors utilize this function more frequently. Specifically, there are 27 instances in NF abstracts (8.16%) compared to 18 instances in NM abstracts (5.44%). Similarly, there are 19 instances in the NNF abstracts (5.74%) and 10 instances in the NNM abstracts (3.02%).

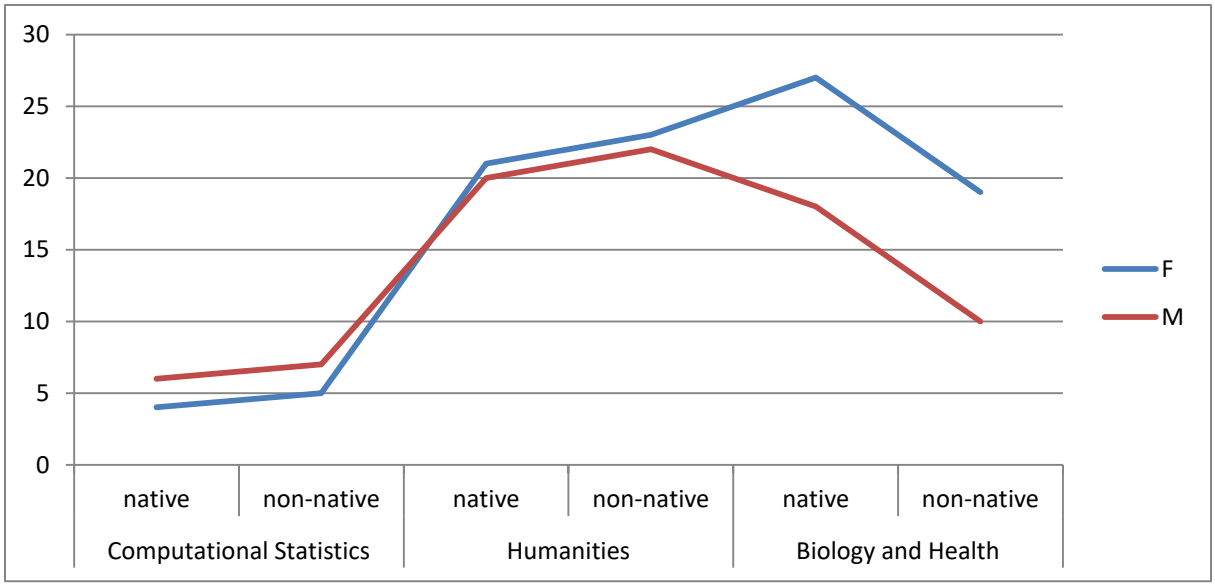

Figure 34 Drawing conclusions function: quantitative use by gender and speaker origin

Moreover, a similar tendency can be observed in the field of Humanities (H), although on a smaller scale. For the F authors, 21 native (6.34%) and 23 non-native (6.95%) abstracts have been identified. Conversely, the M authors have utilized the F5 *Drawing conclusions* function in 20 native (6.04%) and 22 non-native (6.65%) abstracts. The Pearson's Chi-squared test scores for significance in the quantitative variation of the fields concerning gender are as follows: 0.93, 0.99, 0.81, and 0.25. Additionally, the test for significance in quantitative differentiation across the academic fields yields a p-value of 0.54. None of these tests are significant enough to reject the null hypothesis. However, it can be stated that there is no substantial difference between the research groups within the field of Humanities. Although minor differences (0.9%) exist, this finding can provide a basis for further investigation into conference abstracts.

4.9.2 Vocabulary of Evaluation Functions

Table 64 Reviewing functions quantitative use in reference to academic field, gender and speaker origin

Academic field	speaker origin	G	%	F	%	M	%	Total	%	Pearson's Chi-squared test
Computational Statistics	native	1	4.00%	3	12.00%	2	8.00%	6	24.00%	X-squared = 0.73661, df = 1, p-value = 0.3907
	non-native	3	12.00%	1	4.00%	5	20.00%	9	36.00%	
Humanities	native	1	4.00%	1	4.00%	2	8.00%	4	16.00%	X-squared = 1.3333, df = 1, p-value = 0.2482
	non-native	3	12.00%	1	4.00%	0	0.00%	4	16.00%	
Biology and Health	native	0	0.00%	0	0.00%	2	8.00%	2	8.00%	NA
	non-native	0	0.00%	0	0.00%	0	0.00%	0	0.00%	
					NA					
Total	native	2	8.00%	4	16.00%	6	24.00%	12	48.00%	X-squared = 2.7219, df = 2, p-value = 0.2564
	non-native	6	24.00%	2	8.00%	5	20.00%	13	52.00%	

Table 64 and Figure 35 illustrate the quantitative application of the F6 *Reviewing function* within the analyzed corpus. The Pearson's Chi-squared test results for all researched disciplines are as follows: 0.39 for CS and 0.24 for H. There appears to be a general trend among M authors to employ this function more frequently, with 11 instances (44%), compared to only six instances (24%) observed in the opposite group.

The p-value for the total number of texts used in the corpus and research categories is 0.25. Due to insufficient data, it was not possible to run the Pearson's Chi-squared test for significance for the BH and for the cross-field quantitative differentiation. All chi approximations related to F6 *Reviewing function* cannot be considered statistically significant as they yield p-values higher than 0.05 and are based on a small sample size, with digits fewer than 5.

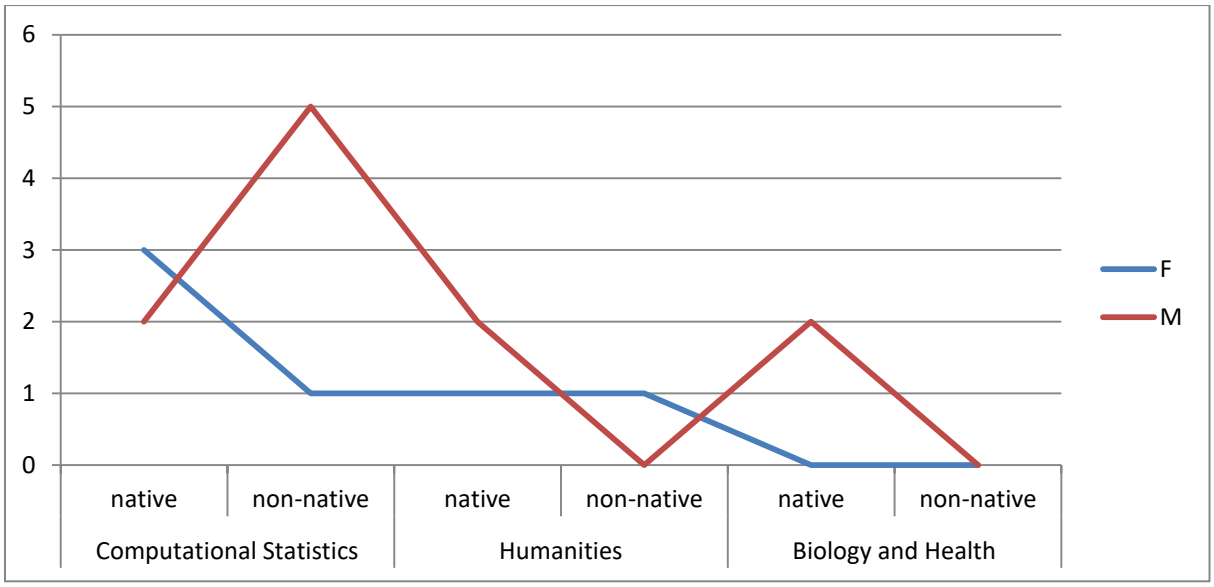

Figure 35 Rewieving function: quantitative use by gender and speaker origin

Additionally, both native and non-native writers utilized the function in a comparable number of texts, specifically, 12 (48%) and 13 (52%), respectively. The highest concentration of examples is observed in the CS field, with 4 (16%) instances attributed to F authors and seven (28%) to M authors. In the H field, there are two examples for each gender (8%). In the BH field, only two examples (8%) have been identified for NM writers.

Table 65 Evaluating function quantitative use in reference to academic field, gender and speaker origin

Academic field	speaker origin	G	%	F	%	M	%	Total	%	Pearson's Chi-squared test
Computational Statistics	native	8	2.71%	4	1.36%	11	3.73%	23	7.80%	X-squared = 0.067812, df = 1, p-value = 0.7945
	non-native	10	3.39%	6	2.03%	10	3.39%	26	8.81%	

Humanities	native	12	4.07%	22	7.46%	19	6.44%	53	17.97%	X-squared = 0.04997, df = 1, p-value = 0.8231
	non-native	28	9.49%	21	7.12%	22	7.46%	71	24.07%	
Biology and Health	native	38	12.88%	14	4.75%	16	5.42%	68	23.05%	X-squared = 4.5651, df = 1, p-value = 0.03263
	non-native	27	9.15%	21	7.12%	6	2.03%	54	18.31%	
				X-squared = 12.89, df = 5, p-value = 0.02443						
Total	native	58	19.66%	40	13.56%	46	15.59%	144	48.81%	X-squared = 1.7224, df = 2, p-value = 0.4227
	non-native	65	22.03%	48	16.27%	38	12.88%	151	51.19%	

Table 65 and Figure 36 present quantitative corpus data regarding the use of the F7 *Evaluating* function, analyzed in relation to combined parameters of gender and speaker origin across all academic fields. The highest frequency of native abstracts employing the F7 *Evaluating* function is observed in the BH field, with 68 instances (23.05%). It follows that the NNF and NM writers utilize the evaluating function more frequently than other groups in both the BH and CS fields. In the BH field, this trend is statistically significant, as indicated by a Pearson's Chi-squared test p-value of 0.03. Moreover, the Pearson's Chi-squared test for significance in the CS field yields a p-value of 0.79, while the H field shows a p-value of 0.82. Additionally, the chi-square approximation for the total quantities across the combined gender/speaker origin parameters results in a p-value of 0.42. Finally, the Pearson's Chi-squared test for the combined parameters of gender, speaker origin, and academic field yields a p-value of 0.02, indicating statistical significance and serving as evidence of this relationship.

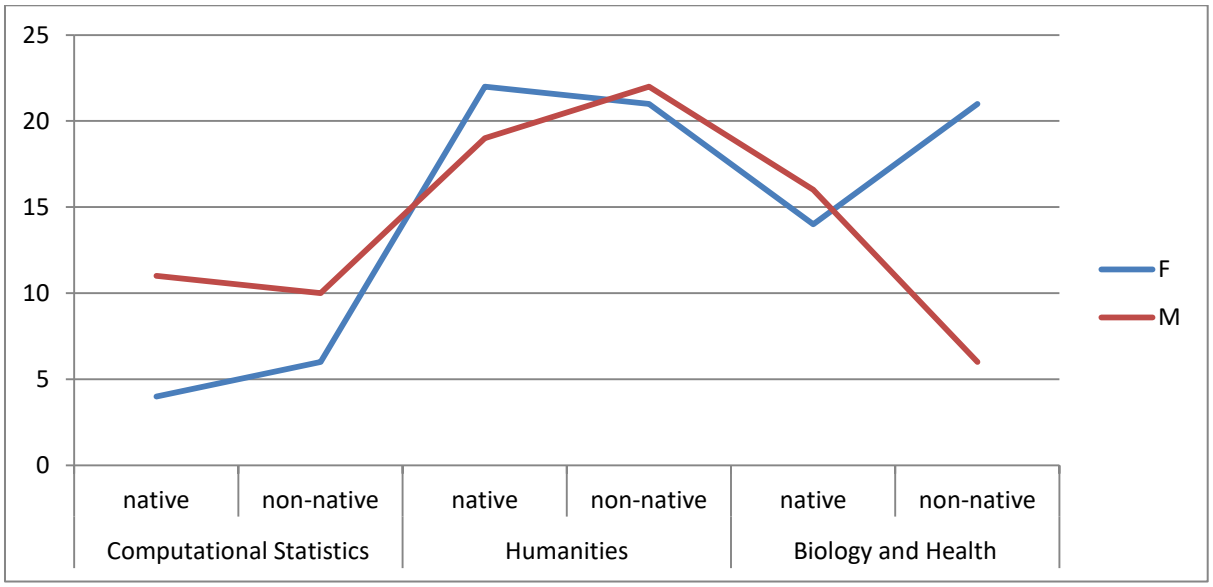

Figure 36 Evaluating function: quantitative use by gender and speaker origin

Additionally, non-native writers in the H field utilize the F7 *Evaluating* function most frequently, as evidenced by its presence in 71 abstracts (24.07%). The distribution of F7

Evaluating examples in the F and M authored abstracts is nearly equal in the H field: 22 native (7.46%), 21 non-native (7.12%) abstracts of F authorship compared to 19 native (6.44%) and 22 non-native (7.46%) abstracts of M authorship.

4.9.3 Vocabulary of Analysis Functions

Table 66 Organizing order of information function quantitative use in reference to academic field, gender and speaker origin

Academic field	speaker origin	G	%	F	%	M	%	Total	%	Pearson's Chi-squared test
Computational Statistics	native	3	3.61%	1	1.20%	3	3.61%	7	8.43%	NA
	non-native	2	2.41%	0	0.00%	0	0.00%	2	2.41%	
Humanities	native	4	4.82%	2	2.41%	7	8.43%	13	15.66%	X-squared = 3.8222, df = 1, p-value = 0.05058
	non-native	16	19.28%	9	10.84%	3	3.61%	28	33.73%	
Biology and Health	native	6	7.23%	9	10.84%	4	4.82%	19	22.89%	X-squared = 0.98194, df = 1, p-value = 0.3217
	non-native	10	12.05%	1	1.20%	3	3.61%	14	16.87%	
				NA, (X-squared = 8.3726, df = 3, p-value = 0.03891)						
Total	native	13	15.66%	12	14.46%	14	16.87%	39	46.99%	X-squared = 8.5996, df = 2, p-value = 0.01357
	non-native	28	33.73%	10	12.05%	6	7.23%	44	53.01%	
				X-squared = 1.061, df = 1, p-value = 0.303						

Table 66 and Figure 37 display the quantitative use of the F9 *Organizing order of information* function. The H field demonstrates the highest number of abstracts utilizing this function, particularly among the NN authors: 28 instances (33.73%). In contrast, the N authors contribute 13 abstracts (15.66%). Conversely, an opposite trend is observed in the BH field, with 19 texts (22.89%) by the N authors and 14 texts (16.87%) by the NN authors. Pearson's Chi-squared tests indicate significance in the overall number of F9 *Organizing order of information* uses concerning the combined parameters of gender (F, M, and G) and speaker origin, with a p-value of 0.01. However, the Pearson's Chi-squared test conducted solely for the F and M groups yields a p-value of 0.30, which does not reject the null hypothesis. In the CS field, there is insufficient data to run a significance test. Although the p-value for the H field is 0.05, this result may be distorted due to the test being conducted on small sample sizes, potentially leading to an incorrect output. Nonetheless, the differentiation in the H field offers a promising basis for further research. Lastly, the BH field significance test result is 0.32. As the conducted tests suggest, the data cannot be considered statistically significant in this study. However, the F9 *Organizing order of information* function may provide a valuable foundation for additional research and yield interesting outcomes in future analyses.

