universität wien

MASTERARBEIT / MASTER'S THESIS

Titel der Masterarbeit / Title of the Master's Thesis

„Informality in academic journals“

verfasst von / submitted by

Erna Hebib, BA

angestrebter akademischer Grad / in partial fulfillment of the requirements for the degree of

Master of Arts (MA)

Wien, 2022 / Vienna 2022

Studienkennzahl lt. Studienblatt / degree programme code as it appears on the student record sheet:	UA 066 812
Studienrichtung lt. Studienblatt / degree programme as it appears on the student record sheet:	Master-Studium English Language and Linguistics
Betreut von / Supervisor:	Univ.-Prof. Mag. Dr. Nikolaus Ritt

Acknowledgments

Language changes due to various external factors. Just like that, this thesis has been influenced by various events in the past and has been shaped into its current form. I definitely could not have finished this work without a few important people who provided enormous amounts of their time as well as support to help me. I would like to use this opportunity to express my gratitude to them.

Above all, I want to thank my supervisor, Nikolaus Ritt. He, along with professor Lotte Sommerer, stoked my interest in historical linguistics and language change. From the beginning, he guided me in the right direction and helped me to shape my topic so that it is appropriate, doable, and meaningful. There were problems in some stages of this research that I managed to solve on my own, but he provided solutions on how to deal with the bigger problems which I believed to be a dead end. Later, he also pointed out areas for improvement and his feedback drew my attention to some things which I did not notice. Owing to him, I had the opportunity to select and work on a topic from the historical linguistics area & I'm extremely grateful for that.

I am also grateful to my parents who supported me from the first semester at the University of Vienna, and who during my last semester were also extremely supportive in various ways. They loved hearing about my topic, every progress I made, and the methods and solutions in this project. I would also like to thank my pet cat, Tabby, for being a constant reminder to take a break now and then and to manage stress by sitting in the garden or attempting to catch the mysterious red dot. Lastly, I would like to thank my best friend, Robert, for his patience and daily support, and motivation. At some point, he even encouraged me to do a follow-up research in a few years. He has been there for me from the first day of this research and listened to my occasional rants.

I'm, again, very grateful to have these people in my life.

Table of Contents

List of Tables & Figures

1. Introduction

The world has developed to a great extent and at a rapid pace in the past 40 years. In addition to an obvious change such as enormous population growth, literacy rates have also increased in the majority of the world. Language does not withstand such global advancements; it changes and develops as well. Moreover, it is undeniable that language changes are directly caused by changes in society. English, being the global lingua franca, is anything but an exception when it comes to change. Academic writing as a genre was established centuries ago and it entails a specific vocabulary and tone. However, genres are subject to change and development as well. A noticeable change in academic writing over the years has been observed and studied by numerous authors. Hyland & Jiang (2017:40) point out that numerous researchers have noticed a progressive shift in written academic texts; the shift develops from a distanced towards an involved writing style. Style is certainly not the only criterion that determines whether a text is formal or informal; in addition to style, vocabulary and other features also increase or decrease the formality of a text. This paper was designed to investigate whether and to which extent informality in written academic texts has increased and is a replication study of Hyland & Jiang's study from 2017. The investigation focused solely on written academic articles in English and the presence of informal features in those texts. The present study investigated 10 informal traits in 16 different academic journals, from four different disciplines, all of which are formal. The study featured journals from two social sciences, linguistics and economics, as well as journals from two natural sciences, namely, mathematics and biology. The study entails data from 1980, 2000, and 2020. The focus is exclusively on the development of informal elements, and will not investigate the vocabulary or any other features which contribute to or diminish the formality of a text. Aside from the aforementioned factors, there are a plethora of other factors which cause a change in a language. For example, since the number of authors has gradually increased since the 1980s, it is plausible to question whether the writing quality has decreased, as quantity and quality are almost inseparable: if quantity increases, it is often at the cost of quality. Technology and communication have developed enormously in the past 20 years and as a result, social media has been present

and influential on individuals for an extended period. While developments in the whole world may have triggered the trend toward informality, the development and availability of various media resources, as well as social media, later on, is potentially what facilitated informality to spread to formal domains. Finally, with a greater number of authors, the need to seek prominence and individuality might arise and thus some authors might employ a specific out-of-the-norm writing style to attract readers. Certain contexts such as social media, blogs, and conversations between colleagues do not require formal language, and whether formality increased or decreased in such contexts is irrelevant for this study. By contrast, this study aims to scrutinize the presence or absence of informal elements in academic journals. Hyland & Jiang's study from 2017 is the closest in parameters and theme to the aim of this study. Hyland & Jiang investigate whether the formality in academic writing is decreasing and in which ways. They compiled their corpus which encompasses three periods 1965 to 1985 and until 2015. They investigated changes between the three periods, yet also the changes throughout five decades. Their corpus entails data from four disciplines: applied linguistics, sociology, electrical engineering, and biology as they deemed them representative of social and natural sciences. They selected six random papers from five journals for each discipline. In the end, their corpus consisted of 360 papers and around 2.2 million words. They used TreeTagger to grammatically tag various parts of speech in their corpus, and they searched for traits of informality using AntConc. Their research is overall rather inexplicit in numerous ways and that raises the question as to how balanced it is overall. Although it is not a severe issue, another problem with their study is the irregular interval between the three periods. The present study will provide a representative overview of the changes which have taken place in the four decades preceding 2020. This research will entail an insight into the various developments in writing styles of different disciplines while also providing a transparent overview of the data, methods, and results. Certain genres entail a larger percentage of elements, which do not contribute to the word count, than others. It is unclear how Hyland & Jiang have dealt with this issue in their study from 2017, yet this non-usable data has been eliminated from the word count in this study. Although this study is similar, in terms of scope, to Hyland & Jiang's study, it is more precise

and detailed as well as slightly more recent and it will feature more regular time intervals. This paper will attempt to solve the problem of the previously mentioned study by providing a detailed and clear answer about the development of informality. Its importance lies in data that will demonstrate in which ways each discipline has become more informal and thus the members of each discipline might recognize and avoid those informal patterns in the future. Finally, it is another piece of literature on this specific contemporary writing issue. The first chapter provides a basic introduction to the term informality and provides a review of studies that have been conducted previously and are related to informality. Here, it is important to note that previous studies, such as Hyland & Jiang's study from 2017 view informality as a construct and that it represents the attitudes of writers and editors amongst others. This delivers necessary background information which is salient to understanding the study. It is important to restate which traits have been used for measuring informality. Thus, section 2 provides an overview of 10 informal traits, and examples, as well as an attempt to explain why each of those features is considered to be informal. The third chapter focuses on the concepts of formality & informality. Different approaches and ideas of informality have been discussed along with a conclusive list of informal features by Chang & Swales (1999). The last section in this chapter presents the research aims, research questions, and hypotheses in detail. Next, the data collection process is explained in detail, as well as the tools which have been used to analyze around 2 million words. Moreover, because the study entails an equal number of disciplines from social sciences as well as from natural sciences and an almost equal number of words from all disciplines and journals, this study employs an extensively balanced approach, to track the developments of informal traits in formal writing. This section features more details on the methods that have been used and the reasons behind why these methods have been selected for this study and a brief description of participants, materials, and the corpus. The data analysis process as well as raw frequencies, and an overview of the informal traits per discipline are also described in this chapter. Lastly, the issues and solutions related to data collection and analysis are explained. Chapter 5 entails a discussion of the results, as well as their meaning and significance. This section provides the normalized frequencies which

provide answers to the research questions and deal with the hypothesis. Finally, the results are related to previous studies. The most problematic areas are highlighted and explained. Limitations and suggestions for further research are also stated in this chapter. This chapter is finally followed by the conclusion, which summarizes the main findings and contributions to this field of research.

2. Preliminaries

In the past and the contemporary period, there was and still is a sufficient amount of style manuals that instruct writers on the basics of academic writing. The mere existence and distribution of these guides, do not guarantee that all authors are familiar with the conventions of academic writing. The informal traits which will be used as a benchmark for measuring informality have been adapted from Chang & Swales (1999) and entail a comprehensive and generally accepted list of elements that count as problematic in academic texts. These informal features were collected from a variety of different guidebooks; the purpose of those guidebooks is to teach how to achieve a high level of formality in writing by using certain grammatical features. (Chang & Swales 1999: 147) This chapter will provide an overview and introduction of each informal element which will be used as a basis to investigate whether academic journals have become more informal. The purpose of this chapter is to provide an overview of each informal feature and to attempt to explain its form, function, and what makes it informal, as well as to provide examples and explain why authors might use a specific feature in their works.

2.1. Research Aim, Research Questions & Hypotheses

The main aim is to compare and contrast the developments of the four disciplines, or within the three time periods. There are a few research questions yet the main research question is: “To what extent on average has informality increased in academic journals?” The first hypothesis is that the informality in natural sciences has increased. The second hypothesis is that informality in social sciences has decreased. This study will highlight which discipline is

leading when it comes to developing informality. Providing feedback on which trait is the most commonly found and most problematic trait is the first step towards improvement. The methodology is straightforward. The frequency of 10 traits considered to be indicative of informality will be scrutinized in three distinct periods: namely, the data is extracted from journals published in the 1980s, 2000, and 2020. The 10 informal traits were adapted from Chang & Swales' (1999:148), and include First-person pronouns to refer to the author, unattended anaphoric pronouns, split infinitives, sentence-initial conjunctions or conjunctive adverbs, sentence-final prepositions, listing expressions, second-person pronouns/determiners to refer to the reader, contractions, direct questions, and exclamations. They will be used as criteria to gauge the level of informality of each journal, in each period. This research will investigate the frequency of informal traits in different disciplines. The first research question is concerned with whether this study replicates the results of the model study. The guiding question of Hyland & Jiang's study, which is "Has academic writing become less formal and, if so, in what ways and in what disciplines?" (2017:43) is the first research question of this study and will also be answered with the data from this study. Hyland & Jiang's answer to the question "Is academic writing becoming more informal?" is "it depends" (Hyland & Jiang 2017: 48), and there is a need for a more precise answer than that. For this study to confirm the findings of the original study, it would have to confirm Hyland & Jiang's findings and overlap with the increases or decreases of individual informal elements. The second research question will scrutinize whether certain disciplines are more informal than others and if they are to what extent do they differ? This will provide a general overview as a starting point. The third research question will investigate whether informality has increased in different disciplines over the years. This will demonstrate the developments in each genre. The fourth research question will focus on the dominant informal traits in each genre. This will answer the question of how informality has increased, or which traits have become more frequent throughout the years. The fifth research question will investigate whether specific informal traits have remained constant in frequency in a specific genre. This will enable the comparison or contrast of different disciplines. FInally, the sixth research question will scrutinize whether

any correlation can be identified between the various journals and countries of publishing or editing.

2.2. Literature review

Almost all of the available research that has been conducted, can be divided into two broad categories: studies on spoken language and studies on written language. Although spoken and written data are equally important, while producing written research articles, the authors have an opportunity to edit their work before it is published. When it comes to spoken data, editing is not an option; speakers would potentially also edit their spoken data if it were an option. In certain cases, authors hand in their work due to deadlines and do not always submit a version of a text that they are content with; however, written data is already edited and reflects in most cases an author's comfortable and confident language use. That is the main reason why the data used for this research is exclusively from written sources, and why this study is primarily concerned with written texts. There are a plethora of studies that focus on various aspects of the English language and how they changed over the past few decades. Moreover, there are fewer studies that focus on the changes and developments of one genre. This, to an extent, makes sense, as certain traits are genre-specific; these traits also help to distinguish genres from one another or to categorize subgenres into one group. These traits fix genres in time and contain their essence; this is also true for academic texts. Mair for example (1998:153) has noted that informal elements, which are commonly found in spoken texts, are contemporarily commonly found in written texts. Thus, conducting a study on written data is plausible. Constantinou et al. also noted that in the UK, among 585 extracts from student essays, formality has decreased over 10 years (Constantinou et al 2019, cited in Akhtar & Riaz 2019:6). Since those are native speakers, it is plausible to investigate whether this trend is found in other written data, such as journals which are the basis of this paper. Similarly, Akhtar & Riaz (2019: 17) have found that the informality in the essays of EFL undergraduates has not increased or decreased, yet the overall level of formality is low. Academia entails numerous rules and prohibitions, and that is to retain quality and prestige. On "a cline of openness to innovation ranging from

"agile" to "uptight" genres", Hundt and Mair (1999: 221) state that academic genres are on the more traditional side, which is old-fashioned and not prone to change. This is plausible, yet there is more proof that informality is increasing in formal genres. Because numerous authors have written about the increase in informality, it is plausible to assume that it is not happening in one specific discipline, age group, or region but globally. Melissourgou & Maruster (2017) focused on informal traits, provided by Chang and Swales (1999), in academic research articles from Medicine, Philosophy, Economics, and Business, their findings almost entirely overlap with the findings of Hyland & Jiang's 2017 study. Some authors managed to specify where this development is most prominent and where the opposite is happening. Although there are general ideas about what is formal and what is informal, they are not equally employed in all departments. Hyland & Jiang (2017:48) state that informal features have decreased in social sciences, yet they have increased in natural sciences. The study which is published in 2017 and titled *Is academic writing becoming more informal?* is one of the most salient studies for this paper. Hyland & Jiang have compiled three corpora for three time periods: 1965, 1985, and 2015. Their study provides an overview of informal traits throughout 50 years and the aim is to investigate if and how academic writing has become more informal. They selected six random papers from each science, including applied linguistics, sociology, electrical engineering, and biology. Since their aim was also to provide a representative study, they selected disciplines from soft and hard sciences. Their study entails over 360 papers and around 2.2 million words, they used TreeTagger to tag their corpus, and AntConc to find the instances of informal traits. Both authors checked all instances to confirm that they were representative of the informal feature. As a benchmark for grading informality, they adapted a list of informal traits identified and compiled by Chang and Swales (1999), because these are following their experience with informality in texts from students. They disregarded one feature which, according to them does not occur often, and the final list entails the main problems highlighted in writing manuals. After tracking the number of each informal feature in each genre and period, they normalized their frequencies by 10.000 words. The final numbers revealed that informality has increased by around 2% and that informal features grew

especially in the hard sciences, whereas the articles in social sciences became somewhat more formal. (Hyland & Jiang 2017:44) Their study also highlights three main informal features which mainly contributed to the changes: first-person pronouns, unattended anaphoric references, and sentences beginning with conjunctions. Moreover, they proved that informal features do not have a linear development nor increase or decrease equally in a period nor among the selected disciplines. Another study that is salient for this paper is Alipour & Nooreddinmoosaa's 2018 study, where they focused on informal traits in native and non-native speaker writings. They conducted a quantitative and qualitative study of informal traits in a corpus that consists of 200 articles from applied linguistics written by native as well as non-native speakers. They, just like Hyland & Jiang used the same informal features proposed by Chang & Swales in 1999. Similarly, they used AntConc and checked the frequencies within 1.000 words. They found that native speakers use informal traits more frequently than non-native speakers. (Alipour & Nooreddinmoosaa 2018:367) According to their findings, sentence-initial conjunctions were the most commonly occurring informal features, along with first-person pronouns, listing expressions, second-person pronouns, split infinitives, and direct questions. These informal traits were most prominent in their findings. In contrast to that, exclamation marks were the least commonly used informal traits among both types of speakers, along with stranded prepositions, contractions, and unattended anaphoric pronouns. (Alipour & Nooreddinmoosaa 2018: 362) The aforementioned two studies are similar to the present study, and the purpose of this study is to continue the research on informality in academic texts and relate the findings to form a wider picture of the development of informality in academia. The mentioned two studies are relatively recent, and this will study will provide also an overview of the recent changes. However, the previous two studies entail some flaws. The former study is relatively obsolete, as the most recent period is 2015. The balance of the data in the study is questionable: this is because they selected six papers from each journal, yet the articles potentially vary in length to a great extent. Moreover, it lacks numerous details on the individual traits in each discipline, and the fact that they did not normalize their data by per million makes data comparison difficult. Hyland & Jiang selected non-uniform periods: their

first period is 20 years apart from the second period, and the period from their second to their last period is 30 years apart. The last issue, which they also explicitly mentioned, is that some of their data is from 1967 instead of 1965. Furthermore, the study does not explain the data analysis process, nor does it mention the size of the individual sub-corpora. Moreover, unusable data, such as formulas, which do not contribute to the word count, appear to be ignored. The issue with Alipour & Nooreddinmoosaa's 2018 study is that it focuses on a small native and non-native scale which is, without doubt, intriguing, yet does not provide an overview on a global scale. Finally, the present study is also balanced, strict, and straightforward and the aforementioned flaws have been avoided. Moreover, the present study deals with data from random open access journals. It also focuses on informality in journals from different countries.

2.2.1. Formality & informality

The first step is to define formality, as well as to contrast it to informality, as informality is a salient term in this paper. Bennet (2009:50) studied academic style manuals and found that the majority of authors of those manuals affirm that “academic writing is by nature formal and technical”. Kuo (1999:122) pointed out that Academic writing was distinguished by impersonality. By contrast, according to the Online Cambridge Dictionary, informality is the fact of not being formal or official, or of being suitable for friends and family but not for official occasions. Although academic writing is known to be formal and detached in nature, further studies confirmed that informality is expanding in academic contexts. Leedham (2011:258) has shown in her study that there is a significant amount of informal features in the essays of undergraduates. Leedham’s study was published during a time when Facebook and other social media were at the peak of their popularity. Thus it is not surprising that that generation was influenced by an omnipresent social media service. In contrast to this, Belal (2014:36) suggests that social media can be beneficial to students and their writing skills if used properly. Hyland & Jiang (2017: 48) have also conducted a study on informal elements in academic journals and found that academic writing is becoming more informal. As there are at least 50 different disciplines, it is unlikely that informality has

increased at the same rate in all of those disciplines. Hyland & Jiang also provided some clarification for the development of this trend. Namely, that the increase of informal elements is genre-specific and dependent on the individual feature. (Hyland & Jiang 2017: 48) In other words, certain genres have become more informal over time than others, and one trait of informality became more dominant in specific genres. Finally, Basal & Bada (2012:1778) mention that numerous studies, based on data from corpora, have demonstrated the presence of the authors in academic texts, which is traditionally rejected in scientific writing. This is not just a discrepancy, it also has consequences. Namely, this imbalance between contemporary developments and traditional rules may cause insecurity among writers and the decision on whether to mention themselves in their work or not do so. (Basal & Bada 2012:1778) Consequently, this might cause confusion and the inability to make the right choice. There are numerous studies on formality as well as informality in academic writing. However, there are also numerous inconsistencies when it comes to deciding what is formal and what is informal. To explain the approach as well as introduce the central topic of this study, an evaluation of previous research is necessary. Formal language is used in important texts and mainly in academia. There are numerous reasons behind such a strict approach to writing. One of the reasons why formal language is dominating the scientific domain is because it's void of any ambiguity. (Heylighen 1999:5) Academia mainly focuses on knowledge and original ideas and unambiguous language maintains order and originality among billions of authors. In contrast to informal language, formal language is "less direct", "detached", "impersonal" and "objective". (Heylighen & Dewaele 1999:10) As formal language is mostly used to target an academic audience, there is no need for an amicable or personal tone. Kane warns in his Essential Writing Guide that an author must examine the context and audience as well as the level of formality which they seek, as a detached point of view seems more formal than a personal point of view. (Kane 2000: 75) Bennet was one of the numerous authors who studied writing manuals; he concluded that based on 20 manuals, 16 point out clearness and precision as the main aim, while 10 manuals mentioned brevity, and 7 also noted detachment and neutrality as fundamental traits of writing in academia. (Bennet 2009:45) Although academic writing is

strict and complicated, it is crucial. Kane (2000:179) warns that formality might seem too complex for the contemporary reader, on the contrary, formality must not be compared to unnaturalness, and naturalness on the other hand must not be thought of as the sole archetype. Hyland & Jiang point out that formality within the rules of academic writing entails treating the audience as equals, no ambiguity, no proximity, and a fabricated egalitarianism. (Hyland & Jiang 2017:41) However, academic language is even more complex than that. Heylighen (1999:5) was among the first to list the benefits of formal expressions over those which depend on context; he also adds that the lack of ambiguity is the reason why ideas and predictions are extensively formalized in scientific contexts. For him, the most salient and visible benefit of formal language is its characteristic to withstand changes throughout time. He argues that "The longer we desire our expressions to remain meaningful, the more formal we should try to make them." (Heylighen 1999:5) The second benefit of formal language is according to Heylighen (1999:6) the ability to be communicated to the widest audience. Thus it is the scientist who generates phrases and ideas which every person can comprehend. Finally, the third and last advantage of formal language is the ability to test whether it is true or false; this is also what instigates the formality in the expression of scientific ideas. Finally, all of these traits together facilitate the production and quality of knowledge. Testability enables one to choose quality ideas and decline inappropriate ones; storability guarantees the maintenance of appropriate ideas and universality signals that knowledge can be distributed. (Heylighen 1999:6) Moreover, seeking authenticity and fact is expressed through objectivity, and formal language rejects any biased stances to create the appearance of a neutral unknown author who is focused on intercommunication with similar individuals. (Hyland & Jiang 2017:41) Although it is complex, academic language is rather fair. It is a field where precise and unbiased knowledge sharing is paramount to all social characteristics, such as gender or expertise for example. (Hyland & Jiang 2017:41) Aside from promoting equality, academic language is also unbiased. Formality implies that authors detach their identity before producing academic texts. (Hyland & Jiang 2017:41) Thus, being fair and unbiased, all authors in academia have equal opportunities if they adapt a formal writing style. Hyland &

Jiang point out that "The conventions of formality mean that, as far as possible, authors leave their personalities at the door when they sit down to write". (2017: 41) These are all also potentially reasons why various indicators of informality have been proscribed from formal domains. After highlighting one of the most important traits of formality, which is impersonality, it is important to mention that different authors and editors have different opinions and attitudes towards what is informal and what is not. The best example of this is Chang & Swales' list of informal traits which have been mentioned in various style manuals (1999:148); the items on this list are ordered by their number of instances in style manuals and some are more common than others. Moreover, after defining "informality" it is important to introduce and explain the two types of formal language. Heylighen & Dewaele (1999:5) distinguish "surface formality" which is "characterized by attention to form for the sake of convention or form itself" and "deep formality" which is "attention to form for the sake of unequivocal understanding of the precise meaning of the expression". This paper is concerned with the former. Aside from the aforementioned traits of formality, such as the lack of ambiguity, informality is much broader than just the opposites of the mentioned, it is necessary to define informality. Similar to formality, informality can be expressed in many ways. Coffin et al. provide potentially the best and most comprehensive view of the features of informality: "the use of technical, elevated or abstract vocabulary, complex sentence structures and the avoidance of the personal voice". (Coffin et al. 2003: 28, cited in Hyland & Jiang 2017:42) Alipour & Nooreddinmoosaa (2018:351) warn that informality should not be defined as the opposite of formality. Although the focus of this particular work is on informality, it is undeniable that both formality and informality are equally important in written and spoken text. Formality is characterized by precise vocabulary to avoid misunderstandings and vagueness, however, informality lacks any stiffness and supports a non-tense and friendly role. (Hyland & Jiang 2017:41) The provided definitions are accurate, yet simplified. Informality is "the absence of full grammatical sentences, a decrease in concern about punctuation, and a high tolerance for typographic and spelling errors" (Coffin et al. 2003:141) Moreover, formality occurs most commonly, but not exclusively, in prestigious and academic contexts, while informality occurs in less prestigious

and non-academic contexts. Informality is not negative or detrimental in any way; it is just that both concepts serve different purposes and contexts.