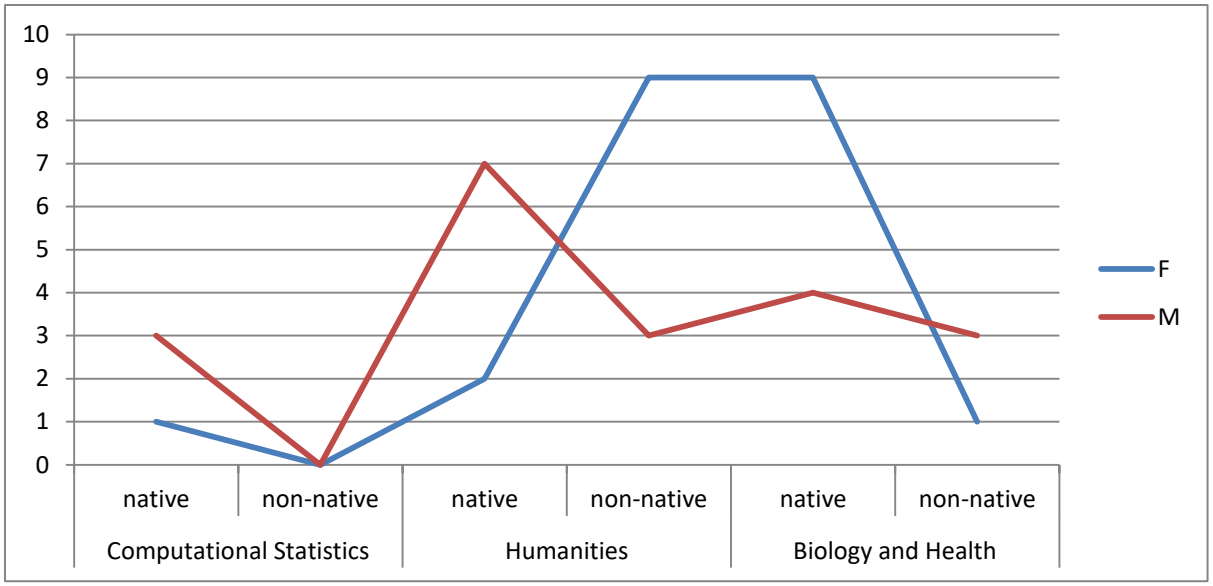

Figure 37 Organizing order of information function: quantitative use by gender and speaker origin

As illustrated in Figure 37, the NNF and NM writers in the H field utilize the F9 *Organizing order of information* function in nine instances (10.84%) and seven instances (8.43%), respectively. In the BH field, the N writers demonstrate a higher frequency of F9 *Organizing order of information* function use, with nine instances in female-authored abstracts (10.84%) and four instances in male-authored abstracts (4.82%). Notably, there are no recorded instances of F9 *Organizing order of information* function use in the NNF and NNM abstracts within the CS field. However, limited evidence suggests that the NM authors do employ the F9 *Organizing order of information* function, with three instances (3.61%).

Table 67 Reinforcing research process function quantitative use in reference to academic field, gender and speaker origin

Academic field	speaker origin	G	%	F	%	M	%	Total	%	Pearson's Chi-squared test
Computational Statistics	native	4	6.56%	3	4.92%	6	9.84%	13	21.31%	X-squared = 0.27692, df = 1, p-value = 0.5987
	non-native	6	9.84%	2	3.28%	7	11.48%	15	24.59%	
Humanities	native	0	0.00%	0	0.00%	4	6.56%	4	6.56%	X-squared = 2.7533, df = 1, p-value = 0.09706
	non-native	7	11.48%	5	8.20%	2	3.28%	14	22.95%	
Biology and Health	native	5	8.20%	2	3.28%	1	1.64%	8	13.11%	X-squared = 0.035556, df = 1, p-value = 0.8504
	non-native	2	3.28%	3	4.92%	2	3.28%	7	11.48%	
				X-squared = 8.5796, df = 5, p-value = 0.1271						
Total	native	9	14.75%	5	8.20%	11	18.03%	25	40.98%	X-squared = 1.2228, df = 2, p-value = 0.5426
	non-native	15	24.59%	10	16.39%	11	18.03%	36	59.02%	

Table 67 and Figure 38 show the quantitative usage of the F10 *Reinforcing research process* function within the corpus. There is an observable tendency for non-native writers to employ this function more frequently, with a total of 36 instances (59.02%), distributed as follows: 15

instances (24.59%) in the CS field, 14 instances (22.95%) in the H field, and seven instances (11.48%) in the BH field. The Pearson's Chi-squared test results for the specified fields are as follows: p-value of 0.59 for CS, 0.09 for H, 0.85 for BH, and 0.54 for the total quantities. Regrettably, none of these p-values are statistically significant within the scope of this study. The chi-square approximation for the usage of the F10 *Reinforcing research process* function, considering the combined parameters of academic discipline and gender, resulted in the p-value of 0.12. Furthermore, all Pearson's Chi-squared tests related to academic fields and cross-discipline analyses may be inaccurate due to insufficient quantitative data (values lower than 5). Overall, the F10 *Reinforcing research process* function appears to be a promising topic for further research on abstracts or other genres of academic texts, particularly in the H field, where it shows potential variation concerning the speaker origin parameter.

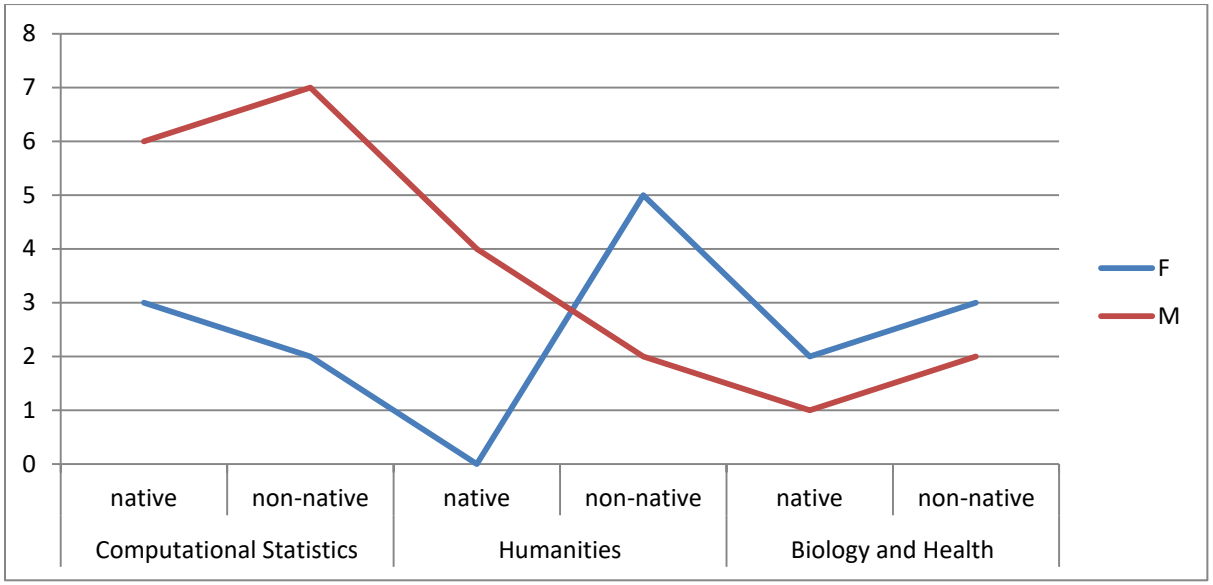

Figure 38 Reinforcing research process function: quantitative use by gender and speaker origin

Regarding the gender parameter, the quantitative distribution of F10 *Reinforcing research process* function examples in the CS field indicates that the M authors utilize the function more frequently, with six instances (9.84%) for the NM authors and seven instances (11.48%) for the NNM authors. Similarly, in the H field, there are four instances (6.56%) for the NM authors and two instances (3.28%) for the NNM authors, in contrast to five instances (8.20%) for the NNF authors. In the BH field, an opposite trend seems to be present, though the data quantity is insufficient to conclusively support this observation. Overall, when considering the gender parameter, the total number of F10 *Reinforcing research process* function examples in the corpus is higher for M authors in both speaker origin categories, totaling 22 instances (36.06%), with 11 instances for the NM authors and 11 instances for the NNM authors.

Table 68 Academic Language Functions quantitative use in reference to speaker origin in all academic fields

Function/Academic Field	Computational Statistics		Humanities		Biology and Health		Total
	quantity	%	quantity	%	quantity	%	
F1	42	13.46%	143	45.83%	127	40.71%	**312**
N	27	8.65%	62	19.87%	75	24.04%	164
NN	15	4.81%	81	25.96%	52	16.67%	148
F2	88	21.15%	165	39.66%	163	39.18%	**416**
N	41	9.86%	73	17.55%	90	21.63%	204
NN	47	11.30%	92	22.12%	73	17.55%	212
F3.1	56	18.54%	101	33.44%	145	48.01%	**302**
N	28	9.27%	42	13.91%	89	29.47%	159
NN	28	9.27%	59	19.54%	56	18.54%	143
F3.2	65	16.62%	154	39.39%	172	43.99%	**391**
N	33	8.44%	65	16.62%	97	24.81%	195
NN	32	8.18%	89	22.76%	75	19.18%	196
F3.3	38	25.50%	77	51.68%	34	22.82%	**149**
N	19	12.75%	25	16.78%	22	14.77%	66
NN	19	12.75%	52	34.90%	12	8.05%	83
F4	12	5.24%	66	28.82%	151	65.94%	**229**
N	3	1.31%	26	11.35%	82	35.81%	111
NN	9	3.93%	40	17.47%	69	30.13%	118
F5	47	14.20%	141	42.60%	143	43.20%	**331**
N	20	6.04%	56	16.92%	85	25.68%	161
NN	27	8.16%	85	25.68%	58	17.52%	170
F6	15	60.00%	8	32.00%	2	8.00%	**25**
N	6	24.00%	4	16.00%	2	8.00%	12
NN	9	36.00%	4	16.00%	0	0.00%	13
F7	49	16.61%	124	42.03%	122	41.36%	**295**
N	23	7.80%	53	17.97%	68	23.05%	144
NN	26	8.81%	71	24.07%	54	18.31%	151
F9	9	10.84%	41	49.40%	33	39.76%	**83**
N	7	8.43%	13	15.66%	19	22.89%	39
NN	2	2.41%	28	33.73%	14	16.87%	44
F10	28	45.90%	18	29.51%	15	24.59%	**61**
N	15	24.59%	4	6.56%	8	13.11%	27
NN	13	21.31%	14	22.95%	7	11.48%	34

Table 68 and Figure 39 illustrate the quantitative differences in the use of academic language functions between the N and NN authors across the researched corpus in various academic fields. It can be observed that the quantities for the NN authors are highest in the H field, except for the F6 *Reviewing* function, where the highest figures are recorded in the CS field. Additionally, the highest figures for the N authors are present in the BH field.

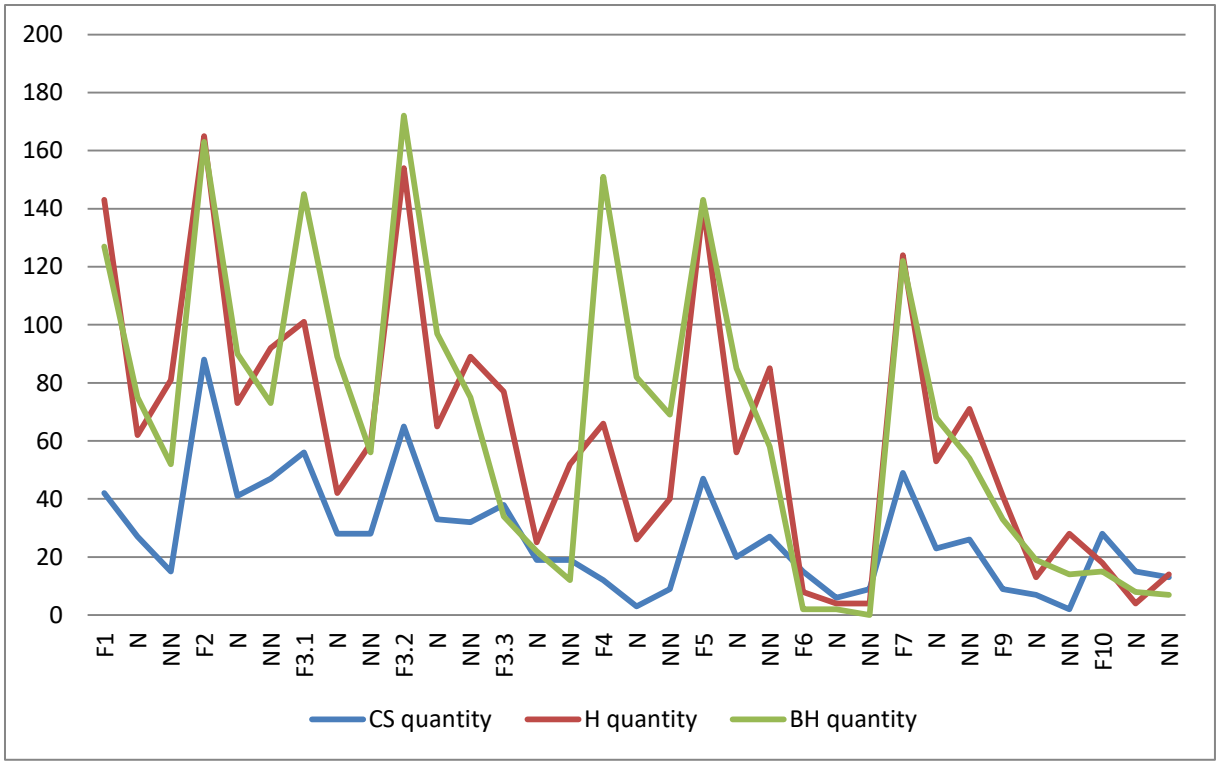

Figure 39 Academic language functions: quantitative use compared to the parameter of speaker origin in all academic fields

It is worth noting that the CS field shows equal figures for both the N and NN writers in the following functions: F3.1 *Presenting methodology* and F3.3 *Simulate situation/develop model*. Moreover, there are similarities between the N and NN authors in the CS field for the F3.2 *Design the experiment*, F7 *Evaluating*, and F10 *Reinforcing the research process* functions, with 33 and 32, 23 and 26, as well as 15 and 13 instances, respectively, for N and NN authors.

Table 69 Pearson's Chi-squared test for significance in academic functions use in reference to all native and non-native authors across all academic fields

F1	**CS**	**H**	**BH**	**Pearson's Chi-squared test**
N	27	62	75	X-squared = 9.3224, df = 2, p-value = 0.009455
NN	15	81	52	
F2	CS	H	BH	
N	41	73	90	X-squared = 4.2177, df = 2, p-value = 0.1214
NN	47	92	73	
F3.1	CS	H	BH	
N	28	42	89	X-squared = 9.5509, df = 2, p-value = 0.008434
NN	28	59	56	
F3.2	CS	H	BH	
N	33	65	97	X-squared = 6.5671, df = 2, p-value = 0.0375
NN	32	89	75	
F3.3	CS	H	BH	
N	19	25	22	X-squared = 10.607, df = 2, p-value = 0.004974
NN	19	52	12	

F4	CS	H	BH	
N	3	26	82	X-squared = 6.8814, df = 2, p-value = 0.03204
NN	9	40	69	
F5	CS	H	BH	
N	20	56	85	X-squared = 11.869, df = 2, p-value = 0.002646
NN	27	85	58	
F6	CS	H	BH	
N	6	4	2	X-squared = 2.5641, df = 2, p-value = 0.2775
NN	9	4	0	(may be insufficient)
F7	CS	H	BH	
N	23	53	68	X-squared = 4.2394, df = 2, p-value = 0.1201
NN	26	71	54	
F9	CS	H	BH	
N	7	13	19	X-squared = 8.7537, df = 2, p-value = 0.01256
NN	2	28	14	(may be insufficient)
F10	CS	H	BH	
N	15	4	8	X-squared = 5.028, df = 2, p-value = 0.08094
NN	13	14	7	

Table 69 and Figure 40 present the Pearson's Chi-squared tests for significance based on the quantitative data obtained during the academic function research, comparing native and non-native groups of writers across the researched academic fields. The most significant outcomes of the statistical testing are observed in F1 (0.009), F3 (0.004-0.03), F4 (0.03), F5 (0.002), and F9 (0.01), as they fall below the significance threshold of 0.05. However, it should be noted that the last test may be insufficient due to a small data sample.