2.2.2. Studies on individual informal traits

There are numerous studies on individual informal traits in language. Leedham (2011:258) reports a high frequency of specific informal elements which are typical of spoken texts, as well as connectors and both 1st person pronouns and 2nd person pronouns. Also, Pitarch (2016) focused on the academic language in banking; they investigated personal pronouns in 64 About us pages written in English and Spanish from the US, the UK, and Spain. Their findings indicate that possessive pronouns had the highest frequency. Swales (2005:9) has investigated anaphoric *this* and found that it is least commonly found in biology and medicine and most commonly found in philosophy, physics, and then applied linguistics. Moreover, in the case of first-person pronouns, Harwood (2005) has conducted a qualitative corpus study based on the use of first-person pronouns I and us in four disciplines. Harwood acknowledges that academic writing is usually free from personal pronouns (2005). Yet Harwood also acknowledges that there is a need among academic writers to promote themselves and highlight the importance of their work (2005:1209). His findings prove that personal pronouns can be found in hard and soft sciences and that aside from self-promotion, they can be used in various ways to facilitate the writing and make the text more cohesive. (Harwood 2005) Moreover, some authors also published works on unattended anaphoric pronouns. Swales (2005) has studied unattended this and determined its frequency in ten different disciplines, as well as how common it occurs clause-initially, and some of its collocations. Swales also points out its numerous flaws, yet he finally concludes that extensive use of attended *this* facilitates the reading process and decreases the ambiguity. Social sciences and humanities have higher frequencies of this, and that signifies a slightly higher degree of informality. (Swales 2005:9) Split infinitives have also been researched by some authors. Supakorn (2013) points out at the beginning of his study the controversial issue of the split infinitive. He proceeds to list reasons why it

should be banned, as well as why it is beneficial. His study was conducted on split infinitives in essays and writings of three levels of English learners. He found that the most advanced learners use the split infinitive most commonly, and concluded that less advanced learners might lack the confidence to split an infinitive. He also concludes that as the level of proficiency increases, the diversity of split infinitives increases as well. Supakorn (2013:31) found that as the level of proficiency in English increases, the frequency of split infinitives in their writing also increases. Moreover, Bell 2007 was one of the authors who focused on sentence-initial conjunctions. More precisely, he focuses on sentence-initial *And* and sentence-initial *But*. He outlines a significant difference in the use of both sentence-initial conjunctions in different disciplines of academic texts and finds that *And* is the most frequently used additive marker, while *But* is the 2nd most common connective used to contrast. He also highlights three salient functions for *And* and *But.* Zhao's (2017) study scrutinized the use of conjunctions among native and non-native authors. She regarded conjunctions as markers of linguistic proficiency. The factors that cause preposition stranding have also been studied. Dimitriadis (2007) studied the factors which increase the stranding of prepositions. More importantly, she found that more formal papers do not entail as many stranded prepositions as other papers do. She also explains "preferred preposition stranding" which are instances where stranding a preposition is acceptable or as mentioned preferred. Direct questions count as another problematic aspect of academic writing. However, Hyland (2002:11) found that approximately 1/3 of all questions facilitated forming the discourse, whereas 20% facilitated highlighting the author's aim. Finally, he also states that questions can be used to gain the reader's attention by addressing them directly via question. (Hyland 2002:11) Traditionally, questions are personal and count as problematic in academic writing. Yet, according to Hyland (2002:23), the amount of personality in discourse is governed by questions, and thus they vary in frequency in various contexts. (Hyland 2002:23) Praminatih, Kwary, and Ardaniah (2018) found that after first-person pronouns, second-person pronouns are the 2nd informal trait for EFL undergraduate students in their essays. However, all of those researches are about informal characteristics

but are not conclusive, as all of these studies focus on an individual informal trait at one point in time, often in a smaller corpus, one discipline, or between two groups.

2.2.3. Disagreement & Criticism

Although there are numerous guides and handbooks and numerous opinions have been mentioned, there is overall no agreement as to what is acceptable and what is not in formal writing. Chang & Swales compiled a list of informal features among 40 style guides and recorded the number of instances each feature was mentioned (1999:148); these numbers vary from 2 to 15. Formality has also been criticized and while certain traits might be informal to some authors, others fail to understand the issue with those traits. After all, formal language has some limitations. Formally conveying ideas is rather complicated, and an absolute formalization is impossible. (Heylighen 1999:6) Moreover, not every writer or researcher is fond of a prescriptivist approach. For example, certain authors desire to explicitly mention themselves and become prominent to the reader. Thus, on one hand, there is a fixed genre supported by generations of prescriptivists, and on the other hand, there are generations of modern writers and the preference to be prominent and unique in academia.

2.3. Informal elements

Each section in this chapter will focus on one of the following features: first-person pronouns, unattended anaphoric pronouns, split infinitives, sentence-initial conjunctions or conjunctive adverbs, sentence-final prepositions, listing expressions, second-person pronouns/determiners to refer to the reader, contractions, direct questions and exclamations respectively.

2.3.1. First-person pronouns

The first set of informal items on the list are first-person pronouns; Chang & Swales have listed first-person pronouns as the most commonly mentioned item in the 40 guidebooks

which they analyzed for informal features. (1999:148) They are also the cause of the largest controversy when it comes to informality in academia. Geertz has described academic writing as "author –evacuated" (Geertz 1988, cited in Msuya 2016:55). In other words, academic writing is extensively impersonal. The issue with personal pronouns, when it comes to formality, is already mentioned in their name: these pronouns are personal, and academic writing is known to be detached. However, first-person pronouns are also extensively useful in writing, as there is no synonym for personal pronouns. A potential alternative to *I* would be *The author* however if the author is writing the article and writes about an author, the reader might become puzzled or might even question the validity of the article. The two first-person pronouns which have been investigated for this study are *I* and *we*. Although both *I* and *we* can be used in combination with numerous other words, authors in extensively long studies most often include phrases such as *the generalizations I have already made* or *we summarized the results of three studies*. In addition, Basal & Bada (2012: 1784) found in their study that various native, as well as non-native speakers of English, often use *I* and *we* in academic writing. Numerous authors use *I* or *we* in at least one section of their paper. Thus, it appears that the use of first-person pronouns in academic texts is not motivated or affected by an author's mother tongue. Although first-person pronouns should be avoided, there are numerous other instances when authors might use *I* or *we* to convey an idea; this is because it is rather difficult to form a sentence without using pronouns. For example, if an author attempts to describe the procedure of an experiment, they will likely have to use *I* to explain the steps of their study in detail. Similarly, there is also an explanation as to why one or two authors might resort to using *we* in their study. *We* potentially occurs in studies that are co-authored by two or more authors, especially when the authors want to announce the results of a study. Moreover, both authors write the contents of a study, and this is not done independently. Thus, both authors potentially seek to be the more prominent persona in their study, and to solve this, they might agree on using *we*. It is difficult to imagine an entire co-authored study without first-person pronouns, and paraphrasing might not be an option for some authors. In addition, Basal & Bada (2012: 1787) mention that first-person pronouns are in certain cases

used by authors to “indicate their authorial presence”, and this argument is plausible. An author might be extensively satisfied with their text so they might want to mention themselves in the academic work. However, this is the opposite of what the style manuals mention: “Write your paper with third-person narration that avoids “I believe” and “It is my opinion.”. (Lester & Lester 2015: 184) Various guidebooks warn writers that *I* or *we* is inappropriate and should be avoided in academic writing: “The total paper is considered to be the work of the writer. You don’t have to say ‘I think’ or ‘My opinion is’ in the paper. (...) Traditional formal writing does not use I or we in the body of the paper” (Spencer & Arbon 1996:26, cited in Koutsantoni 2007: 173) Furthermore, style manuals demonstrate various impersonal constructions which could be used to avoid personal pronouns, and it is indeed possible to write a paper or an article without using personal pronouns. Basal & Bada (2012:1778) point out that when it comes to the frequency of personal pronouns in academic texts, there is no set norm, but rather only advice on how often to use them. There is a lack of rules which define in which instances personal pronouns can be used. Thus, members of different disciplines decide on the use of personal pronouns in their field, and the frequency of personal pronouns thus differs among different disciplines. (Basal & Bada 2012:1778) To sum up, the issue is that it is difficult to avoid using personal pronouns in academic writing, but there is a wide range of reasons why authors might use personal pronouns in academic writing, and pride might be
one of them.

2.3.2. Unattended anaphoric pronouns

Unattended anaphoric pronouns are another set of elements that are considered informal in academic writing. They are the second most commonly mentioned informal element according to Chang & Swales (1999:148). Unattended pronouns are usually related to a preceding word or phrase, hence the name “anaphoric”. This study has investigated the following unattended anaphoric pronouns: *this, that, these, those,* or *it*; these can also be attended in certain cases. Due to their similar form and function, this chapter will not discuss each anaphoric pronoun in detail but regards them as a group of informal traits. An

example would entail a newly introduced statement or topic, such as the fact that *if the temperature drops, ice melts* and an author might want to elaborate on this statement: thus, they might start the subsequent sentence with *This*, referring back to the ice melting, and might add additional information such as *This can be prevented by...* or *This will cause...* T*his*, as well as other anaphoric pronouns, facilitates the argumentation and cohesiveness of the text. Without a doubt, unattended anaphoric pronouns are flexible when it comes to referents and their use in the language. However, there is a reason why they are considered inappropriate in academic writing. Swales (2005:1) points out that the major problem with *this* is its vagueness. As it has already been noted, academic writing is characterized by precise language. For example, the phrase *This is frequently the case* could refer back to numerous referents, and it is difficult to decide, based on context, which word or topic it was meant to connect to. Swales (2005:1) also highlights that *this* is rather common in academic writing. This is because the word *this* is acceptable in specific contexts. This is the case, for example, in parts of sentences such as *This research has…* or *This study is*. The reason why *this* is frequent is that it has numerous grammatical functions, as demonstrated by examples. Moreover, authors who do not specialize in language potentially do not distinguish between the various instances of *this* and thus do not understand when to use which form. This is also the case for *that, those, these* as well as *it*; although they are versatile in form and meaning, all of these anaphoric pronouns are vague in terms and reference. It is important to note that in the case of unattended anaphoric pronouns, although all representative instances are listed, it is perhaps impossible to determine which instances are actual anaphoric pronouns and which are determinatives with the ellipsis of the nominal heads. Moreover, aside from vagueness, there is another issue when it comes to *this*. Mauranen (1992: 243) adds that *this* causes an "impression of closeness and solidarity between reader and writer" and this is problematic in academic writing. It is unknown whether writers have noticed this effect, and whether this is the motivation behind using *this* in an academic paper. Finally, *this,* as well as other unattended anaphoric pronouns such as *that, those, these, and it*, are overall vague in terms of meaning, yet

versatile in terms of function. This combination of feature and form is potentially the reason why authors use unattended anaphoric pronouns correctly, yet also incorrectly.

2.3.3. Split infinitives

Split infinitives are another informal feature that is relevant for this study; thus, their frequency in academic texts will be investigated. Chang & Swales recorded those split infinitives take the third place for being the third most frequent informal trait mentioned in writing manuals. (1999:148) However, in the case of split infinitives, there is no fixed phrase or word which will be located in the text. By contrast, it is a large set of possible combinations. More precisely, split infinitives consist of *to,* an adverb, and a verb, and consequently, thousands of different word combinations are possible for this group of informal features. An example of a split infinitive is *to boldly go*. It is essentially an infinitive, such as *to go* that has been split by an adverb, such as *boldly*. Moreover, explaining why a split infinitive is considered informal might be challenging. Leech, Deuchar, and Hoogenraad explain that "some people (especially older people who have gone through a traditional education) feel quite strongly that we should not split our infinitives" (Leech, Deuchar, and Hoogenraad 2006: 184, cited in Supakorn 2013: 23). It seems that in the case of split infinitives, they are perhaps not inherently informal, but simply considered inferior. Supakorn (2013:24) points out that an adverb that divides *to* from an infinitive should not be considered a mistake, because the word *to* is not a component of the infinitive phrase. By contrast, some authors, such as Bryson, state that "there is no logical reason not to split an infinitive". (1990:35) Certain authors agree with this statement and suggest what could happen to split infinitives in the future. For example, Stageberg & Oaks (2000:333) speculate that over time, the frequency of split infinitives will increase in all varieties of written English. This study will investigate whether this is the case. Although, Carter & McCarthy's 2006 study concluded that in spoken texts, native English speakers do split infinitives frequently. The fact that spoken language is often informal might be another reason why some authors or prescriptive grammarians consider the split infinitive to be informal. Supakorn (2013:27) has proved in his study that the split infinitive is overall on the

increase in the texts in his corpus. Thus, it appears that the authors are open concerning the use of slightly more complex construction. It also potentially shows that numerous authors prefer to use split infinitives in their writing. There are numerous reasons why an author might use a split infinitive in their work. One reason is that the mother tongue of an author might potentially motivate the use of split infinitives. That is potentially because this construction is rather flexible and straightforward and might cause their writing to appear more proficient, complex, or fluent. Moreover, authors might also feel more confident about their writing if they add an adverb between an infinitive form of a verb. Whereas the former is just speculation, the latter has been confirmed. Confident English learners tend to use split infinitives in their writing more as this is different from the norm, which is the avoidance of infinitive splitting. (Supakorn 2013:27) Finally, academic texts almost without exception have a word limit and a split infinitive will in every case be longer than a non-split infinitive and will cause the text to be more fluent. Split infinitives are complex on numerous levels. However, if they were more complex than functional, their frequency would not increase over the years, and writers would not value them to the extent they do.

2.3.4. Sentence-initial conjunctions

This section will solely focus on *And* and *But* in sentence-initial position. Split infinitives appear on Chang & Swales' list as the fourth most commonly mentioned informal feature in writing manuals (1999:148). Bell (2007: 183) also proved in his study that out of all sentence-initial connectors, sentence-initial *And* and sentence-initial *But* are the most frequent. An example of *And* in sentence-initial position would be: *And if this is true, B is the solution*, and an example of *But* in sentence-initial position follows: *But then the former is more significant.* Similar to split infinitives, sentence-initial *And and* sentence-initial *But* have always been banned by prescriptivist grammarians. (Bell 2007: 183) In other words, neither of these two words in sentence-initial position is inappropriate because of a specific problem, such as lack of preciseness, but rather because they might have not been used in this way traditionally. Although *And* and *But* are considered informal, their versatility is undeniable. One of the functions of *And* is that it can substitute the wort *Finally,* which is a

frequent and salient word in academic writing. For example, the sentence *After the data analysis, the hypothesis was finally confirmed* could be paraphrased as The data was analyzed, and the results confirmed the hypothesis. Here, *And* is used to indicate the end of a list or the last element. Similarly, *But* could be used instead of *However*. Moreover, there are numerous reasons why a certain author might start a sentence with *And* or *But*. Bell also pointed out the three main functions of sentence-initial *And*: to give form to a text by highlighting the last point on a list, for arguments to evolve by the progression of a topic, and to point out a change in an author's point of view with relation to explanatory, contrastive and comments in parentheses related to the preceding text. (2007:193) Similarly, Bell also highlighted the functions of sentence-initial *but*. Namely, he notes that sentence-initial but is used to organize ideas, explain arguments, and change the topic area. (Bell 2007:195) Thus, both conjunctions are useful. Moreover, *But* is used most frequently to continue a line of argumentation, *And* is mostly used to mention the last item on a list. (Bell 2007: 184) Without a doubt, both *And* and *But* are rather convenient in academic writing. Bell (2007: 184) confirms that both *And* and *But* are common in academic writing, however, he points out that *But* is more common than *And*. This is plausible because academic texts usually include more arguments than lists. Finally, Bell (2007: 99) highlights the importance of *And* and *But* by pointing out a unique feature, namely that they "provide special features of cohesion that alternative forms of coordination do not" and even adds that the reason why numerous authors use them in academic writing, although advised not to, is because of the "linguistic privileges they allow" in contrast to stylistic reasons. Moreover, the use of *And* or *But* in sentence-initial position is not limited to the mentioned functions. In contrast to Bell's 2007 study, this study will investigate the overall frequency of these two conjunctions, and will not investigate their frequencies. Based on the arguments above, *And* and *But* are highly beneficial in academic writing. Moreover, it appears that their functions in a text outweigh their assumed informality. This is proved because although the use of these two conjunctions was not supported throughout centuries, they are still commonly found in various texts.

2.3.5. Preposition stranding

A sentence-final preposition or stranded preposition is, as the name suggests, a preposition in the ultimate position of a sentence. Chang & Swales have listed stranded prepositions as the 5th out of 10 informal traits which are commonly mentioned in style guides. (1999:148) Preposition stranding is the separation of a preposition from its object; more precisely, the preposition remains situated while the object moves. (Alaowffi & Alharbi 2021:943) *Which country is Vienna in?* and *A preposition is something you should not end a sentence with* are examples of stranded prepositions. Moreover, Biber et al (1999: 105) provide a definition: "if it is not followed by its complement or, where the preposition is bound to a preceding verb, by the prepositional object" a preposition is stranded. With regards to reasons why a sentence-final preposition might be considered informal, they are similar to split infinitives and sentence-initial conjunctions. In other words, it is owing to prescriptivist grammarians who consider stranded prepositions informal and inferior. (Adejare 2021:41) In other words, stranding a preposition was not commonly found in traditional texts, thus it is not viewed as traditional or appropriate. Moreover, Adejare (2021: 41) adds that "stranding is associated with speech and informal style". Based on the aforementioned features, it seems that numerous informal traits originated in spoken language, and thus, if they occur in written language, they are considered informal. Bergh & Seppänen (2000: 307) confirm that preposition stranding occurs more frequently in speech than in writing. Moreover, in Hyland & Jiang's study (2017:45) which examined informal traits from 1965 until 2015, the frequency of stranded prepositions decreased, before slightly increasing in the end. This study will investigate the development of prepositions in academic contexts in a slightly later period than Hyland & Jiang's. There are numerous reasons why authors might strand prepositions; both native and non-native speakers of English might use stranded prepositions due to the reason that they cause them to appear more experienced and their writing to appear more complex and modern, or even more fluent. There is, however, a case in language when the stranding of prepositions is preferred: namely, in the case when the preposition is a part of a phrasal verb (Huddleston & Pullum 2002: 269). Phrasal verbs usually entail or end with a preposition. Dimitriadis (2007:3) provides examples in her paper

and these include: account *for, ask for, look out for, rely on,* and *run through*, among others. Similarly, there are specific instances when stranding of prepositions should be avoided: that is in case the preposition "precedes a content clause" (Huddleston & Pullum 2002: 630) Huddleston and Pullum (2002: 630) also provide the following example for their argument: *To whom did she declare that she was not going to take any more abuse?* Moreover, the prepositions which were investigated include: *to, on, for, with,* and *of.* Although these conventions are plausible, this study will not investigate in detail whether the occurrences of stranded prepositions, which do occur in academic texts, are preferable or not, as this is not the main goal nor related to one of the research questions which is focused on the frequency of informal traits from 1980 until 2020.

2.3.6. Listing expressions

Listing expressions, as the name suggests, are used after not listing all items in a sentence and wanting to let the audience know that a list is not finite. Chang and Swales provided a list of the most common informal traits which are mentioned in style manuals and listing expressions take the 6th place. (1999:148) The following example includes a rather common listing expression: *If the weather is nice, we can go out and play with the dog, the sand, the ball, and so on*. The most popular listing expressions in English are: *and so on, and so forth,* and *et cetera* which is sometimes abbreviated to *etc*. Expressions such as *firstly, second*, and *Lastly* are found in academic writing yet these expressions do not count as problematic. This is due to the reason that these expressions are not vague but rather precisely state the order of arguments or examples. In addition, they are cohesive devices. By contrast, listing expressions are without doubt vague, and that might be the main reason why listing expressions might be considered informal. However, within academic articles which are usually a few thousand words long, writers might prefer to use *and so on* to end a list without actually ending it. Ending a sentence with *and so on* might mean “and anything else related or similar”. Moreover, it does not indicate how many unnamed items remain. Consider the following example: *There were many different pets at the pet fair: cats, dogs, parrots, fish, and so on*. In this case, *and so on* could potentially have up to at least 20

different additional referents. Thus, using an extension particle in writing does not specify anything further aside from what has been mentioned. From a semantic point of view, this phrase is almost useless as it does not provide any precise additional information, aside from the fact that the list is not complete. In addition, although formal words are usually derived from Latin, in the case of *et cetera*, this argument is invalid. Dubois confirms that and states that "etc. or and the like elicits disfavorable judgment of written style". (1992:198) This is also potentially related to or based on the opinions of prescriptive grammarians, although it is not mentioned in the literature. Moreover, listing expressions were one of the informal traits that Chang & Swales (1999: 148) found in 4 out of 40 style manuals and used in their study to measure the level of informality in academic texts. In addition, *and so on, and so forth* and *et cetera* seem appropriate in a casual conversation between two acquaintances because they do not have a strict but a relaxed tone. It is important to mention one of Dubois's findings. Namely, Dubois (1992: 196-197) also noticed that if a speaker uses *et cetera*, it is usually before they change the topic. Although the literature on this topic is scarce, this statement seems plausible. This is only due to the reason that *and so on* and *and so forth* seem relaxed, whereas *et cetera* seems slightly tenser and as if it highlights the end of a specific topic. Although lists do appear in academic writing, using any listing expression at the end of a sentence in an academic text is frowned upon. Finally, listing expressions are considered informal and will thus be investigated to estimate, with other informal traits, to what extent informality has increased in academic writing.