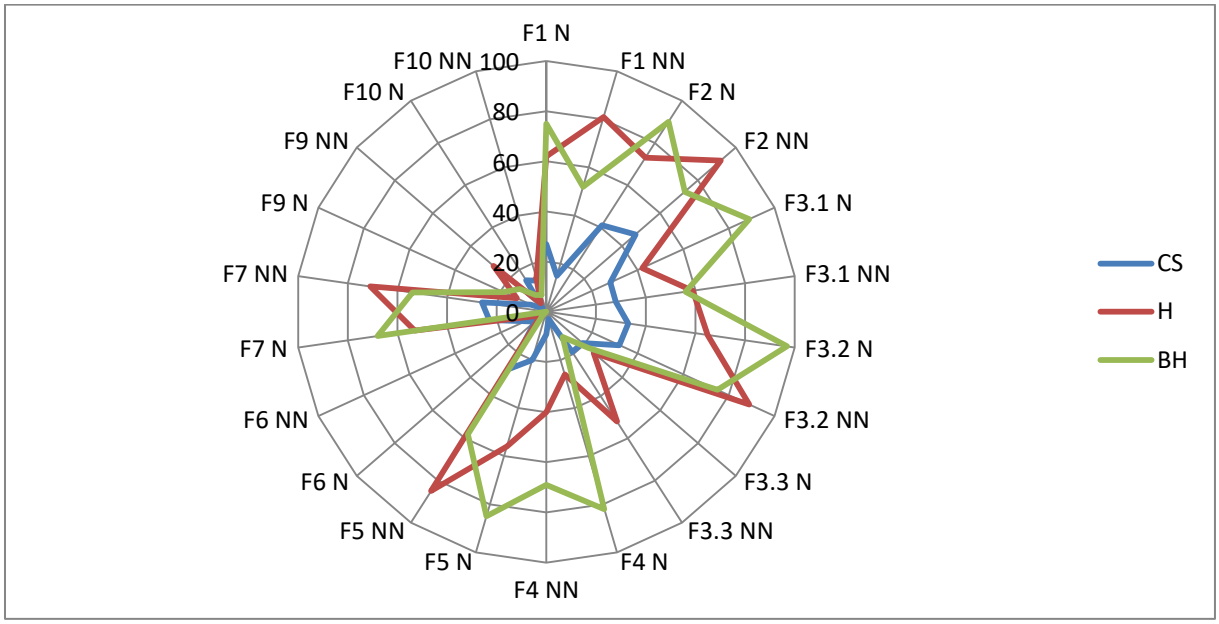

Figure 39 Academic functions use compared to the parameter of speaker origin in all academic fields

For the first, third, fourth, fifth, and ninth functions, i.e., formulating, analyzing, reporting results, drawing conclusions and organizing order of information, native authors tend to use these functions more frequently in the BH abstracts. Conversely, in the H field, the opposite trend is observed for the NN authors. Interestingly, the use of the third function in the CS field shows equal figures for both the N and NN authors. Additionally, in the case of the first function, CS field trends appear to correlate with those in the BH field, whereas trends for the fifth function seem to align with those in the H field. Nevertheless, across all functions in the H field, whether statistically significant or not, except for F6, the NN writers tend to use more functions of each type. In contrast, N writers lead in the use of all function types in the BH field.

Discussion

Parts of speech discussion

The analysis of conference abstracts reveals the following outcomes regarding the use of particular parts of speech. For nouns, the mean value for the entire corpus is 106.79. Female authors average 98.69 nouns, while male authors average 110.90 nouns, with the group mean value being 109.59. In the verbs category, the mean value for the whole corpus is 32.19 verbs, with female authors averaging 32.53 verbs and male authors averaging 30.56 verbs, resulting in a group mean value of 33.09. Section 5.3 details the use of adjectives, showing a mean value of 31.235 for the entire corpus, with female authors averaging 30.12 adjectives and male authors averaging 30.25. In comparison, the group mean value stands at 32.69 adjectives. Lastly, the analysis of adverbs indicates a mean value of 8.785 for the corpus, with female authors averaging 8.57 adverbs and male authors averaging 8.97, resulting in a group mean value of 8.80 adverbs.

The Academic Word List (AWL) division section presents the following findings: Female authors use more AWL key nouns (349) than male authors (321), with the key token ratio slightly higher at 1408 to 1373. The frequency of AWL nouns used in the corpus is nearly identical when considering the speaker origin parameter, with 256/715 types and tokens used by female authors and 249/727 by male authors. Non-native female writers use 268/693 AWL noun tokens and types, whereas male writers use slightly fewer, at 244/646.

AWL verb types in the corpus show that female authors use significantly more tokens than male authors (1085 to 946), although the number of key types used is nearly equal (354 for females and 346 for males). Interestingly, male writers appear to use verbs with prefixes more frequently than female writers; however, there is insufficient quantitative data to verify this tendency. Native writers use almost the same number of verbs, with female authors using 257 and male authors using 255, corresponding to 566 and 517 tokens, respectively. The non-native division shows a slight quantitative difference in verbs used by female writers, who tend to use more tokens (519) and types (247) compared to male writers' texts (429 tokens and 233 types), with the number of types differing by only 14 verbs. Additionally, female authors use more past participle verbs than male authors, with values of 10 to 7, 11 to 6, and 8 to 4 based on the first 20 most frequent verbs used in the corpus. This finding aligns with Dunat's (2019a) studies on text structure and gendered lexical items, albeit in a different discourse context.

Adjectives in the corpus show slight variations in AWL vocabulary use between

female and male authors and native and non-native authors. The AWL adjectives section indicates that male authors tend to use more adjectives than female authors, with a token and type ratio of 702/169 compared to 656/175. This finding concurs with Dunat's (2018) research. Non-prefixed adjectives are more frequently used by female authors (3.20%) than by male authors (2.42%). Conversely, regarding the speaker origin parameter, male authors use more adjectives, with 370 tokens compared to 322 used by female authors, and the number of types differs by four points (127 for males and 123 for females). Among non-native writers, female authors use more types of adjectives (133) compared to their male counterparts (124), though both groups use nearly the same number of tokens (334 for females and 332 for males). For further information on AWL adjectives used in other corpus settings, see Dunat (2020).

The corpus analysis of adverbs shows that male authors use adverbs more frequently than female authors, with a type and token ratio of 58/129 compared to 45/112. Native groups of writers use 36 and 40 AWL adverbs in the female and male groups, respectively, with the number of tokens differing by only four points (67 and 71 instances). Lastly, the use of adverbs among non-native writers shows a tendency towards more frequent use by male authors, with 36 types and 58 tokens, whereas female authors use 29 types of adverbs and 45 tokens.

Academic field division discussion

It is significant to note that the quantities given in brackets are normalized to the same number of texts and rounded accordingly, from 0.5 and above to 1, to represent the average quantitative use by all research groups equally.

In all researched academic fields, women use AWL nouns more frequently than men. Notably, the highest number of AWL nouns is used by females in the Humanities field, with a type and token ratio of 258/734 (86/245 normalized) instances. In comparison, male writers use more tokens but fewer AWL key nouns, with 244/758 (81/253) instances. In the Computational Statistics field, males use AWL noun tokens more frequently, with 132/251 (66/126) instances, whereas females use statistically more types, with 77/126 instances. Additionally, in the Biology & Health field, females use 206/548 (69/183) AWL nouns, while men use 168/364 (84/182); the number of tokens is almost identical. Similar tendencies are observed in the speaker-origin native groups, where native females in the Humanities field use fewer noun tokens than men, with 152/311 (101/207) and 177/373 (118/249) instances, respectively. This trend is also seen in the Biology & Health field, where females use 164/342

(109/228) nouns, while males use 124/241 types and tokens. Conversely, in Computational Statistics, female writers exhibit 50/62 (100/124) instances, compared to 75/113 instances for males. Interestingly, in the non-native groups, females use more types in all fields and more tokens in Humanities and Biology & Health. Specifically, females use 54/64 (108/128) in Computational Statistics, 193/423 (129/282) in Humanities, and 118/206 (79/137) in Biology & Health, while males use 95/138, 177/385 (118/257), and 83/123, respectively.

AWL verb usage in academic fields, divided by gender and speaker origin, indicates that females use more verb types overall (Computational Statistics and Humanities), with a similar trend for tokens in all fields except Computational Statistics. For females, verb usage includes 57/78 in Computational Statistics, 259/548 (86/183) in Humanities, and 199/459 (66/153) in Biology & Health. For males, the usage is 112/164 (56/82) in Computational Statistics, 247/509 (82/170) in Humanities, and 152/273 (76/137) in Biology & Health. Additionally, speaker origin classification reveals that males use more verb types and tokens in Computational Statistics and more types in Biology & Health, with 79/90 (40/45) and 190/176, respectively. Females, in the same order, use 37/41 and 142/265 (95/177). In Humanities, females use 166/260 (111/173), while males use 157/251 (105/167). Among non-native writers, females use AWL verbs more frequently than males, with 36/37 (72/74) in Computational Statistics, 170/288 (113/192) in Humanities, and 118/194 (79/129) in Biology & Health. Males use 58/74, 164/258 (109/172), and 74/97 in the same fields. Lastly, females use more past participle verbs across fields, with higher occurrences in Computational Statistics and Biology & Health based on the first 20 most frequent nouns used. Information about correlations between the vocabulary of the research process used in the conference abstracts can be found in Dunat (2021), and Polish conference abstracts in Dunat (2023).

Regarding adjectives, the analysis shows that men use adjectives more frequently across all fields. In Humanities and Biology & Health, males exhibit higher quantitative use of different adjective types, with 130/391 (43/130) and 80/177 (40/89) instances, respectively. Females show 131/340 (43/113) and 102/252 (34/84) type/token ratios. In Computational Statistics, males use 64/134 (32/67), while females use 45/64, with a slightly higher number of total tokens used by males. Native writers' labeling shows that women use more adjectives overall, with 28/35 (56/70) for females and 47/63 for males in Computational Statistics. In Humanities, males use 94/196 (63/131) and females use 130/337 (87/225). In Biology & Health, females use 102/252 (68/168) while males use 52/111. Non-native texts demonstrate that females use more adjectives overall, particularly in Computational Statistics (27/29 (54/58) for females and 40/71 for males). In Humanities, females use 97/183 (65/122) types

and tokens, while males use 95/195 (63/130). In Biology & Health, females use 67/122 (45/81) while males use 44/66.

Adverbs, analyzed with reference to gender, show that male writers use more adverbs overall, with 14/21 (7/11), 41/64 (14/22), and 27/41 (14/21) instances. For female writers, the frequencies are 5/8, 29/44 (10/15), and 26/59 (9/20). Native and non-native groups show the same trend across all fields. In Computational Statistics, native female writers use 5/5 (10/10) adverbs, and non-native females use 3/3 (6/6), while native males use 10/12 and non-native males use 7/9. In Humanities, native males use 26/33 (17/22) and non-native males use 26/31 (17/22). Native female writers use 16/20 (11/13) adverbs and non-native females use 19/20 (12/13). In Biology & Health, native males use 17/23 and non-native males use 11/19 (14/17), while native females use 21/38 (14/25) and non-native females use 11/19 (7/13).

Academic language functions discussion

The analysis of academic language functions used in the corpus indicates that the most frequent functions across all fields are F2 *Investigating*, F3.2 *Design the Experiment*, and F1 *Formulating*, respectively utilized in 416, 391, and 389 texts. When considering the academic field parameter, the most common function in Computational Statistics is the Investigating function, used in 88 texts. Similarly, in Humanities, it is most frequently used in 165 texts. In the Biology & Health field, the Design the Experiment function is the most prevalent, utilized in 172 texts.

The number of functions used in abstracts within the field of Computational Statistics varies between 1 and 8 functions. Male authors typically use 2 (10 example texts) or 4 (11 example texts) functions per text, while female authors most frequently use 4 (7 example texts) functions. In the field of Humanities, the frequency ranges between 2 and 9 functions overall. Female authors use 6 (14), 5 (13), or 4 (13) functions, whereas male authors most commonly use six functions (18 instances). In the Biology & Health field, the total frequency similarly ranges between 2 and 9 functions per text. Male writers use four functions (14 instances), while female writers use six functions (18 instances). These observations indicate that female and male writers in the Humanities field employ an equal number of functions, with only minor differences for functions 3.2 and 3.3. Furthermore, function 3.3 is more frequently used by female writers in Computational Statistics and by male writers in Biology & Health, on average, and is adapted to functions 9 and 10.

The academic language functions sections reveal the following trends in the quantitative use of the given functions. Firstly, the p-value of 0.09 for Humanities and

Biology & Health fields corresponds to the differences between male native and non-native writers in Humanities, where non-native writers use the F1 *Formulating* function more frequently. In Biology & Health, females utilize the F1 function more often. Considering the speaker origin parameter, native writers employ the F1 function more frequently. Additionally, across academic fields, the quantities are higher for non-native writers in Humanities and native writers in Biology & Health. Furthermore, p-values of 0.93 for Computational Statistics and Biology & Health suggest potential similarities between the research groups, although further research is needed to confirm this.

Secondly, the p-values of 0.91 (Computational Statistics), 0.92 (Humanities), and 0.94 (Biology & Health) suggest similarities in the quantitative use of the second (F2 *Investigating*) function. Furthermore, the p-value of 0.07 indicates a potential statistical difference across all academic fields in the use of the F3.1 *Presenting methodology* function. In the Computational Statistics field, male writers use this function more frequently, with a trend toward non-native male writers. Similarly, in the Humanities, non-native male writers lead in quantity. Conversely, in Biology & Health, female writers are observed to lead in quantity (see Table 59).

Thirdly, the statistical tests for the F3.2 *Design the experiment* function section, where the p-values equal: 0.94 (Humanities), 0.92 (Biology & Health), might suggest some similarities between the native and non-native writers' quantitative use of the F3.2 function. The p-value of 0.16 can, supposedly, mark the differences between gender quantitative use in the academic fields. Although it cannot be considered significant in this research, it can serve as a reasonable basis for further investigation.

The outcome of Pearson's Chi-squared test for the significance of differentiation between academic fields in the F3.3 *Simulate situation/develop model* function shows a p-value of 0.13, suggesting some statistical differences. Specifically, native females and non-native males use this function more frequently in Computational Statistics (see Table 61). Conversely, in the Humanities field, the frequencies are higher overall. In Biology & Health, there are no examples of native male usage; however, native females in this field, similar to Computational Statistics, use the function more frequently. The p-value of 0.95 in the Humanities field indicates that the quantitative use of this function by female and male writers is nearly identical. The smallest p-value for the F3.3 function in the Biology & Health field, 0.10, cannot be considered significant due to insufficient data.

In the F4 *Reporting Results* function, the p-value of 0.15 indicates quantitative differentiation across academic fields. This function is more frequently observed in non-native

females and native males in Computational Statistics and Biology & Health. In the Humanities field, non-native females and males use the function more frequently; however, the test results may be inadequate due to small data samples. Notably, there are more examples of the function used in native and non-native female groups in the Biology & Health field, with a p-value of 0.62 for the specific data.

P-values in the F5 *Drawing conclusions* function section show similarities in the Computational Statistics and Humanities fields, with Pearson's Chi-squared tests yielding p-values of 0.93 and 0.99, respectively. In Computational Statistics, the significance test may be distorted due to the small data sample. At the same time, Pearson's Chi-square test for significance in data differentiation across fields in this section has a p-value of 0.54.

The F6 *Reviewing* section of the research corpus did not provide sufficient data for a significance test. Therefore, it can be stated that the F6 reviewing function is insignificant for the abstract genre in the researched material. Nevertheless, it should be considered in future research across different academic fields.