2.3.7. Second person pronouns

In this section, the focus is on second-person pronouns *you* and *your*, and their presence in academic journals. Unlike other items from this list, Chang & Swales did not mention second person pronouns in their list from 1999; this trait was listed by Hyland & Jiang (2017:44) as a substitute for the category of sentence fragments. Hyland & Jiang pointed out that this was done because sentence fragments "almost never occur" (2017: 44) in their texts. Consequently, this was also adapted to this study, and sentence fragments were not

investigated. As for 2nd person pronouns, authors might use them in instances such as *You must be familiar with this theory* or If *your knowledge of this subject is sufficient…* Similar to first-person pronouns, the issue with second-person pronouns is the fact that they are extensively personal, and thus their presence in a genre that is impersonal is considered to be problematic. In the case of first-person pronouns, it might be difficult to avoid using them in academic writing, as the author is writing from his point of view. By contrast, in academic writing, there is hardly, if any reason why an author might want to establish a personal relationship or directly address a reader whom the author personally does not know. Finally, there is no imaginable scenario where a reader is relevant to a specific study to the extent that they are mentioned in that study. Moreover, thousands of other readers will also read and potentially cite it as a reference. In contrast to first-person pronouns, in the case of 2nd person pronouns, there are numerous ways to form a sentence without using *you* or *your*. In certain cases, however, it is impossible to form a sentence without 2nd person pronouns. De Cock mentions in her study a potential reason why that is the case: namely, she mentions three different ways of using *you*. (De Cock 2016:364) Although it is important and engaging to mention the different ways in which this personal pronoun can be used, it is salient to mention that the result of this study will solely focus on frequency, and will not investigate which of the following function each of the examples fulfills. The first example that De Cock (2016:364) mentions is the "universal reading" of *you* which has already been taken into consideration and explained. This is potentially the most common and most familiar use of *you* for numerous writers. The second *you* which De Cock (2016: 364) mentions is when "the speaker is talking about him/herself in the 2nd person. This *you* is rather similar and potentially interchangeable with *one*, in examples such as *One does not know what to do in that case*. De Cock (2016: 364) adds that this *you* could have more referents which she notes could be "me or and/or everybody". De Cock (2016: 365) notes that based on studies in various European languages, the frequency of 2nd person singular forms has increased. Moreover, certain authors may mirror what Fairclough (2001:62) calls “synthetic personalization” which includes addressing a wide audience through involving language and treating them all as independent individuals. This is often done in marketing

and aids to draw the reader into the topic or central matter. Finally, there are numerous reasons why second-person pronouns occur in academia.

2.3.8. Contractions

Contractions are two words that have been abbreviated into one and are connected by an apostrophe. Chang & Swales have listed contractions as one of the most commonly mentioned features in writing manuals and thus contractions were investigated in this study. (1999:148) Harley (2006: 283) explains that "a single phonological word that is created when a listeme that can be pronounced as an independent phonological word is reduced and attached to another independent word" is a contraction. Examples of contractions include: *I'm, can't, didn't, it'll,* and *isn't* amongst others, and are usually used at the beginning of a sentence, as in: *This can't be the case* or *The experiment didn't provide an answer to the posed question*. In the English language, there are potentially around 100 different contractions; this is because a plethora of words can be fused and turned into contractions. Kjellmer (1998:155) provides a detailed overview of contraction in English and states that the verb contraction and *not* contraction are the two principal types of prepositions in the English language. These are potentially also the most common types of contraction in English; however, it is important to mention that this study will not investigate in detail the type of the prepositions which occur in academic text, but only their frequency. This is related to the main research question, which is concerned with the formality in academic journals in the past decades, rather than the frequency of various types of contractions. For Kjellmer, the former "type occurs when an auxiliary is cliticised to a preceding subject, often or mostly a pronoun" (1998: 155) The prototypical examples for these kinds of contractions include: *I'm, you're, she's* and *they'd.* This kind of contraction includes pronouns. The second type of contraction is slightly different. Kjellmer (1998:155) explains that the other kind "occurs when *not* is cliticised to a preceding auxiliary as when are not becomes aren't and when had not becomes hadn't". This type of contraction is also called "negative contractions" (Kjellmer 1998:155) and includes an auxiliary verb and the negation not. Examples of this kind of contraction include: aren't, haven't, and didn't.

Moreover, Kjellmer (1998:155) introduces two additional types of contractions, "such as non-pronominal verb contraction" which he also calls "subsidiary types", as well as "isolated types". The examples of the former include: should've seen, the store's closed, or it'll end eventually and examples of the latter include: *d'y ou, 's me* or *let's.* (Kjellmer 1998:155) Similar to other informal traits, numerous guidebooks advise writers to avoid contractions: for example, Chang & Swales conducted their study while focusing on contractions as one of the informal traits to investigate the level of informality in academic texts. They also pointed out that they found contractions mentioned in three writing manuals. Moreover, authors might intend to gain prominence and might abbreviate a few consequent words in a paper, such as *Let's assume they'd've been present at…* Finally, other authors might copy or produce similar phrases, believing that it is appropriate. Furthermore, contractions involve, as the name implies, shorter versions of words. Although the meaning of a word does not change, there is no reason why formal language would favor the use of contractions. It is not the case that a contraction can be used to conceal some of the original meaning or create any sort of ambiguity, however, there is also no reason to avoid using the full set of two words that make up a specific contraction.

2.3.9. Direct questions

If viewed from a traditional perspective, academic texts are managed by questions. (Hyland 2002:12) Academic writing entails salient research questions. However, some authors sometimes -also ask rhetorical questions, indirect questions, or even direct questions in academic contexts. Kane (2000:90) explains that topic sentences might sometimes be formed into rhetorical questions, to attract a reader's attention. By contrast, indirect questions are slightly different: they are often hedged and more polite, such as *Could you tell me where the nearest bookstore is?* Although they are questions and require an answer, they are posed as statements. (Kane 2000:385) Moreover, indirect questions are also slightly different in form: indirect questions end with a full stop, rather than a question mark. In addition, some authors promote the use of questions in essays, to obtain interest. (Kirszner & Mandell 1987:67, cited in Hyland 2002:531) Other authors also highlight an

important function of questions. Namely, Swales (1990:156) notes that questions in the introductions of research articles can be a “minor way of establishing a niche”. There is one final reason why questions are beneficial in writing. Hyland (2002:530) found that questions are found in cases where "writers seek to explicitly establish the presence of their readers in the text". There is an important reason behind this. Namely, to introduce the reader to the writer's point of view. (Hyland 2002:530) In addition, there is another less obvious way in which questions benefit writers and readers. Hyland (2002: 14) found that in his corpus, approximately 30% of all questions were used to coordinate the flow of information in a text, and this was mostly done by students. This is a rather rational and advanced way of using questions. Moreover, as it has been mentioned at the beginning of this section, academia entails questions and thus authors might consider that their work will be more prominent if they include questions. Hyland (2002: 16) adds a similar point and suggests that the use of questions by an author demonstrates that the writer is familiar with the kind of reader their work is targeting. Questions in a text have an important and beneficial role. However, numerous authors warn against the use of questions and advice writers to avoid any type of questions in their writing (Swales & Feak 1994: 74, cited in Hyland 2002:3). This is potentially due to the reason that different kinds of questions establish a personal connection between the addresser and addressee. In contrast to this, a recent study found that questions are commonly found in various genres of academic prose. (Hyland 2002:530) This might suggest that the benefits of asking a question in academic texts outweigh the sole statement that they are informal, based on potentially obsolete approaches. Although different kinds of questions have been reviewed in this section, this research will focus on direct questions in academic writing; this is due to the reason that the list of informal features which has been adapted from Swales & Chang’s study focuses only on direct questions because numerous writing manuals listed direct questions as informal traits (1999:148).

2.3.10. Exclamations

Exclamations are marked by an exclamation mark at the end, rather than a period. Exclamations are the least frequently mentioned item on Chang & Swales' list of informal features which were collected from various writing manuals. Exclamations often, but not exclusively, have an almost identical purpose as statements; namely, their purpose is to inform, yet in a different way. The purpose of exclamation marks is to highlight the importance or significance of a subject or a text. (Kane 2000: 386) Aside from that, exclamation marks can also be used to express an emotion, such as excitement for example: *Wow!* Similar to other informal traits, numerous academic writing guidebooks advise against the use of exclamation marks; Chang & Swales (1999:148) identified them as one of the traits which count as problematic in their list which was based on style manuals. Moreover, certain elements are considered informal because prescriptive grammarians advise against those elements; however, this is not the case with exclamation marks. In other words, no literature which promotes or emphasizes the use of exclamation marks in academic writing has been found. On the contrary, numerous works and authors provide reasons why exclamation marks count as problematic and almost void of any value. Kane (2000:222) for example, argues that a well-trained writer does not rely on exclamation points in their writing. Although this approach resembles prescriptivism, there is a reason why this is the case. This is because exclamation points are minor parts of a text and almost immediately become insignificant. (Kane 2000:222) Being ineffective is only one of the reasons. Moreover, Kane (2000: 386) adds that an exclamation point as a means of accentuation is restricted. In other words, using more exclamation marks subsequently or in a set of phrases is also not effective, as a well-phrased statement for example. In other words, using exclamation marks also suggests that the writer is not familiar with other, potentially more prominent, ways of emphasizing their words. Exclamation marks even decrease the value of an author's writing. Furthermore, exclamation points are usually used at the end of imperatives and thus hint at a tone of command. (Kane 2000:386) A commanding tone is potentially authoritative, yet it is impersonal and rather unprofessional. All of the mentioned traits are rather characteristic of casual and informal

conversations, and this is perhaps a domain where exclamation marks are heavily used and effective. Thus, exclamation marks should not appear frequently in academic contexts. Alipour & Nooreddinmoosa conducted a similar study on informal features in students' writing; they found in their study (2018:358) that out of all informal features which they analyzed, exclamation marks had the lowest frequency. This shows that although exclamation marks are detrimental and not necessary in academic writing, they still do occur in low frequencies. It is unknown whether there is a specific group that is fond of using exclamation marks. However, there is no difference when it comes to the frequency in the use of exclamation marks between native and non-native speakers. (Alipour & Nooreddinmoosa 2018:362) Although this study is not concerned with the demographics of the authors, the frequency of exclamation marks among different disciplines will be scrutinized. The results of different periods will be compared to investigate which authors are using exclamations more than in the previous decades and thus contributing to the informality of their works.

2.4. Summary

Although the presented views on informality are mainly prescriptivist approaches, informal traits in texts are not necessarily detrimental. The overuse of informal traits may cause a text to be incomprehensible, yet the absence of informal traits may generate a text which is potentially complex and, again, too difficult to comprehend. If used in moderation, informal traits may benefit a text and make it reader-friendly, and this is the functional side of informal traits. Lastly, different genres use different vocabulary and language. Thus, the language use among different genres varies.

3. Methodology

The major objective of this study is to look at the development of informal elements in academic journals. To provide an answer to the aforementioned research question, which is "To what extent on average has informality increased in academic journals?" a study similar to Hyland & Jiang's 2017 paper has been conducted to investigate informal traits in academic journals. This study focuses on the analysis of 10 informal elements in 40 years

among four unrelated academic disciplines. Although informality can be expressed in numerous ways, for example by using inappropriate and informal vocabulary, or a personal and direct writing style, this approach, which focuses on an extensive list of informal traits, is rather straightforward and encompasses a majority of elements from the informal domain. Scrutinizing the presence or absence of those informal traits will provide an answer to the research question. If the research focused on and analyzed vocabulary, it would potentially entail at least 20 informal lexemes. As authors from different disciplines use overall different vocabulary, it would be difficult, if not impossible, to create a list of informal words which occur in all four disciplines. Thus this is not the optimal way to approach this topic. Furthermore, if the analysis of the writing style of at least 100 authors per discipline was used as an approach in a study to provide an answer to the same research question, it would be extensively strenuous, if possible at all. Consequently, this approach, which focuses on frequencies of informal traits is highly advantageous, straightforward, and objective. This chapter will focus mainly on the data; namely, on the process of data extraction, data categorization, and how unusable data was handled. All programs which have been used for the analysis are mentioned in the analysis tools section, followed by a minimal description of participants and materials which have been adapted from other authors. Before the very end of this chapter, one section will provide an introduction to the corpus that has been compiled for this study. Although the study was conducted successfully, the last section in this chapter will provide more insight into the various complications concerning data.

3.1. Corpus

Academic journals have been selected as a basis for analysis as they should be free of informal elements. This is because academic articles should be void of any informal elements. Because no existing corpus was suitable for this study, a corpus has been compiled; it consists of articles from open-access academic journals. To compile a balanced corpus, articles from two social and two natural sciences, linguistics, economics, mathematics, and biology respectively were extracted and adapted into data sets. The data

that represents the social sciences entails articles from the following eight journals: *De Economist*, *Journal of Economics (J Econ)*, *Review of World Economics*, *Cambridge Journal of Economics*, *Journal of Linguistics*, *Theoretical Linguistics*, *Annual Review of Applied Linguistics and Language*. For the natural sciences, the data from *Journal of Experimental Biology*, *The Quarterly Review of Biology*, *The Biological Bulletin*, *The American Naturalist*, *Advances in Mathematics*, *Acta Mathematica*, *Manuscripta Mathematica* as well as *Mathematical Notes* was extracted. Although all journals have articles in English, they are published by publishers from different countries. Based on the publishing house and editors, 6 journals are from the USA, 5 are from Germany, 4 are from the UK, and one journal is from the Netherlands. Moreover, it is interesting to note that all of the journals from table 1 are still active and publishing journal volumes. The corpus entails data from four journals per discipline, with approximately 40.000 words per journal. Moreover, instead of focusing on an equal number of articles per journal, the total word count was regarded during data extraction. This is due to the reason that not all journals have a uniform word limit per article. To retain a balanced corpus, an attempt was made to extract an average total of 40.000 words per journal, although certain journals had longer articles on average, finally exceeding the set word limit. Furthermore, the self-compiled corpus entails data from three time periods: 1980, 2000, and 2022. It consists of a random sample of articles from four disciplines and three-time periods. As the data is representative of the average language used in each journal, it is possible to use this corpus for other studies. The total word count of the corpus exceeds 2.000.000 words, and it equally represents all four disciplines. To be precise, the entire corpus encompasses 2.126.685 words, from which 1.745.136 words are usable data. One issue, that Hyland & Jiang (2017) did not mention in their study, is the elimination of unusable data, such as quotations, formulas, equations and mathematical constants, and example sentences. In the data that has been extracted from journals, the unusable data ranges from 5% to 45%. While a manual data clear-up of 2.000 words would not pose a problem, deleting formulas which add up to 18.000 words would have been overly strenuous. Consequently, a random sample of 10.000 words as well as one single complete article was extracted from each discipline, and the unusable data was

manually cleared from those samples. The resulting word count was compared against the starting word count and thus a percentage of usable words was calculated. This process was repeated four times, for all four individual disciplines.

Table 1: Overview of the word count per journal, year, and discipline

JOURNAL	Running time	DISCIPLINE	1980	2000	2020	Total WC per journal	Total WC per discipline
Journal of Experimental Biology	1923 – present	Biology (85%)	44.124 /37.505	45.749 /38.887	44.029 /37.425	133.902 /113.817	538.132 /457.412
The Quarterly Review of Biology	1926 – present	Biology (85%)	43.409 /36.898	42.145 /35.823	52.018 /44.215	137.572 /116.936	
The Biological Bulletin	1897- present	Biology (85%)	45.660 /38.811	42.754 /36.341	41.531 /35.301	129.945 /110.453	
The American Naturalist	1867 – present	Biology (85%)	40.992 /34.843	43.846 /37.269	51.875 /44.094	136.713 /116.206	
Advances in Mathematics	1961 – present	Mathematics (55%)	47.433 /26.088	45.597 /25.079	45.260 /24.893	138.290 /76.060	553.507 /304.429
Acta Mathematica	1882 – present	Mathematics (55%)	43.568 /23.962	45.773 /25.175	53.489 /29.419	142.830 /78.556	
Manuscripta Mathematica	1969 – present	Mathematics (55%)	42.702 /23.486	44.315 /24.373	50.352 /27.694	137.369 /75.553	
Mathematical Notes	1967 – present	Mathematics (55%)	42.604 /23.432	45.581 /25.070	46.833 /25.758	135.018 /74.260	
De Economist	1852 – present	Economics (95%)	45.520 /43.244	38.999 /37.049	43.124 /40.968	127.643 /121.261	510.875 /485.331
Journal of Economics (J Econ)	1930 – present	Economics (95%)	41.149 /39.092	40.766 /38.728	42.850 /40.708	124.765 /118.528	
Review of World Economics	1913 – present	Economics (95%)	39.708 /37.723	42.247 /40.135	46.560 /44.232	128.515 /122.090	
Cambridge Journal of Economics	1977 – present	Economics (95%)	40.371 /38.352	43.351 /41.183	46.230 /43.919	129.952 /123.454	
Journal of Linguistics	1965 – present	Linguistics (95%)	45.346 /43.078	44.950 /42.703	45.808 /43.518	136.104 /129.299	524.171 /497.964
Theoretical Linguistics	1975 – present	Linguistics (95%)	54.083 /51.379	49.922 /47.426	38.728 /36.792	142.733 /135.597	
Annual Review of Applied Linguistics	1980 – present	Linguistics (95%)	41.068 /39.015	44.102 /41.897	43.088 /40.934	128.258 /121.846	
Language	1925- present	Linguistics (95%)	31.411 /29.840	40.998 /38.948	44.667 /42.434	117.076 /111.222	
							2.126.685 /1.745.136

Table 1 presents a comprehensive overview of the raw word count, as well as the usable word count and a percentage of usable words per discipline. The percentage of usable words ranges from 55% in mathematics, to 85% in biology, and 95% in linguistics and economics. The fact that not all disciplines contain an equal amount of usable data should not cause any concerns as the finite numbers will be normalized frequencies.

The data also provides the exact number of words from each discipline. Moreover, at different times, it was difficult to obtain data from the 1980s, and finally, numerous prominent journals were excluded from this research as they only provided volumes from 1985 or even 1891 onwards. Similarly, certain journals, such as Language, provided articles from the 1980s yet the word count from available volumes amounted to only 31.411 words. Although this is not equal to the set average of 40.000 words, it was still included in the analysis. This is partly because the subsequent periods entail a sufficient number of words, but also because Language fulfills every other criterion such as being peer-reviewed or continuing to publish after 2020. Moreover, this journal was used for analysis due to the scarcity of data, which is mentioned in the last section of this chapter.

3.2. Data extraction, categorization, and clear-up

The data which was used in this study is exclusively from open-access academic journals. This is because academic journals are supposed to have a high level of formality, and should therefore not entail any informal traits. If formal language was becoming more common in informal contexts such as blogs or comment sections, this trend would not be detrimental. Moreover, academic, peer-reviewed journals have been selected as a basis for data extraction as these are, as already stated, rather formal genres and are supposed to be resistant to change. The reason why academic genres are not prone to change is according to Hyland & Jiang: "in research writing adherence to the conventions of formality suggests impartiality, precision, distance, and a faux egalitarianism, allowing authors to construct themselves and their readers as disinterested specialists". (2017:41) Furthermore, the data which was used for this study is exclusively in English as English is a global language. The

data was collected from journals that started publishing volumes from or before 1980 and continued until 2020. This is because these journals are within the period which was selected as a criterion for this study. The first step of the data extraction process was to find journals that have been published without interruption since 1980. The next step was to download all articles from the first volume and issue of each year: 1980, 2000, and 2020. All of these files were almost exclusively in pdf format; thus their contents were copied and pasted into Notepad to clear them of any formatting. The text was finally transferred into a Microsoft Office Word file. All data sets were kept separate from the beginning: this was done to provide a detailed overview of the developments among disciplines and periods. A strict division of all data sets was beneficial due to the large size of each sub-corpus and it enabled a detailed overview for every sub-group. The average sub-corpus size was set to be 40.000 words per genre, journal, and period. In a few instances, the articles were of shorter length and the total sum of words from a single issue did not reach the set average of 40.000 words per journal and period. In this case, additional articles from subsequent volumes and issues from that year were added. Several journals could not be used for data extraction as the articles contained a low word count, even though these journals fulfilled the rest of the criteria. All 48 files were marked with a single letter, such as *M* for example, which specified to which discipline a text file belonged; in this case, it was mathematics. This letter is followed by an acronym of the title of the journal from which the data was extracted, and the last piece of information in the name of each file was the period from which it originated. Thus, all file names have the following format with slight variations: M-AM-1980. This enabled clear-cut and facilitated access to individual journals, periods, and genres. Collecting data from 16 journals and three time periods each into an individual file was a time-consuming process. In addition, every element which was not part of the running text was eliminated: this for example includes abstracts, acknowledgments, references and appendixes, tables, remarks, and annotations.

3.3. Analysis tools

After the data extraction and classification, three distinct programs were necessary to obtain raw numbers. There have been multiple steps before the raw numbers were obtained: uploading the file, running the scan, identifying a feature based on context, copying, recording the instances in a table, and copying and transferring each example. The first round of analysis, which would result in the raw numbers for 9 out of 10 traits, was performed using an online text-analysis software called SketchEngine; when data analysis has taken place, the basic version of SketchEngine was available to all members from the University of Vienna: The 48 text files were uploaded individually, and the search option was used to locate individual informal traits within approximately 40.000 words. It is important to note that SketchEngine only scanned the text and provided all hits; it did not recognize the word class of any word. The total number of occurrences of an item was recorded in a table in a Microsoft Word file under the column total. All hits were inspected based on their context; if they were representing an informal feature, they were counted, extracted, and transferred into the column actual. These files thus provide the total number of hits, the number of actual representative examples, as well as record all example sentences. All of the following traits are simple in form and were not complicated to find: I, we, you, your, and so on, and so forth, etc., et cetera, But, And, ?, !, ‘, This, These, That, Those, It, To., on., for. with., of. The mentioned lexemes are at the same time the query which was entered into SketchEngine. However, in the case of split infinitives, the text scanning was slightly more difficult. Namely, the first issue is that the self-compiled corpus is not tagged. Thus, the second tool which was necessary to continue with the analysis is TagAnt. It is a program by Laurence Anthony which was used to tag the self-compiled corpus. The second issue was that split infinitives are slightly more complex in structure than the aforementioned informal traits and locating them in the text required a more complex search query. Consequently, SketchEngine was not suitable for this analysis. Split infinitives have a straightforward structure and consist of: *to*, followed by an adverb, which is then followed by a verb. Thus, another tool was necessary to complete the data analysis of all 10 informal traits. Finally, each data set was opened in AntConc and the following

query, which reflects the structure of split infinitives, was used: to_TO *_RB *_Vb. All results were examined one after another, and the numbers of actual occurrences, as well as the representative examples, were transferred into a table in a separate Microsoft Word file. As all the data sets are divided, to compare and contrast the relevant data, the numbers were transferred into tables. To track the development of a specific informal trait in a single journal, the numbers of that feature from different periods were entered into a table. Moreover, to provide an overview of the development of a specific informal element in one discipline, the data sets of four journals were combined and transferred into one table per period. The last step was to calculate the normalized frequency, based on each raw number. The normalized frequencies were manually calculated with the use of a calculator. The following formula was used to calculate the normalized frequencies:

number of occurrences of one trait / total word count of 1 journal from one period= x

x* 1.000.000 = normalized frequency of one trait

The normalized frequencies enabled an objective and correct comparison of the results. Because almost every subsection contains an almost equal number of words, the raw numbers are a good indicator of the ultimate numbers of normalized frequencies. The raw numbers were thus used to review the accurateness of normalized frequencies; if a raw number in a certain column deviated to a great extent from the same column in the table with the normalized frequencies, the normalized frequency was calculated a second time. This approach highlighted a few invalid results which were corrected. For the visual representation of the data, various figures have been created and exported with Microsoft Office Excell.

3.4. The authors of the texts

This study features the works of approximately 500 authors. The study includes male and female authors from different age groups, with different first languages, and education levels as well as native and non-native speakers of English. It is important to mention that

there are no active participants in this study. Aside from the various authors from academic journals who have had their works published in the past by the respective journals, there are no other participants or authors. This study is focused on the development of informality as a whole; it is not concerned with the writing styles of male and female authors or within different age groups. However, as the study focuses on individual informal traits in different academic genres, a distinction has been made between authors from different disciplines, as well as different generations of authors of a specific journal.

3.5. Materials

With regards to materials, this study adopts a single set, consisting of 10 criteria. More precisely, the informality of the data has been gauged based on the frequencies of 10 informal traits. Namely: first-person pronouns to refer to the author, unattended anaphoric pronouns, split infinitives, sentence-initial conjunctions or conjunctive adverbs, sentence-final prepositions, listing expressions, second-person pronouns/determiners to refer to the reader, contractions, direct questions and exclamations which have been adapted from Chang & Swales. (1999:148) All three time periods have been analyzed based on the 10 informal traits. This list was used because this is a replication study, and the original study also used the same list. Moreover, the list is comprehensive and it will provide insight into the developments of various informal features within distinct disciplines in distinct decades, as well as a general picture of informality in different fields over 40 years. Moreover, other studies have used this particular list of informal features as a basis for their analysis in the past for similar studies. The data has been strictly classified since the beginning and kept separate for each journal and the three-time periods. Although the period between the three periods is not extensively large, it is undeniable that the lives of people, as well as

access to literature and education overall, differed to a large extent. The search queries which were used to find the informal traits are presented in table 2.

Table 2: Informal traits and their search queries

Informal feature	Form(s)
first person pronouns	I/we
second-person pronouns/determiners	You/your
sentence-initial conjunctions or conjunctive adverbs	But/And
unattended anaphoric pronouns	This/These/That/ Those/It
listing expressions	and so on/and so forth/etc./ et cetera
Contractions	‘
direct questions	?
Exclamations	!
sentence-final preposition	to/on/for/with/of
split infinitives	To adverb verb

Table 1 shows that there is no one-to-one correspondence and that one informal trait might have more than one form. Although all forms are individual words, if they are representative of one informal trait, their frequencies were summarized and treated as a single raw number. In other words, the analysis does not feature details of the frequency of each form of an informal trait. This is due to the reason that the study focuses overall on informal traits in formal contexts, and not for example, on the various personal pronouns which have occurred in academic writing in the past decades.

3.6. Issues and scarcity of data

Throughout the data collection process, numerous inconveniences have occurred. This study aims to be representative and reliable; thus numerous criteria have been set before the data collection has taken place. Consequently, collecting the appropriate data was more difficult than anticipated and numerous open-access journals had to be excluded from this study. One of the issues is the access to and availability of journals that were published before 2000. For the majority of journals, although they published their first volume in 1875 for example, they almost exclusively provide data from 1995 onwards. Unlike Hyland & Jiang (2017), this study did not include volumes or journals, if they were not dating from the

stated period, which is 1980 in this case. In addition, more than 75% of all journals could not be used for data extraction because they have been discontinued by 2015 or even earlier. Although finding articles that have been written and published in 1980 has been an extensively strenuous process, all 16 journals from this study, have started publishing in the 1980s or before and have continued to publish until 2020 or even after. The second problem occurred with bilingual journals. This is because numerous available journals entail articles in English and, for example, French or Spanish. Including the articles in English, and disregarding the articles in other languages seems like a plausible suggestion for solving this issue. However, although the English articles in a bilingual journal are flawless, their total word count does not amount to the set limit of 40.000 words on average. The total sum of English words from a bilingual journal is approximately 20.000 words, and this is not sufficient. Finally, no articles from bilingual journals have been used during the compilation of the corpus. Peer-reviewed journals are salient for this study. This criterion has prevented a plethora of open access journals from being included in this study. However, investigating informal traits in a journal that has not been peer-reviewed would not yield interesting results. Consequently, all 16 journals which have been used to extract data, are peer-reviewed. Furthermore, another issue is the categorization of articles in journals. It appears that a few decades ago, numerous subgenres were published by one journal. Thus, the variety of articles per journal was wider. This was problematic as the access is already limited by numerous aforementioned factors, and a statistics article in a mathematical journal differs to a great extent from a mathematical article with constants and formulas. Although there might be a single statistics article per volume, it consists of approximately 8.000 words, and due to the scarcity of data, these 8.000 words might finally decide if the remaining data from the entire journal can be used for the analysis or not.

Thus, in certain cases, more or all volumes from one temporal period of a specific journal have been used to obtain data, as not all articles from a single volume have been appropriate in terms of content. Another issue, albeit technical, is that numerous journals consist of scanned images of journal pages. Consequently, they are not machine-readable and cannot be analyzed. Numerous programs and converters were used to attempt and

make the data useable, however, no solution has been successful. As a consequence, numerous, potentially more influential, journals were not part of the corpus. Two disciplines were especially problematic to analyze. The most difficult data sets that have undergone analysis have been extracted from mathematic journals. This is because numerous symbols, such as ? ! or even ‘ are used in various mathematical formulas and equations. Another reason is due to the technical nature of mathematics, there are a plethora of examples and constants which also influence that language use in the respective discipline. Moreover, about analysis, linguistics is almost as complex as mathematics. This is because there are a plethora of rather informal example sentences. If an informal trait occurs in a sample sentence in a linguistics article, it cannot be counted as a representative informal trait as it is not a part of the running text, nor was it written by the authors. Therefore, deciding whether thousands of individually marked occurrences, based on the context, are representable of an informal trait was a tedious process. Lastly, a trend that demonstrates a diminishing number of total words per article has been noticed; it appears that articles tend to become shorter as the year progresses. This is the case for almost all journals, and as a consequence, numerous publishing houses have decreased the word limit, yet increased the number of articles. However, one entire journal was excluded from the data analysis because the last period, 2020, did not entail a sufficient number of words, which was set to be an average of 40.000 words.

3.7. Results

This section will focus on the results of the approach that was explained in the previous section. The raw frequencies of informal traits were distributed into various tables which focus on the total number of informal traits per discipline in all-time spans, a total of informal traits per discipline per year, and individual informal traits per discipline per year. In addition, some tables provide an overview of each informal trait as well as in which discipline it was dominant in three different periods, as well as an overview of informal traits in four journals within one discipline. The raw numbers were subsequently turned into normalized frequencies; each number is dependent on the size of the total corpus of

one discipline, or on the sub-corpora from one discipline, which is divided into three distinct periods. Moreover, each section will focus on a different perspective and might answer one of the research questions. The results will be presented from more general numbers to more specific ones; namely, from a broad overview of informal traits in all disciplines, an analysis and comparison of all informal traits in each period in one discipline, analysis and comparison based on discipline, analysis and comparison based on individual traits in three distinct periods, to an analysis based on journals. Finally, two statistical significance sections follow. These will deal with a statistically significant correlation between period and informality and the relationship between publishing country or editor and journal. Moreover, three detailed tables per discipline, which provide insight into the developments within journals, are attached in the appendix. The number in the following sections are normalized frequencies (per million). Finally, the answer to the first research question is that the study does replicate Hyland & Jiang's findings: "While academic writing is becoming more informal, this is by small margins and depends on the discipline and features being considered." and "Even in our corpus, the process of change is not occurring in all disciplines or in all features." (2017:48) The following sub-chapters will provide more details and proof.

3.8. Analysis and comparison of all informal traits in each discipline

This section will provide a rather general overview of all 10 informal traits in the four selected disciplines. This is necessary to find out which discipline has the lowest and highest number of informal traits overall; this specific section is dealing with quantitative data. This is done to answer the second research question which is: are certain genres more informal than others and if they are to what extent do they differ? For each subject, the normalized frequencies of one informal trait from all three time periods were combined into one number; all numbers are presented in table 3.

Table 3: Total number (normalized frequencies) of informal traits per subject

	1st pers. pron.	2nd pers. pron.	Listing expressions	Initial conj.	Exclamations	Direct questions	Contractions	Unatt. anaph. pron.	Stranded prep.	Split infinitives	Total
BIOLOGY	3.766,84	2,18	6,558	69,95	4,37	170,52	2,18	811,08	2,18	135,54	11.522,84
ECONOM.	5.859,91	103,02	30,90	513,05	26,78	432,69	16,48	1.629,81	65,93	123,62	8.802,18
LINGUIST.	5.962,27	68,27	98,4	815,31	16,06	592,41	46,18	1.494,08	108,44	138,56	9.339,98
MATHEM.	23.808,50	16,42	19,70	558,42	13,13	62,41	29,56	1.658,84	36,13	49,27	26.252,38

Table 3 provides an overview of the total number of occurrences of each informal trait in each discipline. This enables an analysis of the numbers of individual traits, and in this case, mathematics is not always at the top. Namely, in the case of first-person pronouns, mathematics has indeed the highest number of instances: 23.808,5. In the case of second-person pronouns, economics has the highest number of occurrences, which is 103,02. Listing expressions are most commonly found in linguistics, with a total of 98,4. Sentence-initial conjunctions are most prevalent in linguistics with 815,31 instances. Exclamations are mostly found in economics, where they occur 26,78 times per million words. Direct questions occur most frequently in linguistics: a total of 592,41 instances. This is also true in the case of contractions, which are also most commonly, with 46,18 hits, featured in linguistics. Mathematics also entails the highest number of unattended anaphoric pronouns, namely 1.658,84 unattended anaphoric pronouns. Stranded prepositions are also most commonly found in linguistics, in this case, a total of 108,44 times. Finally, split infinitives are most commonly found in linguistics articles and biology articles: with 138,56 and 135,54 results respectively. In other words, it appears that a set of informal traits is characteristic of each discipline. Economists use 2nd person pronouns more often than any other group, while mathematicians use 1st person pronouns and unattended anaphoric pronouns the most, and linguists tend to use listing expressions, sentence-initial conjunctions, exclamations, direct questions, contractions, stranded prepositions, and split

infinitives. While each informal trait occurs at least once in biology, neither informal trait occurs more often in biology than in any other discipline. Moreover, linguists use split infinitives slightly more often than biologists: namely, there are 138,56 and 135,54 split infinitives respectively. Although a table provides a good general overview of the individual traits, a bar chart is more suitable to demonstrate the difference between numbers of informal traits among disciplines.

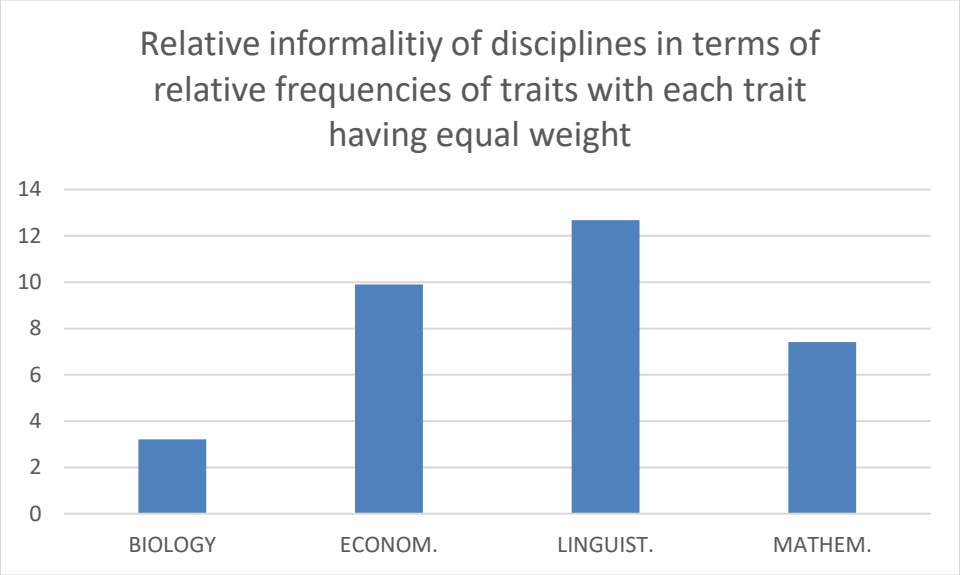

Figure 1: Comparison of the relative numbers of informal elements in each discipline

Based on the largest bar from figure 1, linguistics is the discipline that has the highest number of informal elements. Namely, it is almost 25% larger than the 2nd greatest bar, and it is overall larger than the bars which represent informal traits in economics, mathematics, or biology. Moreover, linguistics entails almost as many informal words as biology and economics combined; namely, the relative frequency of informal traits in linguistics is 12,68 while the relative frequencies for economics and biology are 9.89 and 3.21 respectively. As mentioned, linguistics has the highest relative frequency of informal traits, which is followed by economics, mathematics, and biology; their relative frequencies are 12.68, 9.89, 7.41, and 3.21 respectively.

Finally, it would be interesting to look at the percentages of informal traits in the four individual corpora. In mathematics, informal elements take up 8.62% of all words. Similar to

the aforementioned results, mathematics is followed by biology which consists of a total of 2.52% informal elements in the corpus. Lastly, linguistics and economics have the lowest percentages of informal items: namely, 1.87% and 1.81% respectively. Finally, certain genres are overall more informal than others. For example, mathematics entails the highest number of informal traits in comparison to other disciplines, as well as the highest percentage of informal elements in the corpus of the mathematics discipline. In terms of frequency of informal traits and percentage of informal traits within a specific corpus, it is followed and is followed by biology, linguistics, and economics. This means that economics is overall the least informal discipline. Thus, the second research question has been answered.

3.9. Analysis and comparison of all informal traits in each period in one discipline

This is the most salient section because it will focus on the total of all informal elements in one discipline in one time period. As all 10 selected informal traits are regarded to be equally informal, this will provide an overview of how the frequency of informal traits has changed throughout the three selected periods in each discipline. This is crucial to answering the main research question, which is whether academic genres have become more informed throughout the years. This section will entail quantitative data and the numbers of informal traits will be scrutinized in this section in the following order: mathematics, biology, linguistics, and economics. First is mathematics; figure 2 provides information on the developments of informality in mathematics.

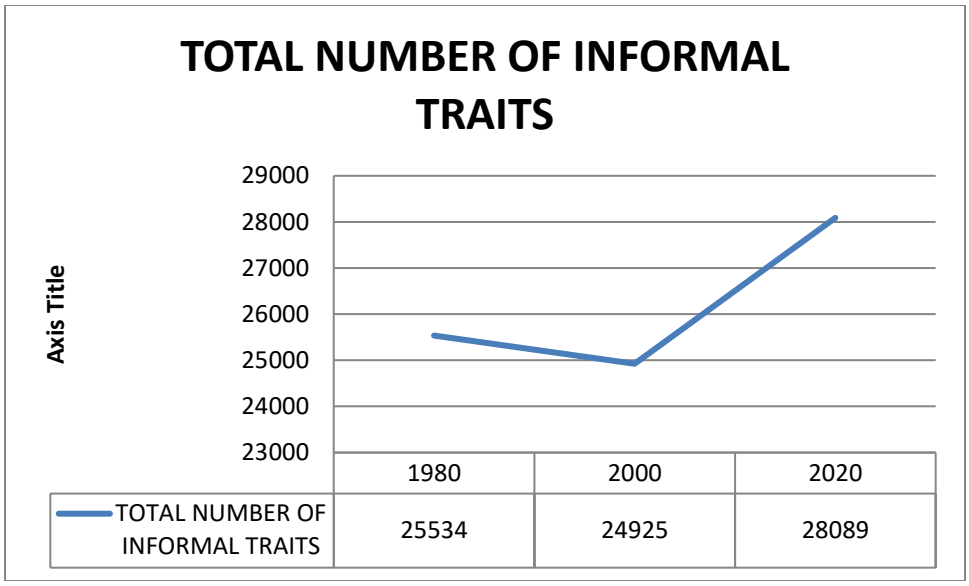

Figure 2: Total of informal traits in mathematics

Informal traits in mathematics slightly decreased in the first two time periods, namely, they dropped from 25.534 in 1980 to 24.925 in 2000. At last, the total of informal traits has increased from 24.925 in 2000 to an all-time high of 28.089 in 2020. On the whole, informal traits in mathematics have increased from 1980 to 2020. However, in the particular case of mathematics, it is important to scrutinize the individual periods and informal traits. Thus, figure 3 has been provided.

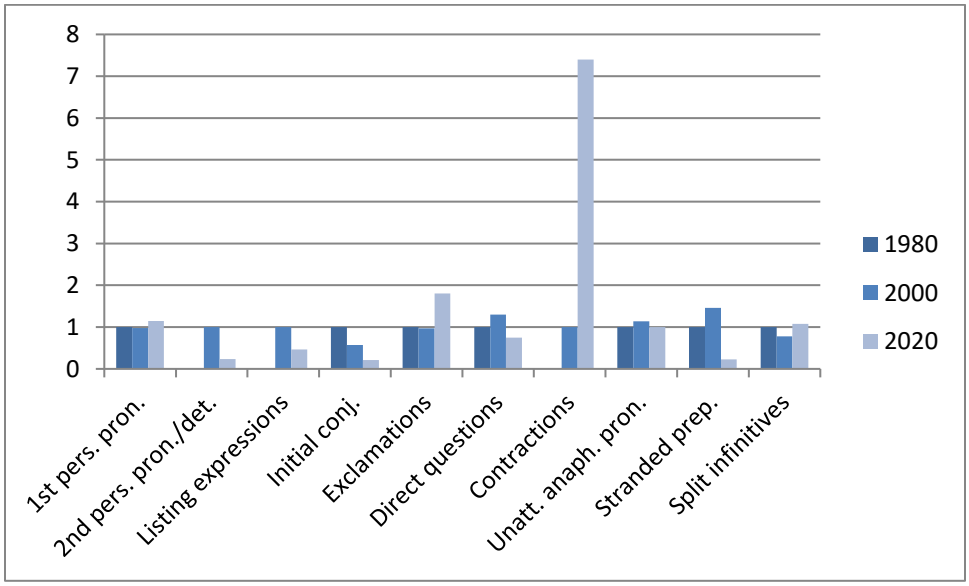

Figure 3 Relative changes in informal traits in mathematics

Figure 3 shows the normalized frequencies of individual informal traits in each period in mathematics. Although informal traits in mathematics have increased overall, that does not suggest that all traits have increased and that none have decreased in frequency.

First-person pronouns have decreased slightly from 0.96 in 1980 to 0.94 in 2000, before rising to 1.09 in 2020. Second-person pronouns did not occur initially in 1980, however, they increased to 2.43 in 2000, before dipping to 0.56 in 2020. Listing expressions were also initially not present, but by 2000 they increased to 2.05 before dipping to 0.94 in 2020. Sentence-initial conjunctions have decreased from 1.67 to 0.96, before decreasing again to 0.35 in 2020. Exclamations have slightly decreased from 0.79 to 0.77, before increasing to 1.43 in 2020. Direct questions have increased from 0.98 to 1.27 before decreasing to 0.73. Contractions were not found initially, however, by 2000 they increased to 0.35, before increasing again to 2.64 in 2020. Unattended anaphoric pronouns have slightly increased from 0.95 in 1980 to 1.08 in 2000, before decreasing again to 0.95. Stranded prepositions have increased from 1.11 in 1980 to 1.63 in 2000, before dropping to 0.25 in 2020. Split infinitives have decreased from 1.04 in 1980 to 0.81 in 2000, before increasing to 1.13 in 2020 again. Moreover, figure 4 shows how informality developed in biology.