It is observed that for the F7 *Evaluating* function, a p-value of 0.03 is obtained, indicating that the quantitative differences in the field of Biology & Health are statistically significant. Specifically, non-native female writers and native male writers use this function more frequently (see Table 65). Moreover, the across-field differences are also significant, with a p-value of 0.02. In the Computational Statistics field, both native and non-native male writers use this function most often, whereas in the field of Humanities, although the data varies on a small scale, native females and non-native males take lead in the qualitative usage of the seventh function.

In the F9 *Organizing order of information* function section, the Pearson's Chi-squared test value of 0.03 corresponds to differences between the Humanities and Biology & Health fields in the quantitative use of this function. Native males in both academic fields and non-native females in the Biology & Health field have the most prominent quantitative use of the ninth function. Additionally, a p-value of 0.05 is observed for the Humanities field, where non-native females and native males use the function most frequently. However, the test outcome may be insufficient due to the limited sample size of acquired data. The overall quantity of corpus examples in this section has a p-value of 0.01, which confirms the trend towards more frequent use of the organizing order of information function by native males, native females, and non-native females in the given fields.

The p-value of the Pearson's Chi-squared for F10 *Reinforcing research process* function, which refers to the differences between the academic fields, is 0.12. The test may be

insufficient due to the small data sample. Nevertheless, it shows that native females and non-native male writers use this function most often in Computational Statistics. The same is true for the Humanities, where non-native females and native males frequently use the function. Similarly, non-native female and male writers in Biology & Health use this function most often. Additionally, the p-value for the Humanities field is 0.09, which, although not conclusive, provides a basis for further investigation to determine whether non-native females and native males generally use the function more frequently.

Discussion of literature

It is hard to find studies that compare gendered language in the academic setting. The situation may be due to the fact that, as Baker (2010, p.4) states:

> "Within the field of Gender and Language then, a number of key approaches have utilized the turn to discourse, including work in discursive psychology which combines elements from conversation analysis, ethnomethodology and rhetorical social psychology. Some researchers have introduced elements of post-structuralist theory or critical discourse analysis into discursive psychology (...). Others have shown how techniques used in Conversation Analysis can be adopted for feminist research (...). feminist approaches to discourse analysis also all place emphasis on intertextuality (relations between texts), interdiscursivity (relations between discourses) and self-reflexivity (...). So while discourse analysis has become popular within Gender and Language, this has tended to be based on detailed qualitative studies using smaller excerpts of texts rather than approaches that involve techniques from Corpus Linguistics (...)."

Similar to Baker I searched for corpus-based studies in the journal *Gender and Language* and found that only 2.43% of the 164 article titles I had access to, between 2013 and 2020, turned out to be the corpus-based studies. Moreover, I checked the titles of 500 articles from Journal of English for Academic Purposes, published between 2013 and 2020. Only 3.6% of the articles use the corpus approach, while 0.8% relate to the academic vocabulary, and 0.8% concern gender. The corpus-assisted studies in the Gender and Language journal are presented by e.g., Kuznetsova, 2014 (Russian verbs); Jaworska & Hunt, 2016 (pronouns); Yating, 2019 (phrase "leftover women"); Candelas de la Ossa, 2019 (pronouns). Only one title used the word academic (Patterson, 2019) but it was a review of a book by Taylor and Lahad (2018). On the other hand, in the Journal of English for Academic Purposes, one can find corpus-assisted works of Lei & Yang (2020), Lan & Sun (2018), Pecorari, Shaw & Malmström (2019), Man & Chau (2018), Lei & Liu (2016) and Chan (2015). Barely six works is, by all means, not much for a spectrum of eight years. In sum, researching gendered vocabulary in an academic setting with the use of corpus tools is almost a non-existent topic. The present study

tried to fill this gap, which was inevitably linked with a very small number of other studies with which to compare the outcomes.

Moreover, the qualitative view on linguistic gender constructions is presented in Motschenbacher's (2009) femininity and masculinity via body part vocabulary, Ergün's (2013) feminist translation and sociolinguistics in dialogue, and Castro's (2013) gender language and translation at the crossroads of disciplines. Nevertheless, they cannot be used to compare with the data presented in this dissertation. Gendered language in use can be found in the studies of, e.g., Moser and Masterson (2013), Ma, Zhang, Anderson, Morris, Nguyen-Jahiel, Miller and Grabow (2016), all focusing on the children's gendered vocabulary. Vocabulary learning strategies with some gender elements can be found in Gu (2002). Other works concerning gender qualitative studies are: Preece's (2018) where gender meets social class, Lillis, McMullan & Tuck's (2018) gender and academic writing, Nygaard & Bahgat's (2018) measuring research productivity, Tuck's (2018) gendering of work, Kitamura's (2014) analysis on gendered bilingualism, Sicurella's "big nouns" in the language of academia.

Conversely, most of the studies on academic vocabulary focus on the distribution of the words list across some specified discipline (Gardner, 2007; Hyland & Tse, 2007; Martinez, Beck & Panza, 2009; Valipouri & Nassaji, 2013; Gardner & Davies, 2013; Wang, 2014; Paribakht & Webb, 2016). Other describe the collocations or clusters with words present on some commonly available word lists (Freddi, 2005; Durrant, 2009). Some scholars (Paquot, 2010; Durrant, 2013; Flowerdew, 2015; Gardner & Davies, 2016; Durrant, 2016; Csomay & Prades, 2018; Pathan, Memom, Memon, Shah & Magsi, 2018) specify how or to what extent the academic vocabulary or vocabulary lists influence students' writing. There are also works which underline the importance of self-mention markers, e.g., generality in student and expert epistemic stance (Aull, Bandarage & Miller, 2017), epistemic stance in science writing (Poole, Gnann & Hahn-Powell, 2019), self-mention markers in doctoral dissertations (Can & Cagnir, 2019). Gao (2016) presents the most frequent linking adverbials, comparing English Native Speakers with Chinese native Speakers in Research Articles from the fields of Physics, Computer Science, Linguistics and Management. Esfandiari & Barbary (2017) presents lexical bundles of English and Persian writers in Psychology research articles. Ackermann & Chen (2013) develop an academic collocation list based on Pearson International Corpus of Academic English. Additionally, there are also articles suggesting that corpus tools can be used for error corrections, as in Bridle (2019) and Dolgova & Mueller (2019). All these examples show that it is hard to find a study or work that could be compared with this dissertation's results, as they all underline different elements.

Conclusions

The analysis of parts of speech in conference abstracts conducted for this research yields several key findings regarding gender and speaker origin differences. Female authors demonstrate a higher frequency of noun and verb usage, whereas male authors use more adjectives and adverbs. Specifically, female authors employ a greater number of Academic Word List (AWL) nouns (349) compared to their male counterparts (321), and exhibit a slightly higher key token ratio, with 1,408 tokens compared to 1,373. The gender division for AWL noun usage within the corpus is nearly equivalent, with females utilizing 256 types and 715 tokens, while males use 249 types and 727 tokens. Non-native female authors use more types and tokens (268/693) than male authors (244/646) for AWL nouns.

When examining AWL verb usage, females use significantly more tokens than males (1,085 versus 946), although the number of verb types is nearly identical (354 for females and 346 for males). Native authors exhibit similar verb usage patterns, with females using 257 types and 566 tokens, and males using 255 types and 517 tokens. In contrast, non-native female authors tend to use more AWL verb tokens (519) and types (247) than their male counterparts (429 tokens and 233 types).

Adjective usage in the corpus shows only slight variations between female and male authors as well as between native and non-native authors, with males generally using more adjectives. Specifically, males use 702 adjective tokens compared to 656 for females, and 169 adjective types compared to 175 for females. This trend is also evident when considering the speaker origin parameter.

In the case of adverb usage, male authors again demonstrate a higher frequency, utilizing 58 types and 129 tokens, in contrast to the 45 types and 112 tokens used by female authors. The native writer subgroups follow this same trend, with males using 40 AWL adverbs compared to 36 used by females.

Regarding noun usage across various academic fields, female authors consistently use AWL nouns more frequently than male authors. In the field of Computational Statistics, however, male authors tend to use more AWL noun types/tokens (132/251) than females (66/126), although females use more noun types (77/126). Among non-native authors, females use more noun types across all fields, with the exception of token usage in the Humanities and Biology & Health fields.

AWL verb usage across academic fields, categorized by gender and speaker origin, indicates that females use a greater number of verb types, particularly in Computational

Statistics and the Humanities. This trend extends to verb tokens in all fields, except for Computational Statistics. In contrast, male authors use adjectives more frequently across all fields, with the Humanities and Biology & Health fields demonstrating more frequent use of different adjective types by males.

Adverb usage, as divided by gender, further underscores the tendency of male authors to use adverbs more frequently. For instance, in Computational Statistics, males use 14/21 adverb types/tokens (7/11 equalized), while females use 5/8 types/tokens (3/6 equalized). In Biology & Health, males use 41/64 adverb types/tokens (14/22 equalized), while females use 29/44 types/tokens (10/15 equalized). This pattern is consistent across all academic fields and for both native and non-native authors.

The academic language functions employed within the corpus reveal that the most frequent functions across all fields are F2 *Investigating*, F3.2 *Designing the experiment*, and F1 *Formulating*, appearing in 416, 391, and 389 texts, respectively. Field-specific analysis shows that F2 *Investigating* is the most common function in Computational Statistics, appearing in 88 texts, while in the Humanities, it is the predominant function in 165 texts. In the Biology & Health field, F3.2 *Designing the experiment* is the most frequently used function, occurring in 172 texts.

For the F7 *Evaluating* function, the p-value of 0.03 indicates statistically significant differences in the Biology & Health field. Non-native female writers and native male writers use this function more frequently. Additionally, differences across academic fields are significant, with a p-value of 0.02. In Computational Statistics, native and non-native male authors most frequently employ this function, while in the Humanities, native female authors and non-native male authors dominate its use.

The F9 *Organizing order of information* function also reveals significant differences between the Humanities and Biology & Health fields, as indicated by the Pearson's Chi-squared test value of 0.03. Native males in both fields, along with non-native females in Biology & Health, show the highest usage of the ninth function.

In summary, the analysis of conference abstracts reveals notable gender-based and speaker origin-related differences in the use of various parts of speech and academic language functions. Female authors tend to use more nouns and verbs, whereas male authors use more adjectives and adverbs. These patterns underscore the importance of considering gender and speaker origin in academic research, highlighting potential areas for further investigation and deeper analysis, particularly in the Humanities and Biology & Health fields.

Answering the hypotheses

The following part is presented in order to give answers to the hypotheses stated in this study:

1. It can be stated that there are differences in the frequency of occurrence of some academic language functions depending on the gender, the origin of authors, and in academic fields, as presented in section 4.9. The following functions were found to be most frequent in the academic field of Humanities: F1 Formulating (45.83%), F2 Investigating (39.66%), F3.3 Simulate situation/develop a model (51.68%), F7 Evaluating (42.03%), and F9 Organizing order of information (49.40%).
2. It can be stated that, as was expected in the preliminary studies, there are differences in academic vocabulary use (formal realizations) for gender and speaker origin, presented in Sections 4.1 to 4.4, 4.5, 5.1 and 5.2. Moreover, there are differences in types of language functions (formal realizations), in texts, for gender and speaker origin, presented in Sections 4.9 and 5.3.
3. It can be claimed that there is a relationship between the functional and formal means of the researched academic language functions, backed by the significant Pearson's Chi-square tests for:
    - The Evaluating function (F7) in the Biology & Health field has a p-value of 0.03, where non-native female writers and native male writers use the function most frequently. Moreover, for the same function, across the field differentiation has the p-value of 0.02 (see Table 65).
    - The Organizing order of information function (F9) between the Humanities and Biology & Health fields differentiation has the p-value of 0.03. Additionally, for the same function, in the Humanities field, the p-value equals 0.05, where the texts of non-native females and native males have the highest quantity. However, the outcome of the two abovementioned tests may be distorted due to the small sample size. Lastly, the Pearson's Chi-squared test for difference in overall quantities of the native and non-native, female and male writers for the ninth function shows the significant p-value of 0.01 (see Table 66).
    - The Investigating function (F1) analyzed with reference to the combined parameters of the academic field and speaker origin resulted in the p-value of 0.009. The higher frequencies of the F1 use occur for native authors in Computational Statistics and Biology & Health, and for non-native authors in

the Humanities field.

- The Analyzing function (F3.1, 3.2, 3.3) analyzed with reference to the combined parameters of the academic field and speaker origin resulted in the p-values of 0.008, 0.03, and 0.004, respectively, where the same frequency of F3 use occurs for native and non-native writers in the Computational Statistics field. Moreover, higher frequencies of the F3 use occur for non-native authors in Humanities and native authors in Biology & Health.
- The Reporting results function (F4) analyzed with reference to the combined parameters of the academic field and speaker origin resulted in the p-value of 0.03, where the higher frequencies of the F4 use occur for non-native writers in Computational Statistics and Humanities; and for native writers the higher frequencies occur in Biology & Health field.
- The Drawing conclusions function (F5) analyzed with reference to the combined parameters of the academic field and speaker origin resulted in the p-value of 0.002, where the higher frequencies of the F5 use occur for non-native writers in Computational Statistics and Humanities; and for native are presented in Biology & Health field (see Table 74).

Limitations

> "we know that a specific event either will happen or will not. The degree of probability, therefore, will be of no use after the truth about the occurrence of the event is known: probability is used as a substitute for truth so long as the truth is unknown. If the event is to happen in the future, the degree of its probability qualifies the reliability of a prediction; and if the event belongs to the past we can also regard the probability as the measure of the quality of a prediction, namely, of a prediction about possible future verifications of a past event the truth value of which is unknown. The criterion for the justification of an interpretation lies in its adequacy for the purpose of prediction." (Reichenbach, 1949, p. 367).

Predictably, the data used in this synchronic study, like in any long-term research, inevitably becomes 'aged'. Nevertheless, its validity is not compromised simply due to its age. It can serve as a basis for predicting future verifications and as a representation of a specific point in time, which will be utilized in my future research career.

It is essential to underscore certain limitations of the present study. Firstly, considerable time was invested in selecting conference books of abstracts that provided appropriate division by gender and speaker origin within three academic fields, spanning conferences across five continents. Secondly, all data were labeled and coded by a single individual using computational linguistics software. Despite meticulous efforts, some unnoticed errors may have occurred, for example, TagAnt software occasionally assigns inappropriate parts of speech to words in sentences.

Thirdly, it was not possible to verify the student–scholar parameter in the abstracts' text and compare it to Hyland & Tse (2005). Due to insufficient data obtained during inquiries, the student-scholar parameter was adjusted to scholar/non-scholar. Efforts were made to balance the number of texts by gender for this label. During the corpus preparation phase, the following results were obtained: 117 examples from females and 108 from males in the scholar group, and 32 examples from males and 23 from females in the non-scholar group. Ultimately, this category was excluded due to significant quantitative differences, specifically 392 scholar abstracts versus 87 non-scholar abstracts.

Fourthly, due to software limitations, double labeling of functions was not possible. Although preliminary research indicated that such instances constituted less than 5%, some unnoticed errors may have occurred, particularly within the F9 and F10 functions. Nonetheless, AntMover software aided in automating corpus data processing. Every effort was made to verify all data labeled by AntMover software, particularly in the context of sentences; however, some errors may still have gone unnoticed.