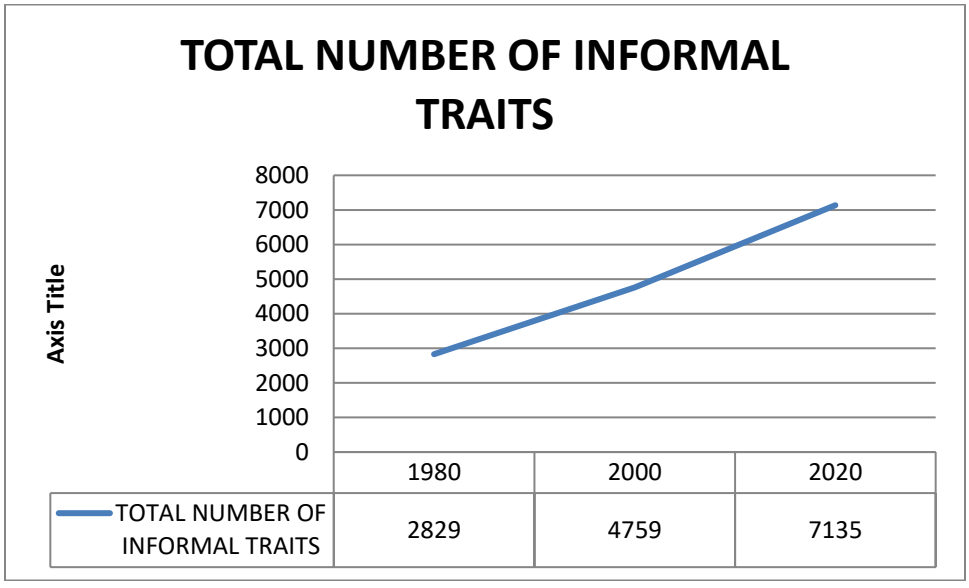

Figure 4: Total of informal traits in biology

As figure 4 shows, informal traits in biology developed steadily from 1980 to 2020.

In biology, the frequency of informal traits has increased from 2.829 hits in 1980 to 4.759 hits in 2000 before increasing again to 7.135 in 2020. Together, these numbers are indicative of an increase in informality in biology. However, just like in mathematics and all other disciplines, that does not suggest that every informal trait has steadily increased. Thus, figure 5 has been provided.

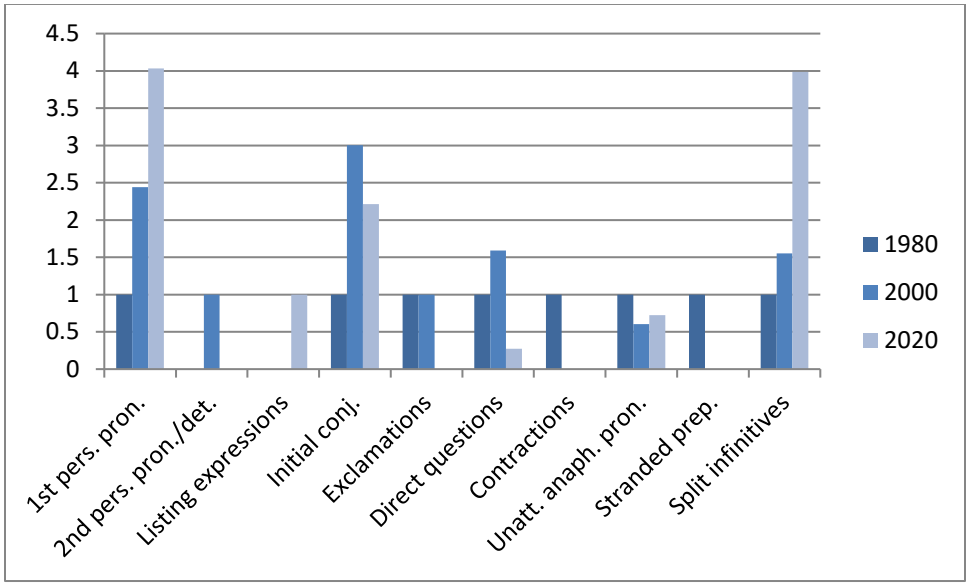

Figure 5 Relative changes of informal traits in biology

Figure 5 shows the increase or decrease of each informal trait. First-person pronouns have increased steadily from 0.40 in 1980 to 0.97 in 2000, before increasing to 1.61 in 2020 again. Second-person pronouns were initially not found in biology in 1980, yet they have increased to 3 in 2000 before dropping again to 0 in 2020. Listing expressions did not occur at all in 1980 nor 2000, yet in 2020 they increased to 3. Sentence-initial conjunctions have increased from 0.48 in 1980 to 1.44 in 2000 before decreasing to 1.06 in 2020. Exclamations have slightly decreased from 1.50 in 1980 to 1.49 in 2000, before completely disappearing by 2020. Direct questions have increased from 1.04 in 1980 to 1.66 in 2000 before dipping to 0.25 in 2020. Contractions have only occurred in 1980 with 3 instances, after that they completely disappeared from the next two periods. Unattended anaphoric pronouns have decreased from 1.28 in 1980 to 0.77 in 2000 before increasing to 0.93 in 2020. Stranded prepositions developed in the same way as contractions: dropping from 3 in 1980 to 0 in the subsequent two periods. Split infinitives have increased from 0.45 in 1980 to 0.71 in 2000 before increasing again to 1.82 in 2020.

Figure 6 shows the trend in linguistics.

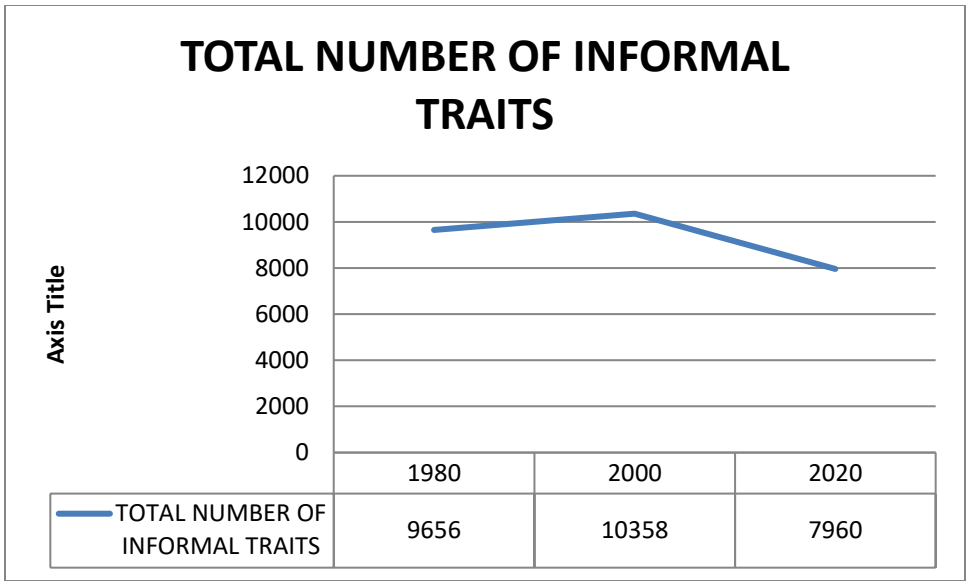

Figure 6: Total of informal traits in linguistics

Based on the numbers from figure 6, informal traits in linguistics have increased at a certain point. Namely, they slightly increased from 9656 in 1980 to 10358 in 2000, before sharply declining to 7960 in 2020. However, in general, the number of informal traits in linguistics has slightly decreased. Figure 7 shows the developments of individual informal traits.

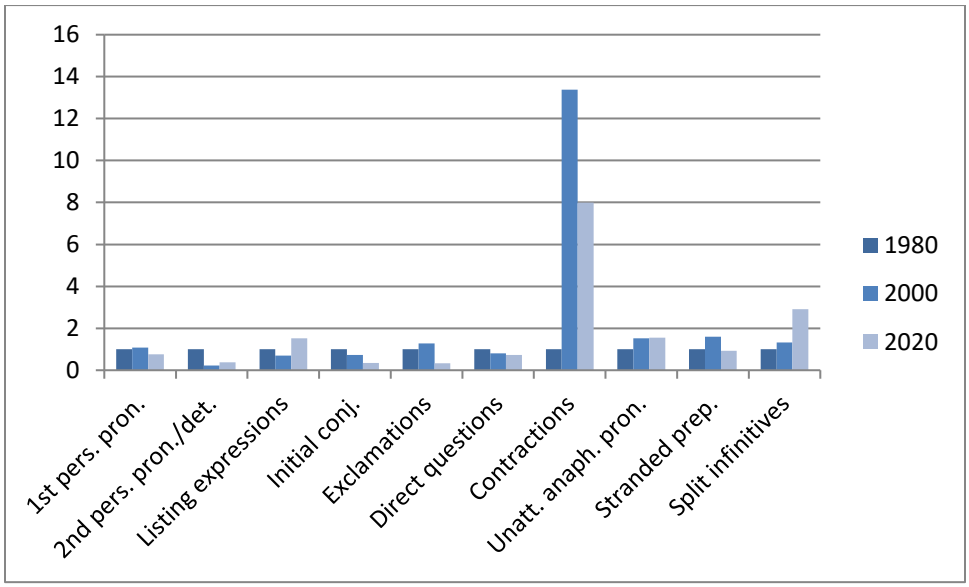

Figure 7 Relative changes of informal traits in linguistics

First-person pronouns in linguistics have increased from 1.05 in 1980 to 1.14 in 2000, before dropping to 0.80 in 2020. Second-person pronouns have decreased from 1.86 in 1980 to

0.42 in 2000, before increasing to 0.70 in 2020. Listing expressions have a similar development to second-person pronouns: they decreased from 0.92 to 0.65, before increasing again to 1.42. Sentence-initial conjunctions have decreased first from 1.44 in 1980 to 1.05 in 2000, before decreasing again to 0.50 in 2020. Exclamations have increased at first, from 1.15 in 1980 to 1.46 in 2000, before decreasing to 0.38 in 2020. Direct questions have continued to decrease: from 1.17 in 1980 to 0.95 in 2000, before decreasing again to 0.86 in 2020. Contractions have dramatically increased from 0.13 in 1980 to 1.79 in 2000, before decreasing to 1.07 in 2020. Unattended-anaphoric pronouns have decreased overall: from 0.73 in 1980 to 1.12 in 2000, before decreasing again to 1.14 in 2020. Stranded prepositions have increased from 0.85 in 1980 to 1.35 in 2000 before dipping to 0.79 in 2020. Lastly, split infinitives have continued to increase: from 0.57 in 1980 to 0.75 in 2000 before increasing again to 1.67 in 2020.

Finally, the developments in economics will be scrutinized in figure 8.

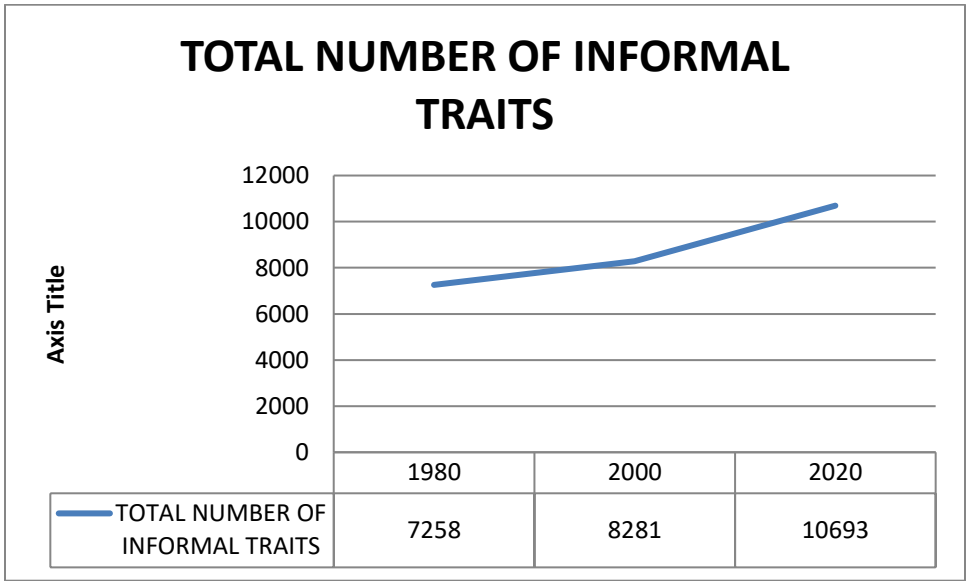

Figure 8: Total informal traits in economics

Informal traits in economics have increased by approximately 12%, before increasing again by approximately 23%. More precisely, informal traits increased from 7.258 in 1980 to 8.281 in 2000 before increasing one more time to 10.693 in 2020. These extreme changes caused a steady increase of informal elements, and as a result, the language in economics articles is

more informal in 2020 than it was 20 or even 40 years ago. Although the frequencies of informal traits in this section are already normalized, they differ to a large extent in quantity. Thus figure 9 has been provided.

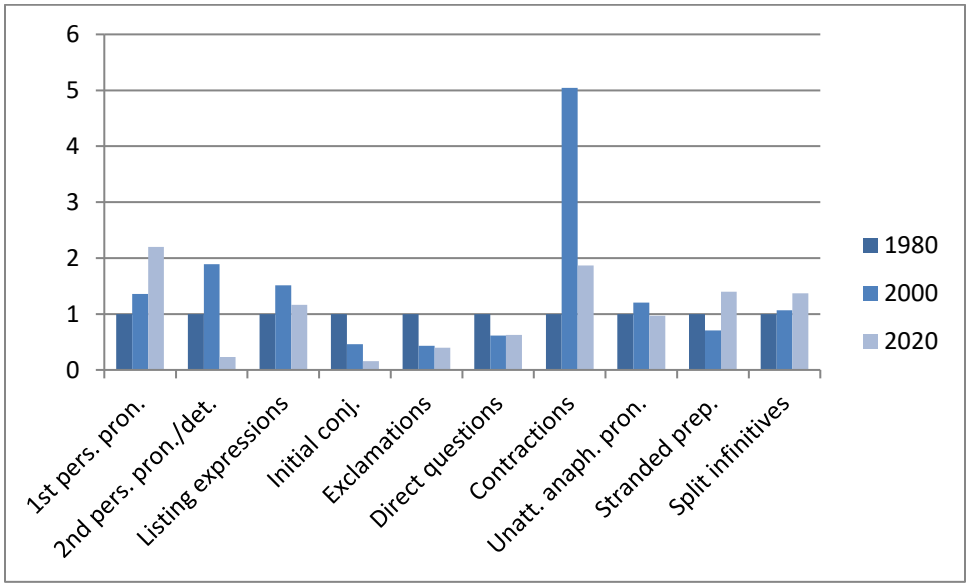

Figure 9 Relative changes in informal traits in economics

Figure 9 presents the developments of each informal trait in economics. First-person pronouns have continued to increase: starting from 0.65 in 1980 to 0.89 in 2000, before rising again to 1.44 in 2020. Second-person pronouns have increased from 0.96 in 1980 to 1.81 in 2000, before decreasing again to 0.22 in 2020. Listing expressions have the same development as second-person pronouns: an increase from 0.81 in 1980 to 1.23 in 2000, before dropping to 0.95 in 2020. Sentence-initial conjunctions have continued to decrease from 1.85 in 1980 to 0.85 in 2000, before dipping again to 0.29 in 2020. Exclamations have also decreased from 1.63 in 1980 to 0.70 in 2000, before decreasing again to 0.65 in 2020. Direct questions have dropped from 1.33 in 1980 to 0.82 in 2000, before slightly increasing to 0.84 in 2020. Contractions have sharply increased from 0.37 in 1980 to 1.91 in 2000, before increasing to 0.70 in 2020. Unattended anaphoric pronouns have increased from 0.94 in 1980 to 1.13 in 2000 before decreasing again to 0.91 in 2020. Stranded prepositions have decreased from 0.96 in 1980 to 0.68 in 2000, before increasing to 1.35 in 2020. Lastly, split infinitives have increased from 0.87 in 1980 to 0.93 in 2000, before increasing again to 1.19 in 2020.

To be able to compare to what extent informality has increased or decreased, a different type of data is necessary. Thus, table 4, which entails percentages of the aforementioned changes, has been provided.

Table 4: Percentage of increase/decrease of informality

SUBJECT/YEAR	1980 – 2000	2000-2020	1980-2020
MATHEMATICS	-2.38%	+12.69%	+10%
BIOLOGY	+68.22%	+49.92%	+152.20%
LINGUISTICS	+7.27%	-23.15%	-17.56%
ECONOMICS	+14.09%	+29.13%	+47.32

The second column in Table 4 shows the percentage of change from 1980 to 2000, and the third column provides the percentage of change from 2000 to 2020, while the third column shows the overall percentage of change from the first period to the last. The most significant change can be observed in biology, where informality has overall increased by 152.20% in four decades. Economics has also become more informal than it used to be initially, with an overall increase of 47.32% from 1980 to 2020. Mathematics has also become more informal, and informal traits overall increased by 10%. In contrast, the informality in linguistics has decreased by 17.56%. To sum up, informal traits have increased in mathematics, biology, and economics. By contrast, informal traits have overall declined in linguistics.

The aforementioned hypotheses will be restated to check whether they are verified or falsified. The first hypothesis is verified: the informality in natural sciences has increased. By contrast, the second hypothesis is that informality in social sciences has decreased, which is only partially true. Furthermore, the answer to the main research question, which is also Hyland & Jiang's guiding question, is that academic genres have become less formal over the years, except linguistics, which has become more formal throughout the examined period. Moreover, this is also the answer to the third research question which focuses on whether informality has increased in different disciplines over the years.

3.10. Analysis and comparison based on discipline

In addition to whether informality has increased or decreased in different disciplines, it is important to scrutinize how the individual informal traits have developed within all

disciplines. This section will focus on individual disciplines and the individual frequencies of informal traits from 1980 to 2020, and it is both qualitative and quantitative. Moreover, whether a certain group of informal elements has increased in all disciplines, an overview of the developments will provide more information about that. The disciplines will be analyzed in a familiar order: mathematics, biology, linguistics, and finally economics. This chapter will also provide an answer to the 2nd part of Hyland & Jiang's guiding question which is concerned with how the different disciplines have changed.

3.10.1. Mathematics

From 1980 until 2020 four informal traits have been prominent and have increased in frequency in mathematics; namely, 1st person pronouns, listing expressions, exclamations, and contractions.

Table 5: Relative frequencies of individual traits in mathematics throughout 40 years

	1st pers. pron.	2nd pers. pron./det.	Listing expressions	Initial conj.	Exclamations	Direct questions	Contractions	Unatt. anaph. pron.	Stranded prep.	Split infinitives
1980	**0.961386**	0	**0**	1.677791	**0.795116**	0.984668	**0**	0.956636	1.117886	1.049746
2000	**0.940563**	2.436931	**2.051125**	0.965088	**0.773522**	1.277029	**0.357109**	1.087541	1.630894	0.816831
2020	**1.098051**	0.563069	**0.948875**	0.357121	**1.431362**	0.738302	**2.642891**	0.955823	0.25122	1.133424

The numbers of these three traits are in bold in Table 5, and they have different developments. In the case of 1st person pronouns, the number has slightly decreased from 0.96 in 1980 to 0.94 in 2000, before finally increasing to 1.09 in 2020. Similarly, exclamations also slightly decreased from 0.79 in 1980 to 0.77 in 2000, before increasing to 1.43 in 2020. Listing expressions have increased from 0 in 1980 to 2.05 in 2000, before plummeting to 0.94 in 2020, which was still a significant frequency. By contrast, contractions have increased overall: rising from 0 in 1980 to 0.35 in 2000, and finally to 2.64 in 2020. These four traits are salient due to their prominent developments. Other informal traits such as 2nd person pronouns, direct questions, and stranded prepositions all share a similar development, which is a sharp increase before a sharp drop. By contrast, sentence-initial conjunctions have continued to decrease from 1.67 in 1980 to 0.96 in 2000, before dropping one more time to 0.35 in 2020. Finally, unattended anaphoric pronouns and split infinitives have changed in frequency throughout the years, yet their developments were

mostly stable without any dramatic changes. Unattended anaphoric pronouns have increased from 0.95 in 1980 to 1.08 in 2000, before dropping to 0.95 in 2020. Split infinitives have decreased from 1.04 in 1980 to 0.81 in 2000, before increasing again to 1.13 in 2020. This answers in part the fourth research question and shows which informal traits have been dominant in mathematics as well as the fifth research question: namely that first person pronouns and unattended anaphoric pronouns have remained almost equal in frequency throughout the years.

3.10.2. Biology

In biology journals, 1st person pronouns, listing expressions, initial conjunctions, and split infinitives have had prominent developments from 1980 until 2020. This can be seen in table 6.