Comparison of the results with other studies

To some extent, the following studies can be compared to the results of this dissertation. Firstly, Csomay and Prades (2018, p. 103) present 70 most frequently occurring academic words in ESL student papers. After comparing the nouns present on Csomay and Prades (2018) list with data presented in this dissertation, one can find the following nouns present in both studies: *process, technology, role, research, analysis, culture, policy*. Looking at verbs and adjectives, one may find: *require, identify, indicate, significant,* and *economic*. Secondly, the work of Durrant (2016, p. 9) presents items "achieving the highest coverage" in the Linguistics discipline, which can be compared with the vocabulary present in the field of Humanities in this dissertation. The words present in both studies are: *focus, approach, research, context, analysis* and *process*—additionally, *structure* and *theory. Structure* is present on the native females list, while *theory* is present on the native males and all females lists. The verbs *identify,* and *focus* are used by females in present form. Conversely, males use them with past participle. Additionally, *occur* (*occurs*) has been found only on the native males list. Thirdly, Martínez, Beck & Panza (2009) study of academic vocabulary in agriculture research articles show the following similarities in the most frequently used vocabulary: *analysis, environment, response* noun (only in Computational Statistics field, male writers), and adjectives *significant* (present in Humanities and Biology & Health fields) and *similar* (in the Humanities field only). Fourthly, Hyland & Tse (2007) engineering part of the corpus relates to computational statistics vocabulary in this paper, with the following elements: *method, function, output* (only for male writers), *require* and *analyze* (only for female writers). Additionally, Hyland and Tse (2007, p. 240) work underlines that the AWL coverage of their corpus equals 9.3% in overall. While AWL coverage in abstracts, i.e., the corpus used in this research, totals 2.17%. Lastly, Gu (2002) has shown that "female students outperformed male students no matter what academic background they were from" (p. 42) in the vocabulary size by approximately 4 points in Arts and 5 points in Sciences (p=0.13). The number of types used by both female and male groups in this study shows that females use 923/3261 types/tokens, whereas males use 894/3150 (p=0.97).

Bondi's (2014) diachronic study describes an increase in evaluative adjectives. She found 80 examples in 1990, 130 examples in 2000, and 129 examples in 2010. The evaluative adjectives quantities in this study showed a further increase of adjectives in abstracts in 2016, i.e., 109 types in Computational Statistics, 269 types in Humanities, and 182 types in Biology & Health. Watinee & Siriluck (2013) noted a similar use of verbs in linguistics and applied linguistics. This study highlights that the quantitative use of verbs is similar for gender

labeling in the Humanities field. The following scholars underlined the length of the abstract, as in Busch-Lauer (2014), Tibbo (1992): 137 tokens in chemistry, 141 psychology, 80 history; Melander (1997): average of 155 tokens, 164 tokens in biology, 153 medicine, 149 linguistics; Orasan (2001): average of 175, 232 in computer science, 215 in chemistry, 196 biology, and 150 linguistics. As observed, there is a diachronic increase in the length of abstracts, not only in other scholars' works but also in this research. The length of abstracts in the researched fields concerning the labeling of the corpus used was presented in Table 17. In conclusion, the length of abstracts increases dramatically compared to other works, especially in the field of Biology & Health. Hyland and Tse (2005) emphasized the preference for nouns over adjectives use and stated that the adjectives constitute 5.2% of the total vocabulary. The percentage of adjectives in the corpus used for this study shows that the AWL adjectives constitute 0.61% of types, and 0.46% of tokens, whereas nouns constitute 1.19% of types and 0.91% of tokens; it supports Hyland and Tse's (2005) findings. Furthermore, Alonso-Almeida (2014) found adverbs to be insignificant in the genre of abstracts. While this study reveals minimal quantities of adverbs in conference abstracts (AWL adverbs constitute 0.15% of types and 0.08% of tokens), it is possible to substantiate this finding and support the claim.

Closing remarks

There are various ways to present, define and analyze the structure and rhetorical functions of abstracts, as this study illustrated in the first two chapters. Previous studies have addressed many aspects of abstracts, particularly in corpus linguistics; however none have focused on academic language functions within abstracts, especially through computational linguistics methods. Chapter Three introduced the analysis design and computational tools (R, AntConc, ProtAnt, TagAnt) employed in this research. Chapter Four, the core of the thesis, provided an analysis of the corpus data, examining differences and similarities in academic language functions concerning gender, speaker origin and academic field parameters. On this basis any scholar can replicate the investigation presented in this dissertation. The primary aims of this study were set as follows:

- Conduct a formal/structural analysis of the academic language used in abstracts by both sexes across the characterized fields and compare them.
- Perform a functional analysis of specified academic language functions in the corpus, comparing them in relation to gender, speaker origin and field of study.
- Investigate the potential connection between formal and functional aspects of specified functions in academic writing.

The quantitative computational approach, employed as the primary method for analyzing the selected data samples, along with the application of Pearson's Chi-square test and various software tools, facilitated the interpretation of specified lexis and academic language functions. This approach was instrumental in achieving the primary objectives of the study and establishing a framework for further analysis.

The analysis of parts of speech in conference abstracts in this dissertation has revealed notable gender-based and speaker origin-related differences. Female authors demonstrate a higher frequency of noun and verb usage, whereas male authors use more adjectives and adverbs. Specifically, female authors were found to use a greater number of Academic Word List (AWL) noun and verb tokens compared to their male counterparts. This pattern underscores the significance of considering gender and speaker origin in academic research, highlighting potential areas for further investigation and deeper analysis.

Considering parts of speech, it has been found that:

- Female authors use more AWL nouns (349) and verbs (1,085) compared to male authors (321 nouns and 946 verbs). Non-native female authors also use more types and tokens of AWL nouns (268/693) and verbs (519) than male authors (244/646 nouns and 429 verbs).
- Male authors tend to use more adjectives (702) and adverbs (129) than female authors (656 adjectives and 112 adverbs). This trend is consistent across both native and non-native groups.
- In Computational Statistics, male authors use more AWL noun types/tokens (132/251) than females (66/126), while females use more noun types (77/126). In the Humanities and Biology & Health fields, females use more noun types and tokens overall.

Regarding the parameters of speaker's origin and academic field, it can be claimed that:

- Native authors exhibit similar verb usage patterns, with females using 257 types and 566 tokens, and males using 255 types and 517 tokens. Non-native female authors tend to use more AWL verb tokens (519) and types (247) than their male counterparts (429 tokens and 233 types).
- Female authors consistently use AWL nouns more frequently than male authors across various academic fields. Notably, females in the Humanities field use 258/734 (86/245) instances, while male writers use 244/758 (81/253) instances.

Academic Lanaguage Functions analysis has shown that:

- The most frequent academic language functions across all fields are F2 (Investigating), F3.2 (Designing the experiment), and F1 (Formulating), respectively utilized in 416, 391, and 389 texts.
- Field-specific analysis shows that F2 (Investigating) is the most common function in Computational Statistics and Humanities, while F3.2 (Designing the experiment) is most prevalent in Biology and Health.

Following the presented hypotheses, it might be stated that significant gender and speaker origin differences exist in the frequency of certain academic language functions. For example, the F1 (Formulating) function is more frequently used by non-native writers in Humanities (45.83%) and by females in Biology & Health. Secondly, there are clear differences in academic vocabulary use based on gender and speaker origin. Females tend to use more nouns and verbs, while males use more adjectives and adverbs. This is evident in sections 4.1 to 4.4 and 5.1 to 5.3. Thirdly, significant Pearson's Chi-square tests support the relationship between functional and formal means of academic language functions. For instance, the F7 (Evaluating) function has a p-value of 0.03 in the Biology & Health field, indicating a higher frequency of use by non-native female and native male writers.

It is hard to find studies that compare gendered language in academic settings. Baker (2010) notes the focus on detailed qualitative studies using smaller excerpts of texts rather than corpus-based approaches. A review of the journal Gender and Language revealed only 2.43% of articles were corpus-based studies, while the Journal of English for Academic Purposes had just 3.6% corpus-based studies. Most studies on academic vocabulary focus on specific disciplines or word list distributions rather than gender differences. Examples include Gardner (2007) on vocabulary in specific disciplines and Hyland & Tse (2007) on word distributions in academic texts.

The present study identified several points for further investigation. The points which demonstrate similarities are listed below:

1. Across the academic fields
    a) F1 between CS and BH,
    b) F2 in all researched fields,
    c) F5 between CS and H.

2. Between gender
    a) F3.3 in Humanities.
3. Between speaker origin
    a) F3.2 in H and BH.

Also the points which demonstrate differences were indentified:

1. Across the academic fields
    a) F3.1 in all researched fields,
    b) F3.3 in all researched fields,
    c) F10 in all researched fields.
2. Between gender
    a) F3.2 in H and BH.
3. Between speaker origin
    a) F4 in CS and BH.

Moving forward, it is crucial to continue examining these patterns to foster a more equitable and representative academic landscape. There are three example routes to pursue. Firstly, further investigation into the underlying factors driving gender and speaker origin differences in academic writing. Secondly, exploring how these differences impact the reception and evaluation of academic work within various fields. And, finally, studying the influence of cultural and educational backgrounds on language use in academic writing. I will definitely explore these routes and investigate them further in my academic career.

References

Adams Smith, D. (1990). Source and derived discourse. In M. A. K. Halliday, J. Gibbons, & H. Nicholas (Eds.), *Learning, keeping and using language* (pp. 415–433). Amsterdam: John Benjamins.

Alexander, P.A., & Jetton, T.L. (2000). Learning from text: A multidimensional and developmental perspective. In M.L. Kamil, P.B. Mosenthal, P.D. Pearson, & R. Barr (Eds.), *Handbook of reading research* (Vol. 3, pp. 285–310). Mahwah, NJ: Erlbaum.

Alonso-Almeida, F. (2014) Evidential and Epistemic Devices in English and Spanish Medical, Computing and Legal Scientific Abstracts: A Contrastive Study. In M. Bondi, & R. Lorés-Sanz (Eds.) *Abstracts in Academic Discourse. Variation and Change.* (pp. 21-42). Bern: Peter Lang.

Bakhtin, M. (1981). *The Dialogic Imagination: Four Essays.* M. Holquist (Ed.). Austin, US-TX: University of Texas Press.

Baßler, H. (2003). Russische, deutsche und angloamerikanische Zeitschriftenabstracts der Soziologie: Worin unterscheiden sie sich? In H. Gruber, F. Menz, & O. Panagl (Eds.) *Sprache und politischer Wandel.* (pp. 189-212). Frankfurt: Peter Lang.

Bauer, L., Nation I.S.P. (1993). Word families. *International Journal of Lexicography 6*, 253-279.

Baumann, J. F., & Graves, M. F. (2010). What is Academic Vocabulary. *Journal of Adolescent & Adult Literacy 54*(1), 1-14.

Bazerman, C. (1984). The writing of scientific non-fiction. *Pre/Text 5*(1), 39-74.

Becher, T., & Trowler, P. R. (2001). *Academic tribes and territories: intellectual enquiry and the culture of disciplines.* Buckingham; PA: Society for Research into Higher Education & Open University Press.

Belić, A. (1956). A Letter on Linguistics. In M. Halle, H.G. Lunt, H. McLean, C.H. Van Schooneveld (Eds.), *For Roman Jakobson* (pp.31-34). Hague: Mouton & CO.

Bhatia, V. K. (1993). *Analysing genre: language use in professional settings.* London: Longman.

Bhatia, V. K. (1997a). Applied Genre Analysis and ESP. In T. Miller (Ed.), *Functional Approaches to Written Text: Classroom Applications* (pp.134-149). Washington, DC: United States Information Agency.

Biber, D. (1988). *Variation Across Speech and Writing.* Cambridge: Cambridge University Press.

Biber D., Conrad, S., & Randi, R. (1998). *Corpus Linguistics: Investigating language structure and use.* Cambridge: Cambridge University Press.

Biber, D. (2006). *University Language: A corpus-based study of spoken and written registers.* Amsterdam: John Banjamins Publishing Company.

Birkeland, H. (1956). Some Reflections on Semitic and Structural Linguistics. In M. Halle, H.G. Lunt, H. McLean, C.H. Van Schooneveld (Eds.), *For Roman Jakobson* (pp.44-51). Hague: Mouton & CO.

Bondi, M. (2014). Changing Voices: Authorial Voice in Abstracts. In M. Bondi, & R. Lorés-Sanz (Eds.), *Abstracts in Academic Discourse. Variation and Change.* (pp. 243-270). Bern: Peter Lang.

Brown, P. & Fraser, C. (1979). Speech as a marker of situation. In K. R. Scherer & H. Giles (Eds.), *Social Markers in Speech* (pp. 33–62). Cambridge: Cambridge University Press.

Brown, P. & Levinson, S. (1987). *Politeness: Some Universals in Language Usage.* Cambridge: Cambridge University Press.

Bucholtz, M. (2003). Theories of Discourse as Theories of Gender: Discourse Analysis in Language and Gender Studies. In J. Holmes & M. Meyerhoff (Eds.), *The handbook of Language and Gender.* (pp. 43-68). Oxford: Blackwell Publishing Ltd.

Busch-Lauer, I. A. (2014). Abstracts: Cross-Linguistic, Disciplinary and Intercultural Perspectives. In M. Bondi, & R. Lorés-Sanz (Eds.), *Abstracts in Academic Discourse. Variation and Change.* (pp. 43-64). Bern: Peter Lang.

Chafe, W. 1986. Evidentiality in English Conversation and Academic Writing. In W. Chafe, & J. Nichols (Eds.), *Evidentiality: The Linguistic Coding of Epistemology.* (pp. 261-272) New Jersey: Ablex.

Chamot, A.U. & O'Malley, J. M. (1990). *Learning Strategies in Second Language Acquisition*. Cambridge: Cambridge University Press

Chamot, A.U. & O'Malley, J. M. (1994). *The CALLA handbook: Implementing the cognitive academic language learning approach.* Reading: Addison-Wesley.

Cohen, A. D., Glasman, H., Rosenbaum-Cohen, P. R., Ferrara, J. and Fine, J. (1988). Reading English for specialised purposes: discourse analysis and the use of student informants. In P. Carrell, J. Devine and D.E. Eskey (Eds.), *Interactive Approaches to Second Language Reading* (pp. 152-167). Cambridge: Cambridge University Press.

Cornillie, B. (2009). Evidentiality and Epistemic Modality. On the Close Relationship between Two Different Categories. *Functions of Language 16*(1), 44-62.

Cowan J. R. (1974). Lexical and syntactic research for the design of reading materials. *TESOL Quarterly 8*(4), 389-399.

Coxhead, A. (2000). A new academic word list. *TESOL Quarterly* (34), 213–238.

Crismore, A., Markkanen, R., Steffensen, M. (1993). Metadiscourse in Persuasive Writing. *Written Communication 10*(1), 39-71.

Dahl, T. (2000). *Lexical cohesion-based text condensation: an evaluation of automatically produced summaries of research articles by comparison with author-written abstracts.* PhD thesis, Department of English, University of Bergen.

Day, R. A. (1988). *How to write and publish a scientific paper.* Cambridge: Cambridge University Press.

De Groot, A.W. (1956). Classification of Cases and Uses of Cases. In M. Halle, H.G. Lunt, H. McLean, C.H. Van Schooneveld (Eds.), *For Roman Jakobson* (pp.187-194). Hague: Mouton & CO.

Dendale, P. (1994). Devoir épistémique: marqueur modale ou évidentiel? *Langue Française* (102), 24-40.

De Saussure, F. (1983). *Course in General Linguistics.* (R. Harris, Trans.). Chicago, IL: Open Court Publishing Company. (Tenth printing).

Dixon, R. (2005). *A Semantic Approach to English Grammar.* Oxford: Oxford University Press.

Donesch-Jeżo, E. (2013). *Akademicki dyskurs medyczny w językach angielskim i polskim w ujęciu lingwistycznym i metodycznym: artykuł naukowy i jego abstrakt.* Kraków: Fundacja Zdrowia Publicznego. Uniwersyteckie Wydawnictwo Medyczne „Vesalius".

Dunat, S. (2015). *Hedging in Spoken American Academic English: Corpus Analysis of 'seem'.* [Unpublished Master's Thesis]. University of Silesia, Sosnowiec.