Table 6: Relative frequencies of individual traits in biology throughout 40 years

	1st pers. pron.	2nd pers. pron./det.	Listing expressions	Initial conj.	Exclamations	Direct questions	Contractions	Unatt. anaph. pron.	Stranded prep.	Split infinitives
1980	**0.40139**	0	**0**	**0.482946**	1.501112	1.048166	3	1.288089	3	**0.45889**
2000	**0.979848**	3	**0**	**1.449269**	1.498888	1.666341	0	0.779776	0	**0.712646**
2020	**1.618762**	0	**3**	**1.067784**	0	0.285493	0	0.932135	0	**1.828464**

The aforementioned four traits are all marked in bold in Table 6. Two traits that had dramatic increases throughout all three periods are 1st person pronouns and split infinitives. First-person pronouns have increased from 0.40 in 1980 to 0.97 in 2000, before increasing again to 1.61 in 2020. Similarly, Split infinitives have increased from 0.45 in 1980 to 0.71 in 2000, before increasing again to 1.82 in 2020. Sentence initial conjunctions have also increased from 0.48 in 1980 to 1.44 in 2000, before slightly decreasing to 1.06 in 2020. Listing expression only increased in 2020 to 3, while having no instances in the previous two periods. Exclamations have slightly decreased from 1.50 in 1980 to 1.49 in 2000, before completely disappearing by 2020. Contractions and stranded prepositions have had the same development: starting with 3 instances in 1980, before completely disappearing by 2000 and having again no instances in 2020. Direct questions have sharply increased from 1.04 in 1980 to 1.66 in 2000, before sharply decreasing to 0.28 in 2020. Unattended

anaphoric pronouns have sharply decreased from 1.28 in 1980 to 0.77 in 2000, before increasing again to 0.93 in 2020. Lastly, 2nd person pronouns have increased from 0 in 1980 to 3 in 2000, before again dropping to 0 in 2020. This is a part of the answer to the fourth research question and specifies which informal traits have been dominant in biology journals. It answers also the fifth research question as no informal traits, aside from second-person pronouns were consistent in frequency in biology throughout 40 years.

3.10.3. Linguistics

The four informal traits which have the highest frequencies in 2020 in linguistics are listing expressions, contractions, unattended anaphoric pronouns, and split infinitives. This can be observed in table 7.

Table 7: Relative frequencies of individual traits in linguistics throughout 40 years

	1st pers. pron.	2nd pers. pron./det.	Listing expression	Initial conj.	Exclamations	Direct questions	Contractions	Unatt. anaph. pron.	Stranded prep.	Split infinitives
1980	1.053826	1.866273	**0.928677**	1.442663	1.151097	1.177428	**0.134142**	**0.735055**	0.851448	**0.572621**
2000	1.144222	0.424404	**0.650499**	1.055032	1.466458	0.956938	**1.794696**	**1.121809**	1.355604	**0.757284**
2020	0.801951	0.709323	**1.420824**	0.502305	0.382445	0.865634	**1.071162**	**1.143136**	0.792948	**1.670096**

The frequencies of the aforementioned three traits are distinctly marked in table 7. For example, listing expressions have slightly decreased from 0.92 in 1980 to 0.65 in 2000, before sharply increasing to 1.42 in 2020. Contractions, on the other hand, have increased significantly from 0.13 in 1980 to 1.79 in 2000, before decreasing to 1.07 in 2020. Unattended anaphoric pronouns have continued to increase from 0.73 in 1980 to 1.12 in 2000, before increasing again to 1.14 in 2020. Split infinitives have increased from 0.57 in 1980 to 0.75 in 2000, before increasing again to 1.67 in 2020.

First-person pronouns have slightly increased from 1.05 in 1980 to 1.14 in 2000, before dropping to 0.80 in 2020. Similarly, exclamations have increased from 1.15 in 1980 to 1.46 in 2000, before decreasing to 0.38 in 2020. Second-person pronouns have sharply decreased from 1.86 in 1980 to 0.42 in 2000, before slightly increasing to 0.70 in 2020.

Sentence-initial conjunctions have continued to decrease, from 1.44 in 1980 to 1.05 in 2000, and 0.50 in 2020. Similarly, direct questions have steadily decreased from 1.17 in 1980 to 0.95 in 2000 and finally to 0.86 in 2020. Lastly, stranded prepositions increased from 0.85 in 1980 to 1.35 in 2000, before decreasing again to 0.79. This highlights which informal traits remained dominant in linguistics and also partially answers the fourth research question. It also provides an answer to the fifth research question: namely, that no informal traits were consistent in linguistics in 40 years.

3.10.4. Economics

Economic journals from all three time periods have been characterized by the following informal traits: 1st person pronouns, unattended anaphoric pronouns, and sentence-initial conjunctions. The numbers of the dominant traits are highlighted in table 8.

Table 8: Relative frequencies of individual traits in economics throughout 40 years

	1st pers. pron.	2nd pers. pron./det.	Listing expressions	Initial conj.	Exclamations	Direct questions	Contractions	Unatt. anaph. pron.	Stranded prep.	Split infinitives
1980	**0.65879411**	0.960228	0.815568	1.852007	1.637509	1.337818	**0.379359**	0.946715	**0.966172**	**0.872203**
2000	**0.893803575**	1.815877	1.233527	0.854423	0.707932	0.821151	**1.913026**	1.137777	**0.681923**	**0.931292**
2020	**1.447402315**	0.223895	0.950904	0.293571	0.654559	0.841031	**0.707615**	0.915509	**1.351906**	**1.196505**

First-person pronouns, contractions, stranded prepositions, and split infinitives have had the most significant developments. First-person pronouns have gradually increased from 0.65 in 1980 to 0.89 instances in 2000 before reaching their all-time high of 1.44 in 2020. Contractions have increased from 0.37 in 1980, to 1.91 instances in 2000, before decreasing again to 0.70 in 2020. Stranded prepositions have, by contrast, gradually decreased; from 0.96 in 1980, which declined to 0.68 in 2000, before increasing to 1.35 in 2020. Split infinitives have slightly increased, starting from 0.87 in 1980 to 0.93 in 2000, before increasing to 1.19 in 2020. Other informal traits have changed in various ways, yet their frequencies are not as significant as the previously mentioned frequencies.

Second person pronouns have increased from 0.96 in 1980 to 1.81 in 2000, before drastically dropping to 0.22 in 2020. Direct question decreased from 1.33 in 1980 to 0.82 in 2000, before slightly increasing to 0.84 in 2020. By contrast, sentence-initial conjunctions

have continued to decrease, starting from 1.85 in 1980 to 0.85 in 2000, before dropping to 0.29 in 2020. Similarly, exclamations decreased from 1.63 in 1980 to 0.70 in 2000, before decreasing again to 0.65 in 2020. Listing expression and unattended anaphoric pronouns have not faced significant changes. Listing expressions have increased from 0.81 in 1980 to 1.23 in 2000, before decreasing to 0.95 in 2020. Similarly, unattended anaphoric pronouns have increased from 0.94 in 1980 to 1.13 in 2000, before dropping to 0.91 in 2020.

Certain informal traits have generally but not exclusively, decreased, and it is exclusively owing to linguistics. Thus, the fourth research question has been answered and the dominant traits in economics have been highlighted. The fifth research question is also answered as unattended anaphoric pronouns and split infinitives have not changed significantly throughout four decades.

3.11. Analysis and comparison based on individual traits in three distinct periods

This section will deal with the frequency of individual informal traits in each period. This will demonstrate which informal trait was dominant in each discipline in a single period and will provide an overview of informal characteristics in each period in each genre. This will provide an overview of the informal traits among different disciplines in one time period, and the frequencies of individual informal elements can be compared.

3.11.1. 1980

Informal traits in journals that were published in 1980 are represented in table 9 and will be scrutinized in this section.

Table 9: Frequencies of informal elements in 1980

	1st pers. pron.	2nd pers. pron./ det.	Listing expressions	Initial conj.	Exclamations	Direct questions	Contractions	Unatt. anaph. pron.	Stranded prep.	Split infinitives
MATHEMATICS	22.821,96	0	0	959,07	10,31	61,87	0	1.588,15	41,25	51,56
BIOLOGY	1.485,9	0	0	33,70	6,75	182,36	6,75	1.046,89	6,75	60,78
LINGUISTIC	6.270,20	128,58	91,84	1.175,66	18,36	698,05	6,12	1.096,06	91,84	79,60
ECONOMICS	3.819,17	101	25,25	965,84	44,18	580,76	6,31	1.546,61	63,12	107,31

Table 9 presents the frequencies of each informal trait in 1980. First-person pronouns and unattended anaphoric pronouns were mostly found in mathematics journals. Contractions were found almost equally in all journals, excluding mathematics. Moreover, half of all traits were most commonly present in linguistics journals, these include 2nd person pronouns, listing expressions, sentence-initial conjunctions, direct questions, as well as stranded prepositions. Lastly, exclamations and split infinitives were most frequently used in economics journals. Moreover, no 2nd person pronouns and listing expressions were found in mathematics and biology, and no contractions were found in mathematics. Finally, although linguistics does not entail more 1st person pronouns than mathematics, it is often leading in terms of the frequency of diverse informal traits in comparison to other disciplines. Thus, 2nd person pronouns, listing expressions, sentence-initial conjunctions, direct questions, and stranded prepositions are most common in linguistics.

3.11.2. 2000

This section shows the most prominent informal traits per discipline in 2000. That will reveal which informal trait was dominant in which discipline in 2000.

Table 10: Frequencies of informal elements in 2000

	1st pers. pron.	2nd pers. pron./ det.	Listing expressions	Initial conj.	Exclamations	Direct questions	Contractions	Unatt. anaph. pron.	Stranded prep.	Split infinitives
MATHEMETICS	22.327,65	40,12	40,12	551,67	10,03	80,24	10,03	1.805,47	60,18	40,12
BIOLOGY	3.627,29	6,74	0	101,13	6,74	289,91	0	633,76	0	94,39
LINGUISTICS	6.808,05	29,24	64,33	859,77	23,39	567,33	81,88	1.672,76	146,22	105,27
ECONOMICS	5.181,57	191	38,19	445,59	19,1	356,47	31,82	1.858,74	44,55	114,58

As table 10 shows, different informal elements were predominant in various disciplines, yet neither trait was most frequently found in biology. First-person pronouns were, yet again, most commonly found in mathematics articles. All remaining 9 informal traits are distributed between linguistics and economics. Listing expressions, sentence-initial

conjunctions, exclamations, direct questions, contractions, and stranded prepositions were all most frequently found in linguistics. Finally, 2nd person pronouns, unattended anaphoric pronouns, as well as split infinitives were dominant in economics. Furthermore, listing expressions, contractions, and listing expressions did not occur in biology. Similar to the previous section, linguistics has again the highest frequency of certain informal traits: listing expressions, sentence-initial conjunctions, exclamations, direct questions, contractions, and stranded prepositions.

3.11.3. 2020

Finally, the frequencies of each informal trait in 2020 will be presented. This will provide a rather recent overview of the informal traits in all disciplines.

Table 11: Frequencies of informal elements in 2020

	1st pers. pron.	2nd pers. pron./ det.	Listing expressions	Initial conj.	Exclamations	Direct questions	Contractions	Unatt. anaph. pron.	Stranded prep.	Split infinitives
MATHEMATICS	26.066,21	9,27	18,56	204,14	18,56	46,39	74,23	1.586,8	9,27	55,67
BIOLOGY	5.992,48	0	18,62	74,51	0	49,67	0	757,59	0	242,18
LINGUISTICS	4.771,56	48,87	140,51	409,34	6,10	513,2	48,87	1.704,56	85,53	232,16
ECONOMICS	8.390,9	23,55	29,44	153,1	17,66	365,1	11,77	1.495,63	88,32	147,21

As Table 11 demonstrates, different informal elements were predominant in various disciplines, yet neither trait was most frequently found in biology. First-person pronouns were, yet again, most commonly found in mathematics articles. All remaining 9 informal traits are distributed between linguistics and economics. Finally, most informal traits were most frequently found in linguistics, and these include listing expressions, sentence-initial conjunctions, exclamations, direct questions, contractions, and stranded prepositions. Finally, 2nd person pronouns, unattended anaphoric pronouns, as well as split infinitives were dominant in economics. Biology was free from certain informal traits, these include

2nd person pronouns, exclamations, and contractions. Based on this, no set of informal traits has been dominant in different disciplines throughout the years, but a variety of informal features were dominated by each genre in each period.

3.12. Analysis based on journals

This section will provide an overview of the total number of informal elements in each journal from all four disciplines. Although all journals are peer-reviewed, the variety of authors potentially influences the distribution of informal traits. This section will answer the last research question: namely, whether any correlation can be identified between the various journals and, for example, the country that is responsible for editing and/or publishing a specific journal. Moreover, this would also provide information on informality among native speakers, from US and UK in this case, and native and non-native speakers, from Germany and Netherlands. This section will attempt to investigate whether the publishing country influences the number of informal traits, as the 16 journals are edited or published by different countries. Namely, *the Journal of Experimental Biology*, *Cambridge Journal of Economics*, *Journal of Linguistics*, and *Annual Review of Applied Linguistics* are published/edited by editors from the United Kingdom. Similarly, *The Quarterly Review of Biology*, *The Biological Bulletin*, *The American Naturalist*, *Advances in Mathematics*, *Review of World Economics* and *Language* originate from the United States of America. Likewise, *Acta Mathematica*, *Manuscripta Mathematica*, *Mathematical Notes*, *Journal of Economics (J* Econ), and *Theoretical Linguistics* stem from Germany. The remaining journal, *De Economist* is published by a publisher from the Netherlands.

The first discipline is mathematics. For this section of the corpus, articles from four journals were extracted: *Acta Mathematica*, *Manuscripta Mathematica*, *Advances in Mathematics*, and *Mathematical Notes*.

Figure 10 shows the total of informal traits in the articles that were extracted from these journals.

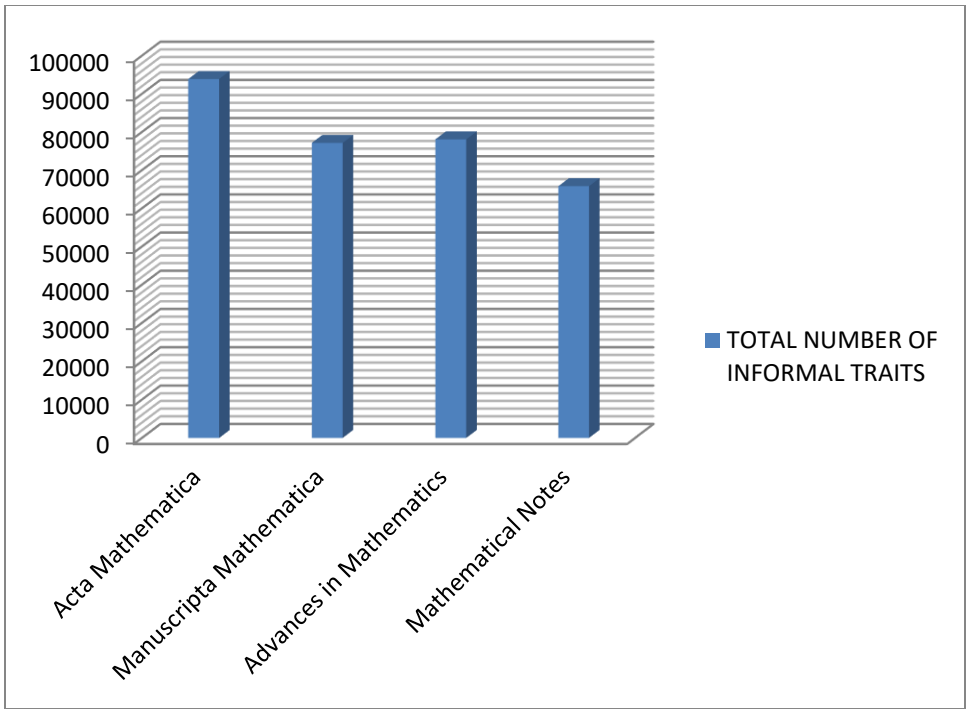

Figure 10: Informal traits in mathematic journals

Out of all journals, *Acta Mathematica* had the highest number of informal traits: 93.885. Advances in Mathematics and Manuscripta Mathematica follow, with 78.077 and 77.147 informal traits respectively. Finally, 65.821 informal traits were found in the journal *Mathematical Notes*.

The second discipline, biology, entails articles from *Journal of Experimental Biology*, *The American Naturalist*, *The Biological Bulletin* as well as *The Quarterly Review of Biology*.
All informal traits in the articles that were extracted from the aforementioned journals can be found in figure 11.

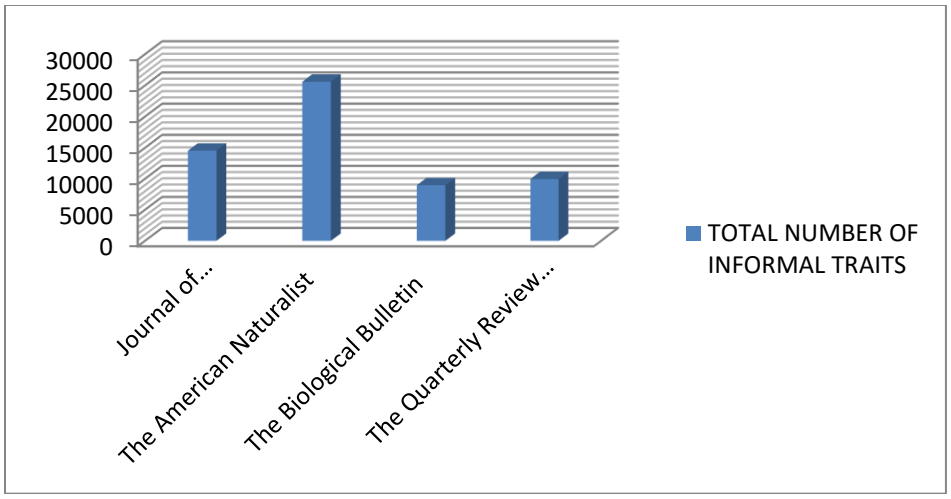

Figure 11: Informal traits in biology journals

As it is clear from figure 11, *The American Naturalist* is the biology journal that entails the largest number of informal traits namely 25.600. The second in line is the *Journal of Experimental Biology* with 14.532 informal traits. Finally, *The Quarterly Review of Biology* entails 9.959 informal traits and *The Biological Bulletin* entails 8.970 informal elements.

Next is linguistics. This discipline is represented by *Annual Review of Applied Linguistics*, *Journal of Linguistics*, *Theoretical Linguistics*, and *Language*. Similar to the previously mentioned disciplines, the frequency of informal traits in individual journals varies, and these are represented in figure 12.

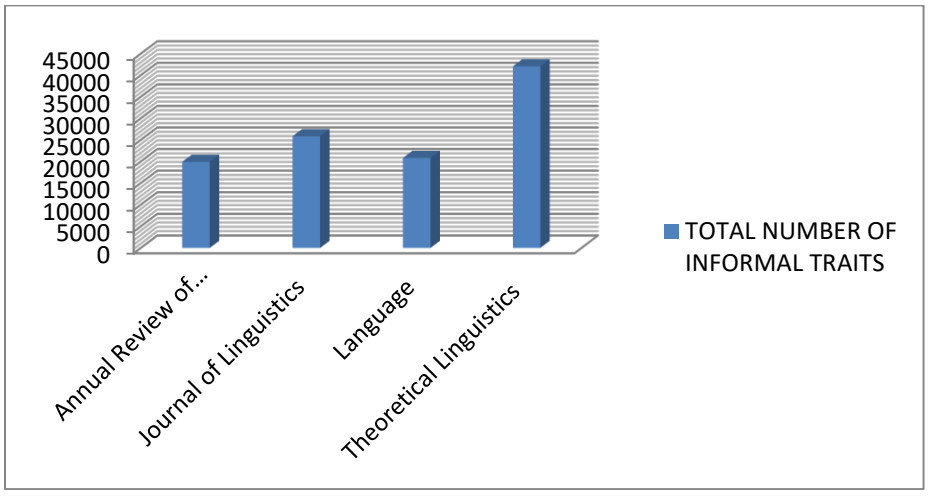

Figure12: Informal traits in linguistics journals

In linguistics, *Theoretical Linguistics* is the most prominent journal in terms of informal traits. It entails 42.076 informal traits in total. Other journals, such as the *Journal of Linguistics* or *Language*, for example, have fewer informal traits: namely, 25.840 and 20.823 respectively.

Similarly, the *Annual Review of Applied Linguistics* entails only 19.908 informal traits.

Lastly, economics articles were extracted from *De Economist*, *Cambridge Journal of Economics*, *Journal of Economics*, and *Review of World Economics*. This discipline has potentially the least amount of discrepancies between informal traits in individual journals. Figure 13 confirms that.

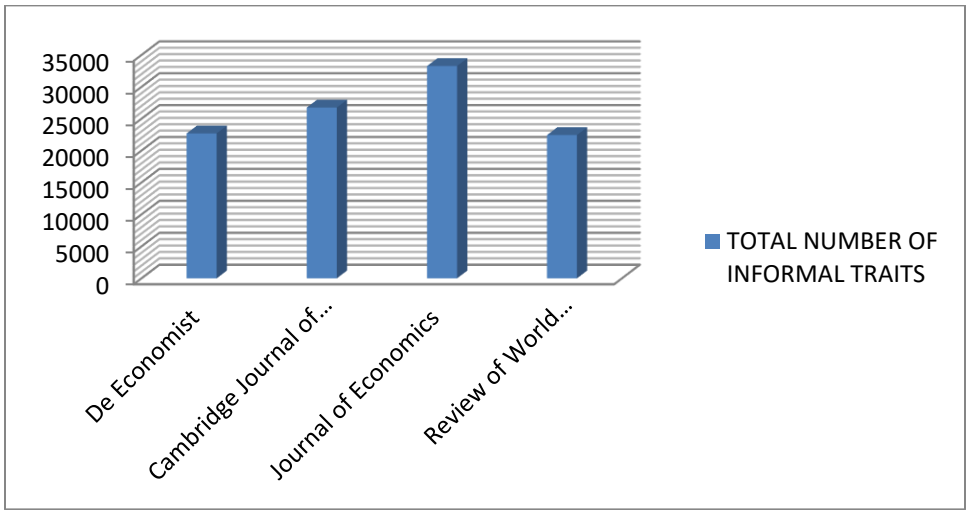

Figure 13: Informal traits in economics journals

The journal's highest number of informal traits, which is 33.269 in this case, is the *Journal of Economics*. Subsequent is *Cambridge Journal of Economics* with 26.755 informal elements. Finally, De Economist, with 22.683, and Review of World Economics, with 22437 informal traits follow. Informality is not equally distributed in one discipline. Although the frequencies of informal traits were also scrutinized based on the country of publishing or editing, no significant points could be concluded, in a sense that journals from a certain country are more or less informal. Thus, the sixth research question has also been answered.