Dunat, S. (2018). Differences of bearable and tolerable lexemes usage in women and men written language. Corpus Study of American newspapers and magazines, based on the Corpus of Contemporary American English. In K. Feruga, A. Ostrowska-Knapik (Eds.) *Język. Tożsamość. Wychowanie.* (pp. 47-66). Praha: Verbum.

Dunat, S. (2019a). Modality and Text Structure in American Senators' Tweets: A Corpus Study of Gendered Lexical Items. *Media i Społeczeństwo* (10), 63-85.

Dunat, S. (2019b). Vocabulary Analysis: A corpus based study of 'analyze' clusters and collocates in academic and spoken discourse. In K. Feruga, A. Ostrowska-Knapik (Eds.) *Język. Tożsamość. Wychowanie II.* (pp. 43-62). Bielsko-Biała: Wydawnictwo Naukowe Akademii Techniczno-Humanistycznej.

Dunat, S. (Ed.). (2020). *A Corpus Study of Adjectives: Students' Studies on Academic English.* Bielsko-Biała: Wydawnictwo Naukowe Akademii Techniczno-Humanistycznej.

Dunat, S. (2021). A corpus-based synchronic study of correlations between the vocabulary of the research process in conference abstracts of three disciplines. In K. Feruga, A. Ostrowska-Knapik (Eds.) *Język. Tożsamość. Wychowanie IV.* (pp. 67-85). Bielsko-Biała: Wydawnictwo Naukowe Akademii Techniczno-Humanistycznej.

Dunat, S. (Ed.). (2023). *Rhetorical moves and research process vocabulary in conference abstracts: A synchronic corpus study of selected Polish books of abstracts; examples in three disciplines.* Bielsko-Biała: Wydawnictwo Naukowe Uniwersytetu Bielsko-Bialskiego.

Eckert, P., & McConnell-Ginet, S. (2003). *Language and Gender.* Cambridge: CUP.

Ehlers-Zavala, F.P. (2008). Teaching adolescent English language learners. In S. Lenski & J. Lewis (Eds.), *Reading success for struggling adolescent learners* (pp. 74–89). New York: Guilford.

Ellis, L., Hershberger, S., Field, E., Wersinger, S., Pellis, S., Geary, D., Palmer, C., Hoyenga, K., Hetsroni, A., Karadi, K. (2008). *Sex Differences: Summarising More than a Century of Scientific Research.* New York: Taylor & Francis Group.

Ervin-Tripp, S. (1972). On sociolinguistic rules: Alternation and co-occurrence. In J. J. Gumperz & D. Hymes (Eds.), *Directions in Sociolinguistics* (pp. 213–250). New York: Holt.

Evans, V. (2009). *Leksykon językoznawstwa kognitywnego.* (M. Buchta, M. Cierpisz, J. Podhorodecka, A. Gicala, J. Winiarska. Trans.) Kraków: Towarzystwo Autorów i Wydawców Prac Naukowych UNIVERSITAS. (Original text published 2007).

Ewer, J.R., & Latorre, G. (1969). *A course in basic scientific English.* London: Longman Group Limited.

Fahnestock, J. (1986). Accommodating science: The rhetorical life of scientific facts. *Written Communication* (3), 275–296.

Firth, J.R. (1956). Linguistic Analysis and Translation. In M. Halle, H.G. Lunt, H. McLean, C.H. Van Schooneveld (Eds.), *For Roman Jakobson* (pp.133-139). Hague: Mouton & CO.

Fløttum, K., Dahl, T. & Torodd, K. (2006). *Academic Voices.* Amsterdam, PA: John Benjamins.

Fløttum, K., Dahl, T. & Torodd, K. (Eds.). (2007). *Language and Discipline Perspectives on Academic Discourse.* Cambridge: Cambridge Scholars Publishing.

Flynn, E. (1988). Composing as a Woman. *College Composition and Communication 39*, 423-435.

Francis, B., Robson, J., Read, B. (2001). An Analysis of Undergraduate Writing Styles in the Context of Gender and Achievement. *Studies in Higher Education 26*(3), 313-326.

Furdal, A. (1977). *Językoznawstwo Otwarte.* Wrocław: Zakład Narodowy im. Ossolińskich – Wydawnictwo.

Gardner, D. (2013). *Exploring Vocabulary: Language in Action.* London: Routledge.

Giannoni, D. S. (2010). *Mapping Academic Values in the Disciplines: A Corpus-based Approach.* Bern: Peter Lang.

Gillaerts, P. & Van de Velde, F. (2010). Interactional Metadiscourse in Research Article Abstracts. *Journal of English for Academic Purposes, 9*(2), 128-139.

Gillaerts, P. (2014). Shifting Metadiscourse: Looking for Diachrony in the Abstract Genre. In M. Bondi, & R. Lorés-Sanz (Eds.), *Abstracts in Academic Discourse. Variation and Change.* (pp. 271-286). Bern: Peter Lang.

Grabe, W. and Kaplan, R. B. (1996). Theory and Practice of Writing: An Applied Linguistic Perspective. New York: Longman.

Graetz, N. (1985). Teaching EFL students to extract structural information from abstracts. In J.M. Ulijin and A.K. Pugh (Eds.), *Reading for professional purposes,* (pp. 123-35). Leuven, Belgium: ACCO.

Grishman, R. (1986). *Computational Linguistics: An introduction.* Cambridge: Cambridge University Press.

Halliday, M. A. K. (1978). *Language as Social Semiotic: The Social Interpretation of Language and Meaning.* London: Edward Arnold.

Halliday, M. A. K. (1985). *An introduction to functional grammar.* London: Edward Arnold.

Halliday, M. A. K. (1988). On the language of physical science. In M. Ghadessy (Ed.), *Registers of Written English* (pp. 162–78). London: Pinter.

Hartley, J. (2004). Current findings from research on structured abstracts. Journal of the Medical Library association. Vol. 93, No. 3, pp. 368-371.

Hausser, R. (2014). *Computational linguistics: Human-Computer Communication in Natural Language* (3rd ed.). Springer.

Heikkinen, H. L. T., Huttunen, R., & Kakkori, L. (1999). ‘‘Ja ta¨ma¨ tarina on tosi. . .’’ Narratiivisen totuuden ongelmasta [‘And this is a true story. . .’: The problem of narrative truth]. *Tiedepolitiikka* (4), 39–52.

Heinke, A. (2018). The Complexity of Language and Learning. Deconstructuring Teachers' Conceptions of Academic Language. *Issues in Teacher Education*. Vol. 27, No. 2, pp. 73-89.

Heinz, A. (1983). *Dzieje językoznawstwa w zarysie.* Warszawa: Państwowe Wydawnictwo Naukowe.

Herbert, R. K. (1990). Sex-based Differences in Compliment Behaviour. *Language in Society 19*, 201-224.

Herrel, A., & Jordan, M. (2001). *Fifty Strategies for Teaching English Language Learners*. New Jersey: Pearson Merril Prentice Hall

Hocket, C.F. (1956). Idiom Formation. In M. Halle, H.G. Lunt, H. McLean, C.H. Van Schooneveld (Eds.), *For Roman Jakobson* (pp.222-230). Hague: Mouton & CO.

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

Hunston, S. (1993). Evaluation and ideology in scientific writing. In M. Ghadessy (Ed.), *Register analysis: theory and practice* (pp. 57–73). London: Pinter.

Hutz, M. (1997). *Kontrastive Fachtextlinguistik für den fachbezogenen Fremdsprachenunterricht. Fachzeitschriftenartikel der Psychologie im interlingualen Vergleich.* Trier: Universitätsverlag.

Hyland, K. (2000). *Disciplinary Discourses. Social Interactions in Academic Writing.* London: Longman.

Hyland, K., Tse, P. (2006) Gender and Discipline: Exploring Metadiscourse Variation in Academic Book Reviews (pp. 177-202) in Bondi, M., Hyland. K. (Eds.) *Academic Discourse Across Disciplines.* Bern: Peter Lang.

Hyland, K. (2014). *Disciplinary Discourses. Social Interactions in Academic Writing* (5th ed.). Ann Arbor, MI: The University of Michigan Press.

Hymes, D. (1974). *Foundations in Sociolinguistics.* Philadelphia, PA: University of Pennsylvania Press.

Johnson, D., Roen, D. (1992). Complimenting and Involvement in Peer Reviews: Gender

Variation. *Language in Society 21*, 27-57.

Jordan, M. P. (1991). The linguistic genre of abstracts. In A. Della Volpe (Ed.), *The seventeenth LACUS forum 1990* (pp. 507–527). Lake Bluff, IL: LACUS.

Kaplan, R., Cantor, S., Hagstrom, C., Lia, D., Shiotani, Y. & Zimmerman, C.B. (1994). On abstract writing. *Text 14*(3), 401-26.

Kourilová, M. (1996). The Non-native Speaker and the Skill of Abstracting. *UNESCO ALSED-LSP Newsletter Copenhagen 19*(2), 12-21.

Kirsch, G. (1993). *Women Writing in the Academy: Audience, Authority, and Transformation.* Carbondale/Edwardsville: Southern Illinois University Press.

Lewis, J., & Reader, T. (2009). How can we help adolescent readers meet the challenges of academic text? In J. Lewis (Ed.), *Essential questions in adolescent literacy: Teachers and researchers describe what works in classrooms* (pp. 101–131). New York: Guilford.

Lorés, R. (2004). On RA Abstracts: From Rhetorical Structure to Thematic Organization. *English for Specific Purposes 23*(3), 280-302.

Lorés-Sanz, R. (2009). Different Worlds, Different Audiences: A Contrastive Analysis of Research Article Abstracts. In E. Suomela-Salmi, & F. Dervin (Eds.), *Cross-Linguistic and Cross-Cultural Perspectives on Academic Discourse.* (pp. 187-198). Amsterdam, PA: John Benjamins.

Lorés-Sanz, R. (2008). Authorial Visibility in Research Article and Research Article Abstracts: The Intergeneric Perspective. In S. Burgess, & P. Martín Martín (Eds.), *English as an Additional Language in Research Publication and Communication.* (pp. 105-122). Bern: Peter Lang.

Łyda, A. (2007). *Concessive Relation in Spoken Discourse: A Study into Academic Spoken English.* Katowice: Wydawnictwo Uniwersytetu Śląskiego.

Łyda, A. (2013). *Reflections on Academic Discourse: Studies in Academic English Lexis.* Katowice: Wydawnictwo Uniwersytetu Śląskiego

Łyda, A., Warchał, K. (2014). Gender and Academicity: Insights from Research Article Abstracts in In M. Bondi, & R. Lorés-Sanz (Eds.) *Abstracts in Academic Discourse. Variation and Change.* (pp. 111-128). Bern: Peter Lang.

Łobocki, M. (2010). *Metody i techniki badań pedagogicznych.* Kraków: Oficyna Wydawnicza Impuls.

Macpherson, R. (2004). *English for Academic Purposes.* Warszawa: Wydawnictwo Naukowe PWN.

Marín-Arrese, J. (2009). Effective vs. Epistemic Stance, and Subjectivity/ Intersubjectivity in Political Discourse. A Case Study. In A. Tsangalidis, & R. Facchinetti (Eds.), *Studies on English Modality. In Honour of Frank R. Palmer*. (pp. 23-52). Bern: Peter Lang.

Martin, A. V. (1976). Teaching Academic Vocabulary to Foreign Graduate Students. *TESOL Quarterly 10*(1), 91-97.

Martín-Martín, P. (2003). A Genre Analysis of English and Spanish Research Paper Abstracts

in Experimental Social Sciences. *English for Specific Purposes 22*(1), 25-43.

Martín-Martín, P. (2005). *The Rhetoric of the Abstract in English and Spanish Scientific Discourse. A Cross-Cultural Genre-Analytic Approach.* Bern: Peter Lang.

McEnery, T., Wilson, A. (2001) Corpus Linguistics, 2nd edn. Edinburgh: Edinburgh University Press.

McEnery, T., Xiao, R., Tono, Y (2006). Corpus-Based Language Studies. London: Routledge.

McConnell-Ginet, S. (2008). Language. In Malti-Douglas, F. (Ed.) *Encyclopedia of Sex and Gender.* (Vol 3. pp. 862–865). Detroit: Macmillan.

Melander, B., Swales, J. M., & Fredrickson, K. M. (1997). Journal abstracts from three academic fields in the United States and Sweden: National or disciplinary proclivities?. In A. Duszak (Ed.), *Intellectual styles and cross-cultural communication* (pp. 251–272). Berlin: Mouton De Gruyter.

Milewski, T. (2004). *Językoznawstwo.* (Wydanie siódme uzupełnione). Warszawa: Wydawnictwo Naukowe PWN.

Miller, C. (1994). Rhetorical community: The cultural basis of genre. In A. Freedman & P. Medway (Eds.), *Genre and the new rhetoric* (pp. 23–42). London: Taylor & Francis.

Mills, S. (2008). *Language and Sexism.* Cambridge: Cambridge University Press.

Mohan, B. (1986). *Language and content.* Reading: Addison-Wesley.

Montgomery, M. (1995). *An Introduction to Language and Society.* London: Routledge.

Moore, D.W. (2008). Program development. In K.A. Hinchman & H.K. Sheridan-Thomas (Eds.), *Best practices in adolescent literacy instruction*. (pp. 313–338). New York: Guilford.

Müller, I. (2008). *Die Übersetzung von Abstracts aus translationswis-senschaftlicher Sicht Russisch-Deutsch-Englisch: eine Untersu-chung am Beispiel von Abstracts aus russischen Fachzeitschriften zur Schweißtechnik und ihren Übersetzungen ins Deutsche und Englische.* Berlin: Frank & Timme.

Myers, G. (1990). *Writing biology.* Madison, WI: The University of Wisconsin Press.

Nation, I. S. P. (2001). *Learning Vocabulary in Another Language.* Cambridge: Cambridge University Press.

Nowak, S. (1985). *Metodologia Badań Społecznych.* Warszawa: Państwowe Wydawnictwo Naukowe.

O'Keffe, A. & McCarthy, M. (2010). *The Routledge Handbook of Corpus Linguistics.* London: Routledge.

Oldenburg, H. (1992). *Angewandte Fachtextlinguistik. 'Conclu-sion' und Zusammenfassungen.* Tübingen: Narr.

Ortega Barrera, I. & Torres Ramírez, A. (2010). Estudio sobre los Abstracts de Artículos de Investigación Informáticos: Evidencialidad y Modalidad Textual. *Revista de Lingüística y Lenguas Aplicadas 5*(1), 141-153.

Palmer, F. (1991). *Mood and Modality.* Cambridge: Cambridge University Press.

Perkins, M. (1983). *Modal Expressions in English.* London: Frances Pinter.

Pilgreen, J. (2007). Teaching the language of school to secondary English learners. In J. Lewis & G. Moorman (Eds.), *Adolescent literacy instruction: Policies and promising practices*. (pp. 238–262). Newark, DE: International Reading Association.

Reichenbach, H. (1949). *The theory of probability.* London: Cambridge University Press.

Reichling, A. (1956). Feature Analysis and Linguistic Interpretation. In M. Halle, H.G. Lunt, H. McLean, C.H. Van Schooneveld (Eds.), *For Roman Jakobson.* (pp.418-422). Hague: Mouton & CO.

Rounds, P. (1982). *Hedging in written academic discourse: precision and flexibility (Mimeo).* Ann Arbor, MI: The University of Michigan.

Rubin, A., & Babbie, E. R. (2008). *Research Methods for Social Work. Seventh edition.* Belmont, CA: Brooks/Cole, CENGAGE Learning.

Rubin, D. L., Greene, K. (1992). Gender-Typical Style in Written Language. *Research in the Teaching of English 26*, 7-40.

Salager-Mayer, F. (1990). Discoursal flaws in medical English abstracts: a genre analysis per research and text type. *Text 10*(4), 365-384.

Salager-Mayer, F. (1992). A text-type and move analysis of study verb tense and modality distribution in medical English abstracts. *English for Specific Purposes* (11), 93-113.

Salkie, R. (1996). Modality in English and French: A Corpus-based Approach. *Language Sciences* (18), 381-392.

Samraj, B. (2002a). Disciplinary variation in abstracts: The case of Wildlife Behaviour and Conservation Biology. In J. Flowerdew (Ed.), *Academic discourse* (pp. 40–56). London: Pearson.

Samraj, B. (2002b). Introductions in research articles: Variations across disciplines. *English for Specific Purposes* (21), 1–17.

Samraj, B. (2005). An Exploration of a Genre Set: Research Article Abstracts and Introductions in Two Disciplines. *English for Specific Purposes 24*(2), 141-156.

Santos, M. (1996). The text organization of research papers abstracts in applied linguistics. *Text 16*(4), 481-499.

Scarcella, R. (2003). *Academic English: A conceptual framework.* The University of California Linguistic Minority Research Institute. Technical Report 2003–1.

Scott, J.A., Nagy, W.E., & Flinspach, S.L. (2008). More than merely words: Redefining vocabulary learning in a culturally and linguistically diverse society. In A.E. Farstrup & S.J. Samuels (Eds.), *What research has to say about vocabulary in-struction* (pp. 182–210). Newark, DE: International Reading Association.

Sinclair, J. (1992). Automatic analysis of corpora. In Svartvik J. (Ed.), *Directions in corpus linguistics.* (pp. 379-397) New York: Mouton De Gruyter.

Snow, C.E., & Uccelli, P. (2009). The challenge of academic language. In D.R. Olson & N. Torrance (Eds.), *The Cambridge handbook of literacy*. (pp. 112–133). New York: Cambridge University Press.

Stąpor, K. (2015). *Wykłady z metod statystycznych dla informatyków z przykładami w języku R.* Gliwice: Wydawnictwo Politechniki Śląskiej.

Steponavičius, A. (2006). *Diachronic Linguistics and Etymology.* Warszawa: Wydawnictwo Uniwersytetu w Białymstoku.

Stevens, S. S. (1951). Mathematics, Measurement, and Psychophysics. In S. S. Stevens (Ed.), *A Handbook of experimental Psychology.* (pp.1-49). New York: John Wiley& Sons. pp. 25, 44-49.

Stotesbury, H. (2003). Evaluation in research article abstracts in the narrative and hard sciences. *Journal of English for Academic Purposes* (2), 327-341.

Stotesbury, H. (2006). Gaps and False Conclusions: Criticism in Research Article Abstracts across the Disciplines. In K. Hyland, & M. Bondi (Eds.), *Academic Discourse Across Disciplines.* (pp. 123-148). New York: Peter Lang.

Swales, J. (1990). *Genre analysis: English in academic and research settings.* Cambridge: Cambridge University Press.

Swales, J. & Feak, C. (1994, 2012). *Academic Writing for Graduate Students.* Ann Arbor, MI: The University of Michigan Press.

Swales, J. & Feak, C. (2009). *Abstracts and the Writing of Abstracts.* Ann Arbor, MI: The University of Michigan Press.

Sztumski, J. (2005). *Wstęp do metod i technik badań społecznych.* Katowice: Wydawnictwo Naukowe „Śląsk".

Tannen, D. (1986). That's Not What I Meant! How Conversational Style Makes or Breaks Relationships. New York: Ballantine Books.

Tannen, D. (1999). *Ty nic nie rozumiesz – You Just Don't Understand.* (A. Sylwanowicz, Trans.) Poznań: Zysk i S-ka.

Tannen, D. (2001). *Talking from 9 to 5. Women and Men at Work.* New York: Quill.

Teubert, W. (2004). Language and corpus linguistics. In M.A.K. Halliday, W. Teubert, C. Yallop, & A. Čermáková (Eds.), *Lexicology and Corpus Linguistics* (pp.73-112). London: Continuum.

Teubert, W., & Čermáková, A. (2004). Directions in Corpus Linguistics. In M.A.K. Halliday, W. Teubert, C. Yallop, & A. Čermáková (Eds.), *Lexicology and Corpus Linguistics* (pp.113-166). London: Continuum.

Tibbo, H. (1993). *Abstracting, Information Retrieval and the Humanities. Providing Access to Historical Literature.* London: American Library Association.

Van Bonn, S. & Swales, J. (2007). English and French Journal Abstracts in the Language Sciences: Three Exploratory Studies. *Journal of English for Academic Purposes 6*(2), 93-108.

Ventola, E. (1994a). Abstracts as an object of linguistic study. In S. Cmejrkova, F. Danes & E. Havlova (Eds.) *Writing vs Speaking: Language, Text, Discourse, Communication*. (pp. 333–352). Tubingen: Gunter Narr.

Ventola, E. (1994b). Problems in producing scientific abstracts in L2 . In S. Cmejrkova & F. Sticha (Eds.), *The Syntax of sentence and text* (pp. 283–303). Amsterdam, PA: John Benjamins.

Vihla, M. (1999). *Medical Writing. Modality in Focus.* Amsterdam - Atlanta, GA: Rodopi.

Voegelin, C.F. (1956). Subsystems within systems in Cultural and Linguistic Typologies. In M. Halle, H.G. Lunt, H. McLean, C.H. Van Schooneveld (Eds.), *For Roman Jakobson* (pp.592-599). Hague: Mouton & CO.

Volosinov, V. N. (1929/1973/1986). *Marxism and the philosophy of language.* (L. Matejka, & I. R. Titunik, Trans.). London: Harvard University Press.

Walesiak, M., Gatnar, E. (2009). *Statystyczna analiza danych z wykorzystaniem programu R.* Warszawa: Wydawnictwo Naukowe PWN.

Wickham, H., Grolemund, G. (2018). *Język R. Kompletny zestaw narzędzi dla analityków danych.* (J. Zatorska, Trans.). Gliwice: Helion. (Original text published 2017).

Online references

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

Ahmad, U. K. (1997). *Scientific research articles in Malay: A situated discourse analysis.* Doctoral dissertation, The University of Michigan. <https://www.proquest.com/docview/304377225>

American National Standard for Writing Abstracts. (1979). NewYork: American Standards Institute (ANSI) Publication. DOI: 10.1109/TPC.1977.6591959

Anthony, L. (2015). *TagAnt* (Version 1.2.0) [Computer Software]. Tokyo, Japan: Waseda University. <http://www.antlab.sci.waseda.ac.jp/>

Anthony, L. (2016). *AntMover* (Version 1.10) [Computer Software]. Tokyo, Japan: Waseda University. <http://www.antlab.sci.waseda.ac.jp/>

Anthony, L. (2017). *ProtAnt* (Version 0.1) [Computer Software]. Tokyo, Japan: Waseda University. <http://www.antlab.sci.waseda.ac.jp/>

Anthony, L. (2018) *AntConc* (Version 3.5.7) [Computer Software]. Tokyo, Japan: Waseda University.< http://www.antlab.sci.waseda.ac.jp/>

Applied Linguistics Association of Australia (ALAA) Annual Conference. List of Abstracts. (2016). Melbourne, Australia: Monash University. December 5-7, 2016. Accessed: 4 Aug. 2018. https://pdfslide.net/documents/alaa-2016-dec-5-7-book-of-abstracts.html

Aull, L. L., Bandarage, D., & Miller, M. R. (2017). Generality in student and expert epistemic

stance: A corpus analysis of first-year, upper-level, and published academic writing. *Journal of English for Academic Purposes, 26*, 29–41. doi:10.1016/j.jeap.2017.01.005

Bhatia, V. K. (1997). Genre-mixing in academic introductions. *English for Specific Purposes, 16*(3), 181–195. doi:10.1016/s0889-4906(96)0003-7

Blanton, L. L. (1994). Discourse, artifacts, and the ozarks: Understanding academic literacy. *Journal of Second Language Writing, 3*(1), 1–16. doi:10.1016/1060-3743(94)90002-7

Bridle, M. (2019). Learner use of a corpus as a reference tool in error correction: Factors influencing consultation and success. *Journal of English for Academic Purposes, 37*, 52–69. doi:10.1016/j.jeap.2018.11.003

Can, T., & Cangır, H. (2019). A corpus-assisted comparative analysis of self-mention markers in doctoral dissertations of literary studies written in Turkey and the UK. *Journal of English for Academic Purposes, 42*, 100796. doi:10.1016/j.jeap.2019.100796

Cambridge University Press. (n.d.) *Cambridge dictionary.* Retreived June, 10, 2024.

Candelas de la Ossa, A. (2019). Exceptionalising intersectionality: a corpus study of implied readership in guidance for survivors of domestic abuse. *Gender and Language, 13*(2), 224–250. doi:10.1558/genl.35094

Castro, O. (2013). Introduction: Gender, language and translation at the crossroads of disciplines. *Gender and Language, 7*(1). doi:10.1558/genl.v7i1.5

Chamot, A., & Robbins, J. (2005). *The CALLA Model: Strategies for ELL Student Success.* Workshoop for Region 10. New York, New York City Board of Education. Accessed: 15 Aug. 2024. <https://www.researchgate.net/profile/Anna-Chamot/publication/268345515_The_CALLA_Model_Strategies_for_ELL_Student_Success/links/553f99ee0cf2736761c03038/The-CALLA-Model-Strategies-for-ELL-Student-Success.pdf>

Chan, T. H.-T. (2015). A corpus-based study of the expression of stance in dissertation acknowledgements. *Journal of English for Academic Purposes, 20*, 176–191. doi:10.1016/j.jeap.2015.09.005

Clyne, S. (2006). *Academic Language Functions.* San Francisco: Colorin Colorado. Accessed: Accessed: 15 Aug. 2024. <https://www.colorincolorado.org/sites/default/files/Academic-Language-Function.pdf>

Csomay, E., & Prades, A. (2018). Academic vocabulary in ESL student papers: A corpus-based study. *Journal of English for Academic Purposes, 33*, 100–118. doi:10.1016/j.jeap.2018.02.003

Dalton-Puffer, C. (2013). A construct of cognitive discourse functions for conceptualising content-language integration in CLIL and multilingual education. *European Journal of Applied Linguistics*, *1*(2), 216 – 253. doi:10.1515/eujal-2013-0011

Dolgova, N., & Mueller, C. (2019). How useful are corpus tools for error correction? Insights from learner data. *Journal of English for Academic Purposes, 39*, 97-108. doi:10.1016/j.jeap.2019.03.007

Durrant, P. (2009). Investigating the viability of a collocation list for students of English for academic purposes. *English for Specific Purposes, 28*(3), 157–169.

doi:10.1016/j.esp.2009.02.002

Durrant, P. (2013). Discipline and Level Specificity in University Students' Written Vocabulary. *Applied Linguistics, 35*(3), 328–356. doi:10.1093/applin/amt016

Durrant, P. (2016). To what extent is the Academic Vocabulary List relevant to university student writing? *English for Specific Purposes, 43*, 49–61. doi:10.1016/j.esp.2016.01.004

EDEN 2016 Annual Conference. Re-Imagining Learning Environments. Book of Abstracts. (2016). Budapest, Hungary: European Distance and E-Learning Network. 14-17 June, 2016. Accessed: 18 Sept. 2016. <https://www.eden-online.org/wpcontent/uploads/2016/07/Annual_2016_Budapest_BoA.pdf>

Ergün, E. (2013). Feminist translation and feminist sociolinguistics in dialogue: A multi-layered analysis of linguistic gender constructions in and across English and Turkish. *Gender and Language, 7*(1). doi:10.1558/genl.v7i1.13

Esfandiari, R., & Barbary, F. (2017). A contrastive corpus-driven study of lexical bundles between English writers and Persian writers in psychology research articles. *Journal of English for Academic Purposes, 29*, 21–42. doi:10.1016/j.jeap.2017.09.002

Flowerdew, L. (2012). *Corpora and Language Education.* Pelgrave, MacMillan. doi:10.1057/9780230355569

Flowerdew, L. (2015). Using corpus-based research and online academic corpora to inform writing of the discussion section of a thesis. *Journal of English for Academic Purposes, 20*, 58–68. doi:10.1016/j.jeap.2015.06.001

Freddi, M. (2005). Arguing linguistics: corpus investigation of one functional variety of academic discourse. *Journal of English for Academic Purposes, 4*(1), 5–26. doi:10.1016/j.jeap.2003.09.002

Gao, X. (2016). A cross-disciplinary corpus-based study on English and Chinese native speakers' use of linking adverbials in academic writing. *Journal of English for Academic Purposes, 24*, 14–28. doi:10.1016/j.jeap.2016.08.002

Gardner, D. (2007). Validating the Construct of Word in Applied Corpus-based Vocabulary Research: A Critical Survey. *Applied Linguistics, 28*(2), 241–265. doi:10.1093/applin/amm010

Gardner, D., & Davies, M. (2013). A New Academic Vocabulary List. *Applied Linguistics, 35*(3), 305–327. doi:10.1093/applin/amt015

Gardner, D., & Davies, M. (2016). A response to "To what extent is the Academic Vocabulary List relevant to university student writing?" *English for Specific Purposes, 43*, 62–68. doi:10.1016/j.esp.2016.03.001

Gu, Y. (2002). Gender, Academic Major, and Vocabulary Learning Strategies of Chinese Efl Learners. *RELC Journal, 33*(1), 35–54. doi:10.1177/003368820203300102

Gutiérrez, K.D. (2008). Developing a sociocritical literacy in the third space. *Reading Research Quarterly 43*(2), 148–164. doi:10.1598/RRQ.43.2.3

Hasko, V. (2012). Qualitative Corpus Analysis. In C. A. Chapelle. (Ed.). *The Encyclopedia of*

Applied Linguistics. Doi:10.10029781405198431.wbeal0974

Hyland, K. (2003). Self-citation and self-reference: Credibility and promotion in academic publication. *Journal of the American Society for Information Science and Technology 54*(3), 251-259. doi:http://dx.doi.org/10.1002/asi.10204

Hyland, K., Tse, P. (2004). Hooking the reader: a corpus study of evaluative *that* in abstracts. *English for Specific Purposes 24*(2), 123-139. doi:https://doi.org/10.1016/j.esp.2004.02.002

Hyland, K., & Jiang, F. (2017). Is academic writing becoming more informal? *English for Specific Purposes 45*, 40-51. doi:http://dx.doi.org/10.1016/j.esp.2016.09.001

HYLAND, K., & TSE, P. (2007). Is There an "Academic Vocabulary"? *TESOL Quarterly, 41*(2), 235–253. doi:10.1002/j.1545-7249.2007.tb00058.x

ICHA The 17th International Conference of Harmful Algae. Abstract Book. (2016). Florianopolis, Santa Catarina, Brazil: N.p. October 9-14, 2016. Accessed: 21 Jul. 2018. <http://www.icha2016.com/program>

International Society of Behavioral Nutrition and Physical Activity 2016 Annual Meeting. Abstract Book. (2016). Cape Town, South Africa: International Society of Behavioral Nutrition and Physical Activity. June 2, 2016. Accessed: 21 Jul. 2018. <https://www.apct.pt/files/annual_meetings/2016/08/31/19/attachments/57c72823e3fda.pdf>

Janzen, J. (2008). Teaching English language learners in the content areas. *Review of Educational Research 78*(4), 1010–1038. doi:10.3102/0034654308325580

Jaworska, S., & Hunt, S. (2016). Differentiations and intersections: a corpus-assisted discourse study of gender representations in the British Press before, during and after the London Olympics 2012. *Gender and Language, 11*(3), 336–364. doi:10.1558/genl.28858