3.13. A qualitative look at one specific trait

This section is concerned about 1st person pronouns in mathematics in particular. First-person pronouns occur more in mathematics than in any other discipline; there is, however, a specific explanation for the use of pronouns in mathematics. Rowland (2000: 215) points out that mathematicians do not use pronouns to express their emotional state or stance, but rather to express their thoughts and reasoning. On closer inspection of the data, the instances of personal pronouns in mathematics include personal pronouns in phrases such as *I checked* or *I calculated*. Similarly, 2nd person pronouns are also used in phrases such as *we prove that* or *we set the value*. This means that the mathematics journals in this corpus are not necessarily informal, as first-person pronouns and second-person pronouns are mostly used to describe the process or explain the steps which the authors have made. This means that although 1st and second-person pronouns are common in mathematics, they are not used to express any personal opinions of the authors. Moreover, Moschkovich (2008: 557) found that in one mathematics classroom, the phrase *I went by* has multiple meanings. This automatically increases the instances of this item, as it can be used in multiple instances or to convey multiple ideas. Moreover, multiple, unprecise phrases with changing meanings did not negatively influence the exchange of ideas. (Moschkovich 2008:578) Therefore, using ambiguous phrases with personal pronouns is crucial for mathematics. This might in turn also influence the writing style to a great extent. Moreover, it was calculated in this study that articles from mathematics entailed only 55% usable data, which is the lowest percentage in comparison to other disciplines. Moreover, mathematicians are potentially attempting to explicitly express their ideas, rather than focusing on academic or formal language.

However, in the case of mathematics, numerous curricula extensively focus on the advancement of a mathematical language: this includes students expressing their opinions, which are potentially informal, in a formal language. (Barwell 2013:73) This is because mathematics includes the exchanging of ideas, which might not always be formal. This is not surprising because the focus is on the idea, topic, or solution to a mathematical problem, and not on the language, which is in that case almost irrelevant. However, Barewell (2013:

79) found in his study that the implementation of the focus on formality was successful, however, it is not a straightforward development from informal to formal and that students use both formal and informal language. While the formal was added the informal were also kept. A similar can be said for unattended anaphoric pronouns.

3.14. Statistical significance

This section will contain a test for normality of data distribution, a statistical significance test, as well as an analysis of variance. Although normalized frequencies have been presented, statistical tests will provide answers to some more complex questions. The mentioned tests will be performed to find out whether there is a statistically significant correlation between the year and frequency of informal traits and to find out whether the publishing country or editors influence the level of informality. Thus, all data has been entered into SPSS, the tests have been performed, and finally, all figures were extracted from the output file.

3.14.1. Statistical significance correlation between period and informality

Before performing any test, it is necessary to see whether the data is normally distributed; this will determine which statistical significance correlation the test is appropriate for this specific data. A normal distribution entails data in which the average scores are more frequent than lower scores (e.g. a bell curve). Figure 14 demonstrates the actual data distribution of informal traits. The red line shows a bell curve; if the bars were aligned under the bell curve, that would represent normal data distribution.

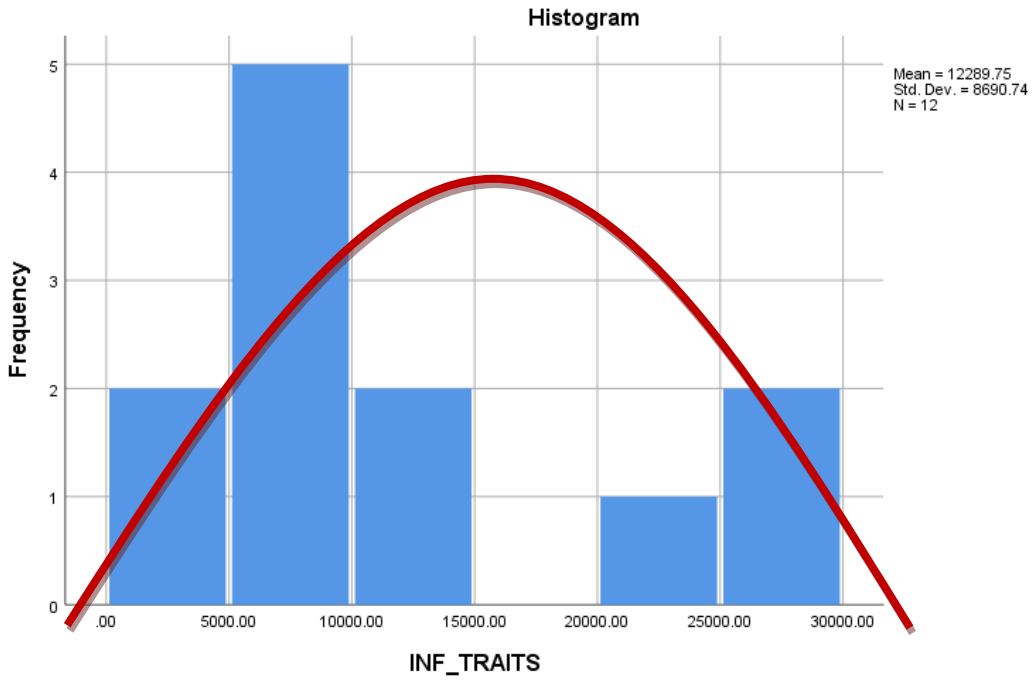

Figure 14: Size of each data set (by informal traits)

Figure 14 demonstrates the distribution of informal traits in 12 data sets: 4 genres and 3 periods. This means that there are, for example, two data sets that have a total of 1 up to 5.000 informal traits, there are 5 data sets that entail between 5.001 and 10.000 informal traits. There are, again, 2 data sets that entail 10.001 up to 15.000 informal traits and there is one data set that entails anywhere between 20.001 and 25.000 informal traits. Lastly, two data sets entail between 25.001 and 30.000 informal traits, however, no data set entails more than 15.001 yet less than 20.000 informal traits. It is clear that the data is not normally distributed as few data sets have an average number of informal traits, and more data sets have either lower or higher numbers of informal traits than average.

There are various tests for normality, however, the main statistical test for testing normality is the Kolmogorov-Smirnov test, as presented in Table 12.

Table 12: Kolmgorov-Smirnov test of normality

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
INF_TRAITS	.323	12	.001	.796	12	.008

a. Lilliefors Significance Correction

In table 12, the informal traits from four disciplines and three-time periods are the dependent variables. Because $p \leq 0.05$, it is safe to assume that the variable is not normally distributed. Thus, based on the performed tests in SPSS, the data is not normally distributed.

It has been established that informality has increased over the years, the next step is to test whether there is a correlation between the period and the number of informal traits. Although the data is not normally distributed, a parametric test has been selected to analyze it. This is because Pearson's correlation test, although typically used for data with normal distribution, can also be used with data that is not normally distributed. Moreover, the alternative, which is Spearman's correlation test, is used for variables that have a rank, whereas Pearson's correlation is used with numerical variables. Table 13 is the output from a Pearson's correlation test that has been performed in SPSS.

Table 13: Pearson's correlation

Correlations

		YEAR	TOTAL_INF_TRAIT
YEAR	Pearson Correlation	1	.861
	Sig. (2-tailed)		.340
	N	3	3
TOTAL_INF_TRAIT	Pearson Correlation	.861	1
	Sig. (2-tailed)	.340	
	N	3	3

In table 13, three distinct periods and the total of informal traits from each period were set as variables, and the correlation between these two was tested. For the Pearson's correlation test, a correlation of 1.0 indicates a perfect positive correlation. Here, a 0.861 Pearson correlation suggests a fairly strong positive correlation. In other words, there is a strong correlation between the year and the frequency of informal traits. Consequently, as one variable increases or shows signs of progress, the other variable increases as well.

3.14.2. Relationship between publishing country or editor and journal

In contrast to the previous section, for this analysis, different variables have been used: namely, the total number of informal traits per journal from table 3 has been entered into SPSS and analyzed together concerning the country of the respective journal. For this, a one-way analysis of variance has been used to compare the aforementioned two samples. This procedure is also called One-way ANOVA, and it is based on comparing whether different means are significantly different from one dependent variable. Thus, table 14 presents the means from different countries.

Table 14: Means of informal traits among groups of journals from different countries

Descriptive Statistics
Dependent Variable: Inf.Traits_journal

COUNTRY	Mean	Std. Deviation	N
US	27644.3333	25623.35581	6
UK	21758.7500	5694.15802	4
Germany	62439.6000	24910.61406	5
Netherlands	22683.0000	.	1
Total	36736.3750	26778.46330	16

Table 14 entails the mean of all informal traits from all journals from one country and it is based on normalized frequency. The fixed factor here is country, and the dependent variable is informal traits. The means of different journals differ to various degrees. It can be concluded that journals that have been published by German publishers or edited by German editors are often the most informal ones: these journals have on average 62.439 informal traits. After that, the second most informal journals were published or edited by editors from the US, and have on average 27.644 informal traits. Although there was only one journal published by a Dutch publisher, it is in third place with regards to informality and has on average 22.683 informal traits. Lastly, the journals which had the lowest average number of informal traits were published by UK publishers and had on average 21.758 informal traits. It is important to mention that the publishing country of the journals does

not necessarily correspond to the authors. In other words, if a certain journal is published by a Dutch publishing house, that does not mean that the articles in that journal were also written by Dutch authors. It is important to mention that the discipline can be seen as a confounding variable. There are four countries and four disciplines; however, the number of journals from a specific discipline is not evenly distributed among the four countries. In other words, it is not the case that one journal from each discipline was published by one country. In some cases, a few journals from the same discipline were published by one country. Similarly, there is only one journal from the Netherlands. Yet, table 14 was obtained from a formula in SPSS and provides an answer to the informality of the publishers in each country and takes into consideration that sometimes there are more or fewer journals from one country.

Moreover, a line graph, Figure 15, will demonstrate more clearly the differences between journals published by various countries.

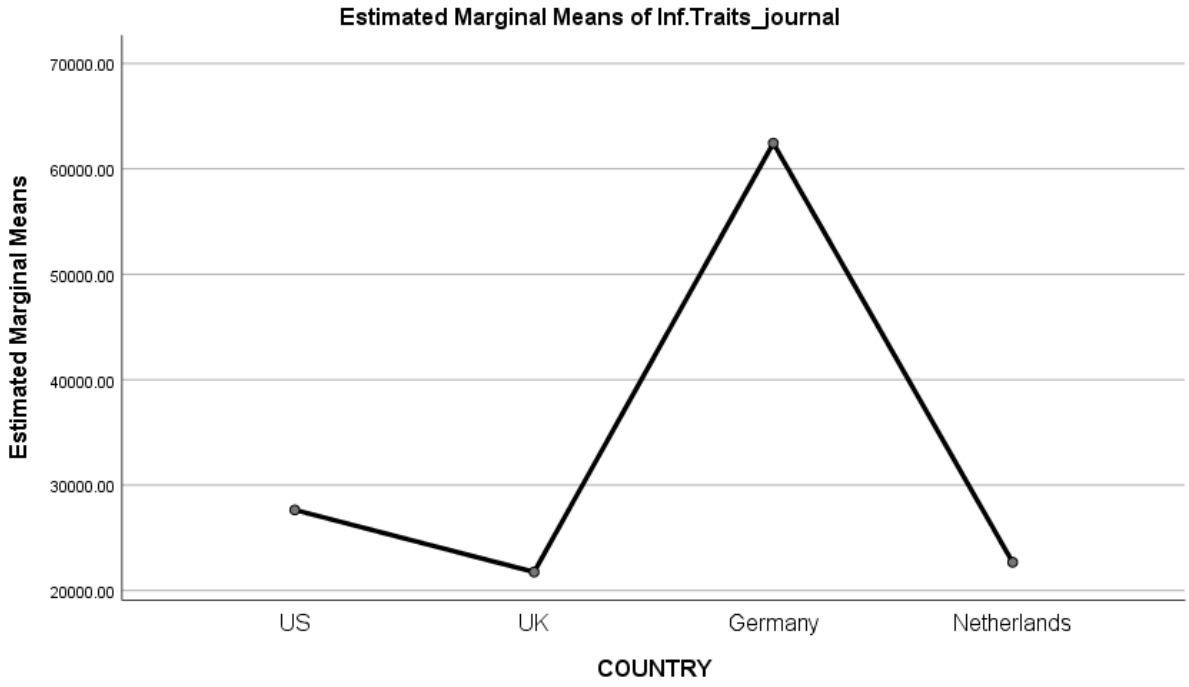

Figure 15: Means of informal traits in journals from different countries

Informal traits are extensively high in German journals, followed by journals from the US, Netherlands, and UK respectively. Thus, there is a correlation between the country of publishing and the level of informality in academic journals.

4. Discussion

Firstly, an analysis and comparison of all informal traits in each of the four disciplines has revealed that the level of informality varies between disciplines. This confirms the findings of the original study which concludes that the changes and developments vary among disciplines. However, as the original study is not clear about its methods, it is difficult to derive any further conclusions. Mathematics, on its own, entails almost as many informal traits as the remaining three disciplines together. This means that mathematics is the most informal discipline out of the four disciplines which were included in this analysis. After mathematics biology is leading in terms of informal traits and is followed by linguistics and economics respectively. The fact that mathematics is the most informal discipline is not surprising; this is because mathematics is complex, and entails a plethora of formulas, predictions, and values that can potentially be difficult to integrate into a text. This study is important to provide an answer to why mathematics is informal as well as to see in which ways different disciplines differ in formality. Moreover, certain traits are characteristic of each discipline. Although biology is the 2nd most informal discipline, neither informal trait is dominant in biology, in comparison to other disciplines. Linguistics, by contrast, has the widest variety of dominant informal traits and these include listing expressions, sentence-initial conjunctions, direct questions, contractions, stranded prepositions as well as split infinitives. All of these were more prominent in linguistics that in any other genre. Lastly, although economics has overall the lowest number of informal traits, it is characteristic of its use of 2nd person pronouns and exclamations.

Analysis and comparison of all informal traits in each period in one discipline. The developments that have taken place within different genres are not unidirectional. Mathematics, biology, and economics have become more informal over the years. In contrast to linguistics, which has become more formal. The most significant development

has been recorded in biology, where informal traits have continued to increase sharply, finally increasing by 152.20%. Relating to the first hypothesis, namely that informality in natural sciences has increased, has been confirmed. It is still necessary to explain the details of the development of this trend. Hyland & Jiang (2017:44) confirm this in their study and state that biology has overall increased in informality "rising by a substantial 24.8%." The most prominent and only informal trait that has continued to increase from 1980 to 2020 are first-person pronouns. With regards to biology, Hyland & Jiang (2017:45) highlight an almost equal trend, stating that in their study informal traits in biology have increased by 213%. In addition, unattended anaphoric pronouns slightly decreased over time in biology: in 40 years, anaphoric pronouns have decreased by 27.60%, remaining the 2nd most dominant trait in this subject. Finally, Hyland & Jiang mention that from 1965 to 2015 split infinitives steadily decreased (2017:46). However, in the last period of this study, split infinitives were the 3rd most dominant informal trait in biology, and data proves that they steadily increased by almost 300% throughout 40 years. Hyland & Jiang 2017:42 provide a rather interesting explanation for informality: it is an approach with nearness to the reader, an openness to a discussion of statements, and an affirmative stance toward individuality. With the increase of first-person pronouns, unattended anaphoric pronouns, and split infinitives in biology, it is not surprising that biology journals have become rather relaxed in style.

It is significant to identify informal traits in academic writing in individual disciplines. If research points out that a certain discipline is becoming more informal, the authors might attempt to control the issue and improve the situation. Furthermore, in this study, informality in linguistics has decreased by 17,56%. Similarly, Hyland & Jiang's study (2017:44) also confirms that Linguistics has become more formal throughout the years by approximately 10.3%. First-person pronouns, for example, have decreased in linguistics. Hyland & Jiang (2017:45) also confirm that 1st person pronouns have declined in linguistics. They also provide the reason behind this development. They also explain that this is "perhaps as a result of the self-consciousness of language-sensitive writers aware of the attention this draws to the individual and the strong claims it makes for agency in research"

(Hyland & Jiang 2017:45) Unattended anaphoric pronouns and split infinitives have increased in linguistics by 55.5 and 191.65 Although informality has increased throughout the years, it is plausible that assume that informal features developed independently. Changes have also taken place within distinct disciplines; informal traits developed independently of one another.

The reasons why individual traits have had different developments may be individual. Moreover, it is only possible to assume why a certain informal trait has developed the way it has. Overall, four informal traits have undergone major changes: 1st person pronouns, sentence-initial conjunctions, unattended anaphoric pronouns, and split infinitives. This is based on their frequency in a specific discipline and on the degree of change. First-person pronouns have increased by 103,25% overall. Hyland & Jiang noticed also noticed an increase, in their case by 45% overall (2017: 45) There are various reasons for this development. The use of first-person pronouns enables authors to create a piece of writing which influences the opinion of readers as well as wins their attention. (Hyland & Jiang 2017:45) This also reinforces the findings of Basal & Bada's study in which it was proved that native and non-native writers use personal pronouns often in their writing. (2012: 1784) This contrasts with what has been initially mentioned, as well as the fact that personal pronouns are rather personal.

Basal & Bada concluded that in academic writing the mentioning of authors is not avoided and that this is an ongoing development. (2012:1779) Moreover, the use of personal pronouns creates a connection between writers and readers, and based on function, their frequency might vary. (Basal & Bada 2012:1781) Basal & Bada (2012:1781) argue that writers feel obliged to support their ideas when facing readers through their writing; the best way of achieving this is by mentioning themselves directly in a text. Basal & Bada also found that “in hard sciences, researchers try to diminish their contribution in terms of their visibility with the use of personal pronouns in their texts. (2012:1783) Finally, it is mostly due to first-person pronouns that academic language has become more personal. Sentence initial conjunctions have decreased by 26,56% on average: this includes *And* and *But*. Due to sentence-initial conjunctions, the aforementioned uses of sentence-initial conjunctions it

has become less argumentative and writers have potentially fewer lists. However, Jiang & Hyland noticed that initial conjunctions and/or conjunctive adverbs such as however, so and indeed have increased by 50% from 1980 to 2015. (Hyland & Jiang 2017:47) This also potentially suggests a more argumentative, or repetitive stance in texts. Unattended anaphoric pronouns have increased on average by 6%. Unattended anaphoric pronouns increase might suggest that authors go back to mentioned points or subjects. Hyland & Jiang (2017:46) report in their study that unattended anaphoric pronouns "declined dramatically across all disciplines". This is true as unattended anaphoric pronouns have slightly declined from 1980 to 2000. However, this study examines a slightly different period, the overall increase from 1980 to 2020. Lastly, split infinitives have increased on average by 134,06%. Moreover, Stageberg & Oaks's (2000:333) speculation that split infinitives will increase in frequency over time has been confirmed. This demonstrates in which way academic language has changed. An increase in split infinitives means that authors want their writing to appear more fluent and proficient. In contrast to the past, editors from all disciplines seem to tolerate contractions and split infinitives more recently. (Hyland & Jiang 2017:47) Another interesting development is that various contractions have increased in social sciences, namely in linguistics and economics. By contrast, contractions decreased in natural sciences, in this case in mathematics and biology. Similarly, 1st person pronouns also increased in most disciplines, however, linguistics was not one of them. Unattended anaphoric pronouns have increased in linguistics, yet overall decreased in all disciplines. Analysis and comparison based on individual traits in three distinct periods. The frequency of individual informal traits in each period differs. However, certain traits were prominent in mathematics and linguistics throughout all three time periods.

In mathematics, for example, first-person pronouns were significantly dominant in all three time periods, always surpassing the total number of first-person pronouns from the remaining three disciplines. Hyland & Jiang (2017:48) also point out that in the overall increase of informality, personal pronouns are one of the main contributors.

Basal & Bada (2012:1778) point out the salience of personal pronouns and add that different disciplines require different amounts of personal pronouns. Aside from shifting

from the language that is characteristic of academic writing, the changes also “allow more personal comment, narration, and stylistic variation” (Chang & Swales 1999: 145). This may suggest that authors are attempting to establish a more direct relationship with their audience by, for example, using personal pronouns and directly including their opinion in their works. Moreover, it is impossible to investigate the types of writing of each department, as well as to what extent they are familiar with writing conventions. The second most significant genre in this case is linguistics. Namely, listing expressions, sentence-initial conjunctions, and direct questions were dominant in linguistics throughout all three time periods. The developments of these informal traits overlap with the findings of Hyland & Jiang’s study (2017: 45). Although listing expressions decreased in Hyland & Jiang’s study (2017:45), as well from 1980 to 2000, they finally increase in linguistics and this is salient to modern authors. They usually serve In the case of sentence-initial conjunctions, although writing manuals overall suggest authors not to use these informal traits, it appears that they are becoming normal in academic writing in English. (2017:46)

This is because direct questions are also tolerated to a greater extent than they were initially Hyland & Jiang (2017:45) Although these traits are not necessarily informal or negative, just classified as informal, it appears that they are useful for writers. In the case of natural sciences, more precisely mathematics, first-person pronouns are dominant and crucial. In social sciences, in this case, linguistics, listing expressions, sentence-initial conjunctions, and direct questions were dominant and salient for this discipline and various authors. It was proved that informality increases with the years.

An analysis based on journals showed that countries from certain journals are more informal than others. Relationship between publishing country or editor and journal. Although german journals were the most informal, it cannot be stated that the native language has an influence on informality in this case. This is because the 2nd most informal journal was from a country with native speakers, which is the US. This is again followed by non-native writers from the Netherlands, and by native writers from the UK.

Based on this, it cannot be stated that overall informal traits are more common in non-native writing. However, if Germany is compared to the US, and the Netherlands is

compared to the UK, in that case, informal authors are more common in non-native academic writing. And, in that case, the findings overlap with Alipour & Nooreddinmoosaa's (2018:367) study which proves that non-native writers use informal traits more extensively. This is potential because their first obstacle is language and formality is already a 2nd level of knowing a language. Finally, Zhao's 2017 study scrutinized the use of conjunctions among native and non-native authors. She regarded conjunctions as markers of linguistic proficiency. However, she found that skills which were gained throughout learning about a discipline are more salient than whether English is the authors' native or non-native language.

Although this study takes into consideration the representativeness by taking an equal number of natural and social sciences, a total of four genres cannot be representative of every existing discipline in academia. Moreover, it is difficult to generalize that every existing discipline from either the social or natural sciences shares the developments which the respective disciplines from this study have. If another study with the same disciplines and periods was conducted but with different journals, it would be possible to conclude how telling the numbers are. This is due to the reason that this random selection of authors might prove one development, yet another group may generate different results.

Based on these findings, the issue between informality as a whole and the individual traits arises and it is salient to address it. Furthermore, charting the development of formality is not meaningful in this case, as academic articles are considered to be academic and formal. Moreover, charting the development of the individual traits is salient because they all contribute to informality, and their increase and decrease also show exactly which trait or area is changing and to what degree. Moreover, if certain traits are functional and beneficial, their increase may not be detrimental. It is important to mention, again, that this paper has considered prescriptivist approaches. However, if half of the sentences in an academic paper start or contain the personal pronoun *I* that would influence the quality of an article in a negative way.

5. Conclusion

This study has proved that the trend which was proposed by Hyland & Jiang (2017:40) that there is a shift in written texts towards an involved writing style is present in academic language and this section summarizes all main findings.

5.1. Summary

The first finding in this study is that informality is not equally distributed among the four selected genres. Moreover, the dominant informal traits are different in each discipline, meaning that each discipline is informal to a different degree as well as in different ways. Perhaps one of the most significant findings is the number of informal elements in mathematics, which is close to the total of informal traits in the remaining three disciplines. Thus, the leading discipline when it comes to informal traits is mathematics, and it is dominated by personal pronouns and unattended anaphoric pronouns. Biology follows after mathematics in the frequency of informal traits, however, neither informal trait has more instances in biology than in any other discipline. Linguistics is the 3rd most informal discipline and it has a variety of dominant informal traits: listing expressions, sentence-initial conjunctions, direct questions, contractions, stranded prepositions as well as split infinitives. Lastly, economics has the least amount of informal traits and exclamations and 2nd person pronouns are prominent in this discipline. Informality in natural sciences has increased. The most significant increase in informality has been recorded in biology: namely, informal traits increased overall by 152.20%. This was mainly caused by first-person pronouns which continued to increase throughout the 40 years, as well as unattended anaphoric and split infinitives. In linguistics, which is a social science, informality has decreased by 17,56%. Moreover, this is the only discipline that has decreased informality. In linguistics, first-person pronouns decreased. Yet, similarly to biology, the use of unattended anaphoric pronouns and split infinitives has increased over the years. Four out of ten informal traits have undergone significant changes. The most significant of them are first-person pronouns, which increased by 103.25% overall. There are different reasons why authors might rely on the use of first-person pronouns, however, their progressive increase

suggests that they are an almost irreplaceable choice for modern authors. By contrast, sentence-initial conjunctions *And* and *But* have decreased by 26,56 overall. This is significant because these two conjunctions are usually used at the beginning of sentences which are used for arguments and lists. Thus, based on the aforementioned functions of *And* and *But*, articles have potentially become less argumentative and writers have potentially fewer lists. When it comes to unattended anaphoric pronouns, they have increased: namely by 6%. And as unattended anaphoric pronouns are used to point back to previous arguments, this also suggests that authors restate previously mentioned points slightly more often than before. Split infinitives have increased dramatically, namely by 134,06. Split infinitives are used for numerous reasons and they add to the fluency of a text, and that is the reason why authors might use split infinitives more frequently than before. These changes show the exact way in which academic language has changed in general. On a more specific note, it can be pointed out that contractions have increased in social sciences and decreased in natural sciences. Linguistics also appears to have developed differently. Namely, 1st person pronouns decreased in linguistics but increased in other disciplines, especially in mathematics, first-person pronouns are dominant and crucial By contrast, unattended anaphoric pronouns have decreased in all disciplines but increased in linguistics. Linguistics was dominant in the frequency of listing expressions, sentence-initial conjunctions, and direct questions.

Finally, it is not possible to conclude that non-native speakers use more informal traits than native speakers. Moreover, it was proved that informality increases, as the years progress, however without a unidirectional development. This study tracked exactly in which way at which point and where the main issue in academic journals was. In addition, the informality in journals from different publishing countries was also analyzed. Therefore, all aims and objectives were achieved.

5.2. Limitations

Compiling a corpus of articles from academic journals and recording the number of each informal trait is an excellent method to measure informality. However, two disciplines from

social sciences and two disciplines from natural sciences are still not representatives, as there are potentially around 50 different disciplines overall. In other words, how this study was conducted is optimal, yet this might still not be finally representative, and that is one of its limitations. Thus, in the future, a study of at least 6 and up to 10 different disciplines might be more representative. Moreover, the data collection process without a doubt was the most difficult part of this research: the problems were caused by scarcity, inadequate, non-available data. In other words, if the access to journals was not limited, the data collection would have been somewhat facilitated. However, after arduous searching and rejecting up to hundred journals, sixteen adequate journals with free access were found. Future studies might facilitate the process of data collection if they focus on periods after 1990, and simply focus on the changes for example from 1990 until 2020. This would show developments, and also the direction of the trend. Future studies might also use the same journals from this study, as they are freely accessible.

References

Adam Kilgarriff, Vít Baisa, Jan Bušta, Miloš Jakubíček, Vojtěch Kovář, Jan Michelfeit, Pavel Rychlý, Vít Suchomel. The Sketch Engine: ten years on. Lexicography, 1: 7-36, 2014. Available under: http://www.sketchengine.eu

Adejare Abonego, Roseline. 2021. "Preposition Pied Piping and Stranding in Academic and Popular Nigerian English Writing". *International Journal of English Linguistics* 11(4), 40-57.

Akhtar, Raheela; Riaz, Mehvish. 2019. "Formality in Academic Writing: Investigating Stylistic Competence of Undergraduate EFL Learners". *University of Wah Journal of Social Sciences* 2(1), 01-20.

Alaowffi, Nouf. Alharbi, Bader. 2021. "Preposition stranding under sluicing: Evidence from Hijazi Arabic". *Journal of Language and Linguistic Studies* 17(2), 941-957.

Alipour, Mohammad; Nooreddinmoosa, Mona. 2018. "Informality in applied linguistics research articles: comparing native and non-native writings". *Eurasian Journal of Applied Linguistics* 4(2), 349-373.

Anthony, L. (2022). AntConc (Version 4.0.11) [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software

Anthony, L. (2022). TagAnt (Version 2.4.0) [Computer Software]. Tokyo, Japan: Waseda University. Available from https://www.laurenceanthony.net/software

Barwell, Richard. 2013. "FORMAL AND INFORMAL LANGUAGE IN MATHEMATICS CLASSROOM INTERACTION: A DIALOGIC PERSPECTIVE."

Basal, Ahmet; Bada, Erdogan. 2012. "Use of first person pronouns: A corpus based study of journal articles". *Energy Education Science and Technology Part B: Social and Educational Studies* 4(3), 1777-1788.

Belal, Anika. 2014. "Influence of digital social media in writing and speaking of tertiary level student". MA thesis, BRAC University.

Bell, David. 2007. "Sentence Initial And and But in Academic Writing". *Journal of Pragmatics* 17(2), 183-202.

Bennett, Karen. 2009. "English academic style manuals: A survey". *Journal of English for Academic Purposes* 8, 43-54.

Bergh, Gunnar; Seppänen, Aino. 2000. "Preposition stranding with wh-relatives: a historical survey". *English Language and Linguistics* 4(2), 295-316.

Biber, Douglas; 문진미디어편집부; Johansson, Stig; Leech, Geoffrey; Conrad, Susan; Finegan, Edward; Bieber, Douglas(ed.). 1999. *Longman Grammar of Spoken and Written English*. Harlow: Pearson Education.

Bryson, Bill. 1990. *Mother tongue.* London: Penguin Books.

Chang, Yu-Ying; Swales, John Malcolm. 1999. "Informal elements in English academic writing: Threats or opportunities for advanced non-native speakers". In Candlin, Christopher N.; Hyland, Ken (eds.). *Writing: Texts, Processes and Practices.* London & New York: Longman, 145-167.

Coffin, Caroline Jane; Curry, Mary Jane; Goodman, Sharon; Hewings, Ann; Lillis, Theresa; Swann, Joan. 2003. *Teaching Academic Writing: A Toolkit for Higher Education.* London: Routledge.

Constantinou, Filio; Chambers, Lucy; Zanini, Nadir; Klir, Nicole. 2019. "A diachronic perspective on formality in students' writing: Empirical findings from the UK. *Language, Culture and Curriculum* 33(1); 1-18.

De Cock, Barbara. 2016. "Register, Genre and Referential Ambiguity of Personal Pronouns: A Cross-Linguistic Analysis". *Pragmatics* 26(3), 361-378.

Dimitriadis, Eva. 2007. "A preposition is something which you should never end a sentence with: A corpus-based study on preposition stranding". MA thesis, Växjö University.

Dubois, Sylvie. 1992. "Extension particles, etc.". *Language Variation and Change* 4, 179-203.

Fairclough, Norman. 2001. *Language and Power.* London: Longman.

Geertz, Clifford. 1988. *Words and Lives: The Anthropologist as Author*. Palo Alto, CA: Stanford University Press. Harley, Heidi. 2006. *English words: A linguistic introduction.* Oxford: Wiley-Blackwell.

Harwood, Nigel. 2005. "'Nowhere has anyone attempted ... In this article I aim to do just that' A corpus-based study of self-promotional I and we in academic writing across four disciplines". *Journal of Pragmatics* 37, 1207–1231.
Heylighen, Francis. 1999. "Advantages and limitations of formal expression". *Foundations of Science* 4 (1), 1-21.
Heylighen, Francis; Dewaele, Jean-Marc. 1999. "Formality of Language: definition, measurement and behavioral determinants". Internal Report, Center "Leo Apostel": Free University of Brussels.
Huddleston, Rodney; Pullum, Geoffrey K. 2002. *The Cambridge Grammar of the English Language.* Cambridge: Cambridge University Press.
Hundt, Marianne; Mair, Christian. 1999. "Agile" and "Uptight" Genres: The Corpus-based Approach to Language Change in Progress". *International Journal of Corpus Linguistics* 4(2), 221-242.
Hyland, Ken. 2002. "What do they mean? Questions in academic writing".Text, 22(4), 529-557.
Hyland, Ken; Jiang, Feng. 2017. "Is academic writing becoming more informal?". *English for Specific Purposes* 45, 40-51.
IBM Corp. Released 2019. IBM SPSS Statistics for Windows, Version 26.0. Armonk, NY: IBM Corp
INFORMALITY meaning in the Cambridge English Dictionary https://dictionary.cambridge.org/dictionary/english/informality (10. July 2022).
Kane, Thomas S. 2000. *The Oxford Essential Guide to Writing.* New York: Berkley Books.
Kirszner, Laurie; Mandell, Stephen R. 1987. *The Writer's sourcebook: strategies for reading and writing in the disciplines.* New York: Holt, Rinehart, and Winston.
Kjellmer, Goran. 1998. "On contraction in modern English". *Studia Neophilologica* 69(2), 155-186.
Koutsantoni, Dimitra. 2007. *Developing Academic Literacies Understanding Disciplinary Communities' Culture and Rhetoric.* Oxford: Peter Lang.
Kuo, Chih-Hua. 1999. "The Use of Personal Pronouns: Role Relationships in Scientific Journal Articles". *English for Specific Purposes* 18(2), 121–138.
Leech, Geoffrey; Deuchar, Margaret; Hoogenraad, Robert. 2006. *English grammar for today: a new introduction.* New York: Palgrave Macmillan.
Leedham, Maria Elizabeth. 2011. "A corpus-driven study of features of Chinese students' undergraduate writing in UK universities". Ph.D. thesis, The Open University.
Lester, James, D.; Lester, James D. Junior. 2015. *Writing Research Papers: A Complete Guide.* 15th edition. Pearson, Harlow. New York
Mair, Christian. 1998. "Corpora and the study of the major varieties of English: Issues and results.". In Lindquist, Hans; Klintborg, Staffan; Levin, Magnus & Estling, Maria(ed/s.). *The major varieties of English: Papers from MAVEN 97.* Växjö: Växjö University, 139-157.
Mauranen, Anna. 1992. Reference in Academic Rhetoric. A Contrastive Study of Finnish and English Writing. In A-C. Lindeberg, N. E. Enkvist, & K. Wikberg (Eds.), Nordic Research on Discourse (Nordtext Symposium 1990).
Melissourgou, Maria N.; Maruster, Laura. 2017. "Changing trends in research articles: investigating informality". Paper presented at 38th TABU Dag, Groningen, Netherlands.
Moschkovich, Judit N. 2008. "I Went by Twos, He Went by One": Multiple Interpretations of Inscriptions as Resources for Mathematical Discussions". *The Journal of the Learning Sciences* 17(4), 551-587.
Msuya, Erasmus Akiley. 2016. Exploring Author Visibility in Academic Writing in EFL Tanzanian Context: A Case Study of University of Dar es Salaam Postgraduate Dissertations and Theses. International Journal on Studies in English Language and Literature (IJSELL) 4(10), 55-61.
Pitarch, Ricardo Casan. 2016. "Case study on banks' webpages: The use of personal pronouns". *International Journal of Language Studies* 10(4), 37-58.
Praminatih, Gusti A.; Kwary, Deny A.; Ardaniah, Viqi. 2018. "Is EFL students' academic writing becoming more informal?". *Journal of World Languages* 5(2), 101–112.

Rowland, Tim. 2000. *The Pragmatics of Mathematics Education Vagueness and Mathematical Discourse.* London: Falmer Press.

Spencer, Carolyn, Arbon, Beverly, 1996. *Foundations of Writing: Developing Research and Academic Writing Skills.* National Textbook Co, Lincolnwood, IL.

Stageberg, Norman C; Oaks, Dallin D. 2000. *An introductory English grammar.* Fort Worth: Harcourt Brace College Pub.

Supakorn, Phoocharoensil. 2013. "The English split infinitive: A comparative study of learner corpora". *International Journal of Research Studies in Language Learning* 2(4), 21-32.

Swales, John Malcolm. 2005. "Attended and Unattended "this" in Academic Writing: A Long and Unfinished Story". *ESP Malaysia* 11, 1-15.

Swales, John. 1990. *Genre Analysis: English in academic and research settings.* Cambridge: Cambridge University Press.

Swales, John; Feak, Christine. 1994. *Academic writing for graduate students: essential tasks and skills.* Ann Arbor: University of Michigan Press.

Zhao, Jun. 2017. "Native speaker advantage in academic writing? Conjunctive realizations in EAP writing by four groups of writers". Ampersand 4, 47-57.

Appendix A: Simplified overview of informal traits in different disciplines

Although this table does not prove extensive details about the numbers, it shows whether a specific trait has increased or decreased in one discipline. A + sign indicates that the feature has increased, while a – sign is indicative that an informal feature has decreased. The number 0 means that there has been no significant change.

Table 15: Entire development of each informal trait in each discipline

Informal trait	1st pers. pron.	2nd pers. pron./ det.	Listing expressions	Initial conj.	Exclamations	Direct questions	Contractions	Unatt. anaph. pron.	Stranded prep.	Split infinitives
MATHEMATICS	+	+	+	-	+	-	+	0	-	+
BIOLOGY	+	0	+	+	-	-	-	-	-	+
LINGUISTICS	-	-	+	-	-	-	+	+	-	+
ECONOMICS	+	-	+	-	-	-	+	0	+	+

Appendix B: Hyperlinks

Table 16 entails the web pages of each journal. These are also the pages from which the data has been collected. Future studies may use this table to collect open access data.

Table 16: Sites used for data collection

JOURNAL	Available under:
Journal of Experimental Biology	https://journals.biologists.com/jeb
The Quarterly Review of Biology	https://www.journals.uchicago.edu/toc/qrb/current
The Biological Bulletin	https://www.journals.uchicago.edu/toc/bbl/current
The American Naturalist	https://www.journals.uchicago.edu/toc/an/current
Advances in Mathematics	https://www.sciencedirect.com/journal/advances-in-mathematics/issues
Acta Mathematica	https://projecteuclid.org/journals/acta-mathematica/issues
Manuscripta Mathematica	https://link.springer.com/journal/229/volumes-and-issues
Mathematical Notes	https://link.springer.com/journal/11006/volumes-and-issues
De Economist	https://link.springer.com/journal/10645/volumes-and-issues
Journal of Economics (J Econ)	https://link.springer.com/journal/712/volumes-and-issues
Review of World Economics	https://link.springer.com/journal/10290/volumes-and-issues
Cambridge Journal of Economics	https://academic.oup.com/cje/issue/46/3
Journal of Linguistics	https://www.cambridge.org/core/journals/journal-of-linguistics/all-issues
Theoretical Linguistics	https://www.degruyter.com/journal/key/thli/html?lang=en
Annual Review of Applied Linguistics	https://www.cambridge.org/core/journals/annual-review-of-applied-linguistics/all-issues
Language	https://www.jstor.org/journal/language

Deutsche Zusammenfassung

Ziel dieser Arbeit war es, die Entwicklung informeller Merkmale in sozial- und naturwissenschaftlichen Disziplinen über einen Zeitraum von 40 Jahren zu untersuchen. Informalität nimmt viele Formen an, vom Vokabular und Stil bis hin zu bestimmten Wortgruppen wie Personalpronomen. In dieser Studie wurde eine abschließende Liste von 10 informellen Merkmalen verwendet, um den Grad der Informalität in jedem Zeitraum zu messen. Darüber hinaus werden alle diese 10 informellen Merkmale als gleichermaßen informell angesehen. Drei Korpora wurden zusammengestellt und bestehen aus begutachteten Forschungsartikeln aus 4 verschiedenen Disziplinen: Mathematik, Biologie, Wirtschaftswissenschaften und Linguistik. Diese drei Korpora sind repräsentativ für die unterschiedlichen Ebenen der Informalität in den Jahren 1980, 2000 und 2020. Dies geschah, weil es für das Schreiben oder die Genres unmöglich ist, sich über Jahrzehnte hinweg dem Wandel zu widersetzen, da die Sprache direkt von der Gesellschaft beeinflusst wird. Informalität in akademischen Zeitschriften ist potenziell schädlich und entspricht nicht dem festgelegten Standard. Daher ist es wichtig, den Grad der Informalität zu messen, die Änderungen zu verfolgen und den Trend in jeder Disziplin zu erkennen sowie alle signifikanten Ergebnisse hervorzuheben. Die Hypothese ist, dass informelle Merkmale in den Naturwissenschaften zugenommen und in den Sozialwissenschaften abgenommen haben.

Die Daten wurden aus vier Zeitschriften aus jeder Disziplin gesammelt und ungefähr 40.000 Wörter aus jeder Zeitschrift extrahiert. Die Korpora wurden mit SketchEngine und AntConc auf informelle Merkmale gescannt, die Häufigkeiten in Tabellen eingetragen und die normalisierten Häufigkeiten berechnet. Das Ergebnis liefert unter anderem einen detaillierten Überblick über die Entwicklung jedes informellen Merkmals in jeder Disziplin, die Gesamthäufigkeit informeller Merkmale zwischen den Disziplinen sowie dominante Merkmale in jeder Disziplin. Letztendlich bewies ein Korrelationstest nach Pearson, dass es eine Korrelation gibt, dass die Informalität im Laufe der Jahre zunimmt. Darüber hinaus verwenden Autoren, die keine englischen Muttersprachler sind, informelle Merkmale in ihrem Schreiben nicht häufiger als Autoren, die Muttersprachler sind. Dieser Ansatz wurde

gewählt, da frühere Studien einen nahezu identischen Ansatz verfolgten und signifikante Ergebnisse hatten. Diese Studie hat sich jedoch auf mögliche Unzulänglichkeiten der vorherigen Studie konzentriert und trägt zu der Forschung bei, die zur Informalität in akademischen Texten durchgeführt wird. Darüber hinaus konzentriert es sich auf eine relativ junge Zeit.

Abstract

This paper aimed to scrutinize the development of informal traits in disciplines from social and natural sciences throughout 40 years. Informality takes many forms, from vocabulary and style up to specific word groups such as personal pronouns. In this study, a conclusive list of 10 informal traits was used to measure the level of informality in each period. Moreover, all of those 10 informal traits are viewed as equally informal. Three corpora were compiled and consist of peer-reviewed research articles from 4 different disciplines: mathematics, biology, economics, and linguistics. These three corpora are representative of the different levels of informality in 1980, 2000, and 2020. This was done because it is impossible for writing or genres to resist change throughout decades as language is directly affected by society. Informality in academic journals is potentially detrimental and does not agree with the set standard thus it is important to measure the level of informality, track the changes and detect the trend in each discipline as well as highlight any significant results. The hypothesis is that informal traits have increased in natural sciences and decreased in social sciences.

The data was collected from four journals from each discipline and approximately 40.000 words from each journal were extracted. The corpora were scanned for informal traits using SketchEngine and AntConc, the frequencies were entered in tables, and the normalized frequencies were calculated. The result provides a detailed overview of the development of each informal trait in each discipline, the overall frequency of informal traits among disciplines as well as dominant traits in each discipline amongst others. Finally, a Pearson's correlation test proved that there is a correlation that informality increases as the years progress. In addition, authors who are non-native speakers of English do not use informal traits more often in their writing than authors who are native speakers do. This approach was selected since previous studies have employed an almost identical approach and had significant findings. This study, however, has focused on potential imperfections of the previous study and contributes to the research conducted on informality in academic texts. Moreover, it focuses on a rather recent period.