Kanoksilapatham, B. (2005). Rhetorical studies of biochemistry research articles. *English for Specific Purposes 24*, 269-292. doi:http://dx.doi.org/10.1016/j.esp.2004.08.003

Keenan E. O. (1975). Dimensions in Sociolinguistics. [Review of the book Explorations in the functions of language., by M. A. K. Halliday, 1973] *Language in Society 4*(2), 247-253. doi:10.1017/S0047404500004735

Kitamura, A. (2014). English mystique? A critical discourse analysis on gendered bilingualism in Japan. *Gender and Language, 10*(1), 1–20. doi:10.1558/genl.v10i1.20692

Kuznetsova, J. (2014). Are verbs politically correct? A corpus study of gender in Russian verbs. *Gender and Language, 9*(3), 391–434. doi:10.1558/genl.v9i3.16650

Lan, G., & Sun, Y. (2018). A corpus-based investigation of noun phrase complexity in the L2 writings of a first-year composition course. *Journal of English for Academic Purposes, 38*, 14-24. doi:10.1016/j.jeap.2018.12.001

Lea, M.R., & Street, B.V. (2006). The "academic literacies" model: Theory and applications. *Theory Into Practice 45*(4), 368–377. doi:10.1207/s15430421tip4504_11

Lei, L., & Liu, D. (2016). A new medical academic word list: A corpus-based study with enhanced methodology. *Journal of English for Academic Purposes, 22*, 42–53. doi:10.1016/j.jeap.2016.01.008

Lei, S., & Yang, R. (2020). Lexical richness in research articles: Corpus-based comparative study among advanced Chinese learners of English, English native beginner students and experts. *Journal of English for Academic Purposes, 47*, 100894. doi:10.1016/j.jeap.2020.100894

Lillis, T., McMullan, J., & Tuck, J. (2018). Gender and academic writing. *Journal of English for Academic Purposes, 32*, 1–8. doi:10.1016/j.jeap.2018.03.003

Lin, A.M.Y. (2016). *Language across the Curriculum & CLIL in English as an Additional Language (EAL) Contexts. Theory and Practice.* Hong Kong: Springer. doi:10.1007/978-981-10-1802-2

Ma, S., Zhang, J., Anderson, R. C., Morris, J., Nguyen-Jahiel, K. T., Miller, B., … Grabow, K. (2016). Children's Productive Use of Academic Vocabulary. *Discourse Processes, 54*(1), 40–61. doi:10.1080/0163853x.2016.1166889

Man, D., & Chau, M. H. (2018). Learning to evaluate through that-clauses: Evidence from a longitudinal learner corpus. *Journal of English for Academic Purposes, 37*, 22-33. doi:10.1016/j.jeap.2018.11.007

Martínez, I. A., Beck, S. C., & Panza, C. B. (2009). Academic vocabulary in agriculture research articles: A corpus-based study. *English for Specific Purposes, 28*(3), 183–198. doi:10.1016/j.esp.2009.04.003

Moser, F., & Masterson, J. (2013). Are there signs of change in gendered language use in children's early reading vocabulary? *Gender and Language, 8*(1), 71–90. doi:10.1558/genl.v8i1.71

MOTSCHENBACHER, H. (2009). Speaking the gendered body: The performative construction of commercial femininities and masculinities via body-part vocabulary. *Language in Society, 38*(01), 1. doi:10.1017/s0047404508090015

Nygaard, L. P., & Bahgat, K. (2018). What's in a number? How (and why) measuring research productivity in different ways changes the gender gap. *Journal of English for Academic Purposes, 32*, 67–79. doi:10.1016/j.jeap.2018.03.009

Merriam-Webster. (n.d.) *The Merriam-Webster dictionary.* Retrieved June, 10, 2024.

Orasan, C. (2001). Patterns in Scientific Abstracts. In P. Rayson, A. Wilson, T. McEnry, A. Hardie & S. Khoja (Eds.), *Proceedings of Corpus Linguistics 2001 Conference.* (pp. , 433-443). Lancaster: Lancaster University. Retrieved from http://ucrel.lancs.ac.uk/publications/CL2003/CL2001%20conference/papers/orasan.pdf

Paribakht, T. S., & Webb, S. (2016). The relationship between academic vocabulary coverage and scores on a standardized English proficiency test. *Journal of English for Academic Purposes, 21*, 121–132. doi:10.1016/j.jeap.2015.05.009

Paterson, L. L. (2019). Feeling Academic in the Neoliberal University: Feminist Flights, Fights and Failures edited by Yvette Taylor and Kinneret Lahad (2018). *Gender and Language, 13*(3), 1–6. doi:10.1558/genl.39038

Pathan, H., Memon, R. A., Memon, S., Shah, S. W. A., & Magsi, A. (2018). Academic Vocabulary Use in Doctoral Theses: A Corpus-Based Lexical Analysis of Academic Word List (AWL) in Major Scientific Disciplinary Groups. *International Journal of English Linguistics,*

8(4), 282. doi:10.5539/ijel.v8n4p282

Pecorari, D., Shaw, P., & Malmström, H. (2019). Developing a new academic vocabulary test. *Journal of English for Academic Purposes, 39*, 59-71. doi:10.1016/j.jeap.2019.02.004

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: http://dx.doi.org/10.1177/1461445607087010

Poole, R., Gnann, A., & Hahn-Powell, G. (2019). Epistemic stance and the construction of knowledge in science writing: A diachronic corpus study. *Journal of English for Academic Purposes, 42*, 100784. doi:10.1016/j.jeap.2019.100784

Preece, S. (2018). Identity work in the academic writing classroom: Where gender meets social class. *Journal of English for Academic Purposes, 32*, 9–20. doi:10.1016/j.jeap.2018.03.004

Sicurella, F. G. (2014). The approach that dares speak its name: queer and the problem of "big nouns" in the language of academia. *Gender and Language, 10*(1), 73–84. doi:10.1558/genl.v10i1.20895

Simon-Maeda, A. (2016). A corpus-based study of the AAAL conference handbook. *Journal of English for Academic Purposes, 23*, 71–82. doi:10.1016/j.jeap.2016.06.001

Stefanowitsch, A. (2020). *Corpus Linguistics. A guide to methodology.* (Textbooks in Language Sciences 7). Berlin: Language Science Press. doi:10.5281/zenodo.3735822

Sweetwater Union High School District & District-Wide Academic Support Teams. (2010). *Academic Language Function Toolkit. A Resource for Developing Academic Language for All Students in All Content Areas.* Accessed: 15 Aug. 2024. <https://www.literacyhow.org/wp-content/uploads/2020/06/Academic-Language-Functions-toolkit.pdf>

Torgesen, J.K., Houston, D.D., Rissman, L.M., Decker, S.M., Roberts, G., Vaughn, S., Vexler, J., … Lesaux, N. (2007). *Academic literacy instruction for adolescents: A guidance document from the Center on Instruction.* Accessed: October 21, 2019. Retrieved from https://files.eric.ed.gov/fulltext/ED578488.pdf

Tribus, A. C. (2017). *The Communicative Functions of Language: An Exploration of Roman Jakobson's Theory in TESOL.* (MA Dissertation). Brattleboro, VT: SIT Graduate Institute. Accessed: 5 Dec. 2017. Retrieved from https://digitalcollections.sit.edu/ipp_collection/723/?utm_source=digitalcollections.sit.edu%2Fipp_collection%2F723&utm_medium=PDF&utm_campaign=PDFCoverPages

Tu, P., & Wang, S. P. (2013). Corpus-based Research on Tense Analysis and Rhetorical Struture in Journal Article Abstracts. In Y. Harada & R. Otoguro (Eds.), *Proceedings of the 27th Oacific Asia Conference on Language , Information, and Computation* (pp. 102-107). Retrieved from https://www.aclweb.org/anthology/Y13-1008.pdf

Tuck, J. (2018). "I'm nobody's Mum in this university": The gendering of work around student writing in UK higher education. *Journal of English for Academic Purposes, 32*, 32–41. doi:10.1016/j.jeap.2018.03.006

useR 2016 Conference. Book of Abstracts. (2016). California, USA: Stanford University. June 27-30, 2016. Accessed: 18 Sept. 2016. <http://user2016.r-project.org/files/abs-book.pdf>

Valipouri, L., & Nassaji, H. (2013). A corpus-based study of academic vocabulary in chemistry research articles. *Journal of English for Academic Purposes, 12*(4), 248–263. doi:10.1016/j.jeap.2013.07.001

Venables, W. N., & Smith, D. M. (2015). An Introduction to R. N.p, n.d. Accessed: 19 Sept. 2016. <http://cran.r-project.org/doc/manuals/R-intro.pdf>

Wang, Y. L. (2014). A Corpus-based Contrastive Study on Academic Vocabulary of RA Abstracts in Applied Linguistics. Applied Mechanics and Materials, 631-632, 1442–1446. doi:10.4028/www.scientific.net/amm.631-632.1442

Watinee, S., & Siriluck, U. (2013). Research Article Abstracts in Two Related Disciplines: Rhetorical Variation between Linguistics and Applied Linguistics. *English Language Teaching 6*(2), 84-99. doi:10.5539/elt.v6n2p84

Yating, Y. (2019). Media representations of “leftover women” in China: a corpus-assisted critical discourse analysis. *Gender and Language, 13*(3), 369–395. doi:10.1558/genl.36223

Streszczenie

Istnieją różne sposoby przedstawiania, definiowania i badania struktury i funkcji retorycznych abstraktów, co ilustrują przytoczone w tej rozprawie prace naukowe innych badaczy. Pomimo tego, że poprzednie badania odnosiły się do wielu celów badań nad abstraktami, nie tylko w zakresie językoznawstwa korpusowego, żadne z nich nie koncentrowało się na funkcjach języka akademickiego w abstraktach. Zwłaszcza z wykorzystaniem metod lingwistyki komputerowej do celów badawczych. Niniejsza rozprawa została rozplanowana następująco: materiał z przygotowanego przez badacza korpusu został oceniony i opisany z pomocą metod lingwistyki komputerowej, w celu porównania słownictwa akademickiego używanego w abstraktach z konferencji naukowych, przez obie płcie w wybranych do badania dziedzinach naukowych. Podstawą do określenia najczęściej używanego słownictwa była metodologia Averil Coxhead (2000). Celem analizy było wykrycie ewentualnych interdyscyplinarnych różnic w częstości występowania funkcji języka akademickiego. Jako wiodący parametr badania przyjęto płeć. Przedstawione dane korpusowe, które posłużyły do analizy leksykalnej, zostały wpierw podzielone na części mowy, a następnie na trzy kategorie: Słownictwo w procesie badawczym, Słownictwo w procesie analizy oraz Słownictwo w procesie oceny. Kategorie te zostały wybrane na podstawie metodologii przyjętej przez Anne V. Martin (1976). Dane do korpusu zostały wybrane wyłącznie z pięciu ksiąg abstraktów, z pięciu konferencji, z których każda odbywała się na innym kontynencie. Dane wybierane były ręcznie przez badacza, aby wyrównać liczbę przykładów, zgodnie z parametrem płci.

Dane korpusu zostały podzielone na trzy dziedziny nauki humanistyczne, nauki biologiczne i zdrowotne oraz nauki wykorzystujące technologie informatyczne. Korpus użyty w niniejszej rozprawie został zaprojektowany w celu określenia ilościowych różnic w wykorzystaniu funkcji języka akademickiego w różnych dziedzinach pisanego języka akademickiego. Dane korpusu zawierają następujące elementy: Źródło; Numer referencyjny i numer strony; Autora, Płeć, Typ mówcy, Instytucję i Pochodzenie mówcy; Dyscyplinę, Tytuł, Abstrakt; Abstrakt oznaczony częściami mowy; Ilość rzeczowników, czasowników, przymiotników i przysłówków użytych w tekstach korpusu. Zakłada się, że w grupach badawczych, oraz między nimi, wystąpią różnice w realizacji funkcji językowych i użytku słownictwa akademickiego.

Praca podzielona jest na pięć rozdziałów. Pierwszy rozdział przedstawia podstawy teoretyczne dotyczące: komunikacji w dyskursie akademickim, gatunku abstraktu, elementów

lingwistyki korpusowej i komputerowej, aspektu płci w badaniach językoznawczych, elementów oceny danych w badaniach statystycznych oraz słownictwa i funkcji języka akademickiego; informacje zawarte w pięciu podrozdziałach pierwszego rozdziału tworzą podstawę najważniejszych pojęć użytych w tej rozprawie. Drugi rozdział zawiera opis praktycznych podejść do analizy abstraktów, nie tylko w badaniach korpusowych. Siedem punktów przedstawia streszczenia prezentowanych artykułów z czasopism, które koncentrują się na perspektywach międzyjęzykowych, dyscyplinarnych, międzykulturowych i diachronicznych, a także na strukturze retorycznej i elementach studiów korpusowych. W rozdziale trzecim przedstawiono metody badań oraz przybliżono programy komputerowe użyte do analizy, tj. programy wykorzystywane w rozprawie (R, AntConc, ProtAnt, TagAnt). Centralna część pracy, rozdział czwarty, prezentuje dane korpusowe, przebadane pod kątem płci i pochodzenia mówców, aby lepiej zobrazować różnice i podobieństwa występujące w badanym materiale. Ilościowa analiza komputerowa była podstawową metodą badania wybranych próbek zebranego materiału. Interpretacje przedstawionych tabel, zawierające dane odnośnie użycia określonej leksyki oraz ilościowego użycia określonych funkcji języka akademickiego przez kobiety i mężczyzn, są weryfikowane za pomocą testu chi-kwadrat Pearsona przeprowadzonego w programie R i oprogramowaniu Ant. Ostatnia część pracy dyplomowej zawiera wnioski odpowiadające interpretacji danych przedstawionych w poszczególnych częściach rozdziału czwartego oraz propozycje dalszych badań w tych zakresach. Poszczególne sekcje tego rozdziału przywołują podobieństwa i różnice danych badawczych, dotyczących płci i pochodzenia autorów (native/non-native). Wyniki badań omówiono w odniesieniu do wybranych źródeł literatury. Wnioski poparte są testami chi-kwadrat Pearsona (w stosownych przypadkach), wyodrębnionymi z oprogramowania R, przedstawiającymi liczbowe różnice między określonymi danymi badanego materiału.

Przeprowadzone testy chi-kwadrat potwierdziły statystyczne kilka zależności. Po pierwsze, użytek funkcji F7, związanej z ewaluacją, jest zróżnicowany pomiędzy badanymi dziedzinami i wpływa na niego płeć ($p=0,02$). Po drugie, istotną wartość testu chi-kwadrat ($p=0,01$) wykazał użytek funkcji F9, odnoszącej się do organizowania kolejności przedstawianych informacji, w odniesieniu do ogólnej liczby rodzimych i nierodzimych pisarzy stosujących tą funkcję w abstraktach w badanych dziedzinach. Po trzecie, funkcja F1, związana z procesami badawczymi, analizowana w odniesieniu do połączonych parametrów dziedziny akademickiej i pochodzenia pisarzy dała wartość p. równą 0,009. Następnie, funkcje F3.1, F3.2 oraz F3.3, związane z prezentowaniem metodologii, badane w odniesieniu do połączonych parametrów płci oraz dziedziny akademickiej, wykazały wartości testu na

poziomach 0,008, 0,03 oraz 0,004. Kolejno, funkcja F4, związana z przedstawianiem wyników badań, jest używana częściej przez pisarzy rodzimych w dziedzinie Biologii i Nauk o zdrowiu oraz przez nierodzimych w pozostałych badanych dziedzinach (p=0,03). Ostatecznie, funkcja F5, odnosząca się do przedstawiania wniosków, uzyskała w badaniu chi-kwadrat wartość 0,002, co przekłada się na częstszy statystyczny użytek owej funkcji przez nierodzimych pisarzy w dziedzinach Statystyki komputerowej oraz Humanistyki w badanym materiale.