



FILOZOFICKÁ FAKULTA
UNIVERZITY KARLOVY
V PRAZE

ÚSTAV ANGLICKÉHO JAZYKA
A DIDAKTIKY

DIPLOMOVÁ PRÁCE

Bc. Veronika Pojarová

Forensic analysis of anonymization strategies in English

Analýza anonymizačních strategií v angličtině

Vedoucí práce:

Praha, 2016

Mgr. Jiřina Popelíková

I would like to thank my supervisor Mgr. Jiřina Popelíková for her support, guidance and encouragement throughout this challenging project. I would also like to thank doc. Radek Skarnitzl for his advice regarding several sources. Finally, thank you to my friends, who helped facilitate the acquisition of texts for the analysis; Elizabet, Eliška, Conor, Laura, Marita, Olivia, Sarah, Scott.

I declare that the following MA thesis is my own work for which I used only the sources and literature mentioned, and that this thesis has not been used in the course of other university studies or in order to acquire the same or another type of diploma.

Prohlašuji, že jsem diplomovou práci vypracovala samostatně, že jsem řádně citovala všechny použité prameny a literaturu a že práce nebyla využita v rámci jiného vysokoškolského studia či k získání jiného nebo stejného titulu.

V Praze, 22. 7.2016

.....

Bc. Veronika Pojarová

Abstract

The objective of this thesis is to identify those specific aspects of written style which native speakers of English modify when attempting to anonymize their texts. The conclusions are based on the analysis of 20 texts by 10 authors, all of whom are native speakers of English. Two texts dealing with the same topic were produced by each participant; one was written as an official letter of complaint, and the other was written as an anonymous letter. The bulk of the results are grounded on a qualitative stylistic analysis of the individual texts, with merely a brief survey of quantitative methods. The purpose of the introductory chapter is to familiarize the reader with the subject of forensic authorship analysis, to provide a brief summary of the current state of research, and to introduce a series of empirical studies. The practical part of the thesis presents the qualitative stylistic analysis, provides a shorter summary of the quantitative analysis, and finally ventures to draw meaningful conclusions from the results. The results showed that the majority of authors manipulated with the style/register of the texts and with the specific lexical choices, whereas none of the 10 authors made alterations to spelling and only 2 authors chose to change the punctuation in the anonymous text. However, factors such as author variables and sample size must be taken into consideration before any definite conclusions can be drawn from the results.

Keywords

anonymization strategies, forensic linguistics, authorship attribution, idiolect

Abstrakt

Cílem této práce je identifikovat konkrétní rysy psaného projevu, které jsou rodilými mluvčími modifikovány při anonymizaci textu. Závěry vychází z analýzy 20 textů napsaných 10 autory, z nichž všichni jsou rodilými mluvčími angličtiny. Každým respondentem byly poskytnuty dva dopisy popisující stejnou situaci, přičemž jeden měl být oficiální stížností, a druhý anonymním dopisem. Většina výsledků se zakládá na kvalitativní stylistické analýze jednotlivých textů, kterou doplňuje krátká prezentace kvantitativních metod. Úvodní kapitola si klade za cíl přiblížit čtenáři problematiku forenzní analýzy autorství, stručně shrnuje současný stav výzkumu v této oblasti a představuje řadu empirických studií. Ve výzkumné části práce je představena stylistická analýza, která je doplněna shrnutím výsledků kvantitativní analýzy, a následná interpretace výsledků přináší odpovídající závěry. Výsledky ukazují, že většina autorů pracovala se stylem/registrem anonymních textů, stejně tak prováděla změny na rovině slovní zásoby, zatímco ani jeden z deseti autorů neupravoval pravopis, a pouze dva autoři v anonymním textu změnili interpunkci. Před ustanovením konečných závěrů je však potřeba vzít v úvahu další faktory, kterými jsou velikost vzorků, nebo řada proměnných, např. motivace autorů.

Klíčová slova

anonymizační strategie, forenzní lingvistika, určení autorství, idiolekt

Table of contents

1	Introduction.....	7
1.1	Idiolect	7
1.2	Style and Variation.....	8
1.3	Methods of Author Identification	9
1.3.1	Orthography	10
1.3.2	Lexical Richness	10
1.3.3	Parts of speech	11
1.3.4	N-grams.....	12
1.3.5	Sentence Length.....	14
1.3.6	Stylistic Analysis	14
1.3.7	Qualitative versus Quantitative Analysis.....	15
1.4	Authorship disguise and imitation	16
1.5	Linguistic self-awareness and competence.....	16
1.6	Similar Experiments.....	17
1.6.1	Textual Fingerprints.....	17
1.6.2	Disguised Voices.....	18
1.6.3	Imitation.....	19
1.7	Hypothesis	20
2	Material and Method	22
2.1	Material	22
2.2	Quantitative analysis	22
2.3	Qualitative analysis	24
3	Results.....	25
3.1	Quantitative analysis	25
3.1.1	Sentence length	25
3.1.2	Sentence Complexity.....	26
3.1.3	Standardized TTR.....	28
3.1.4	Parts of speech	28
3.1.5	N-grams.....	30
3.2	Qualitative stylistic analysis.....	33
3.2.1	Author A	33
3.2.2	Author B.....	34
3.2.3	Author C	36
3.2.4	Author D	36

3.2.5	Author E.....	38
3.2.6	Author F.....	39
3.2.7	Author G.....	42
3.2.8	Author H.....	43
3.2.9	Author I.....	44
3.2.10	Author J.....	46
3.3	Summary.....	48
4	Discussion.....	50
5	Conclusion.....	56
6	References.....	58
7	Resumé.....	60
8	Appendices.....	64
8.1	Appendix I.....	64
8.2	Appendix II.....	65

1 Introduction

1.1 *Idiolect*

At the core of forensic linguistics is “the theoretical position that every native speaker has their own distinct and individual version of the language” (Coulthard, 2004: 31). The idiosyncratic personal dialect or style of one particular person is an *idiolect*, and it is essentially the way in which one uses language in interaction, resulting in a style which is characteristic of the individual. Whereas social or regional (geographical) dialects are held in common by large groups of people, an idiolect is as unique as an individual’s linguistic experience, constantly evolving with dependence on a vast array of factors. Any number of texts produced by different language users can never be identical, a fact due not only to the immeasurable number of possible combinations of words. The most overt and easily perceptible differences will, in all probability, arise in spelling and punctuation. Each text is almost guaranteed to have an inventory of phrasemes specific to itself, the selection of which by the author was far from accidental. It can even be set apart from the others by something as seemingly inconsequential as the author’s lexical choices. All of these aspects of language lend any text, written or spoken, a distinct style. Naturally, the longer the text, the greater the certainty with which it may be told apart from others.

The influence of the speaker’s regional dialect is only one of many possible factors which will almost certainly have an effect on the inventory of the linguistic means by which he chooses to communicate. “[Each] speaker fashions an idiolect from the common language, and in their reception, in so far as each speaker helps to produce the message which he perceives and appreciates by bringing to it everything that makes up his singular and collective experience.” (Bourdieu 1991: 39) Idiolect is dependent on speaker (author) variables such as gender, age, level of education, social class, and personality. However, the reality is much more complex than that, because everyone comes into contact with language in very specific contexts and in varying degrees. Ultimately, one’s idiolect will be a reflection of all the books and magazines that he has read, what he has heard on the television or on the radio, the people who he has spoken to most frequently, and basically any other kind of linguistic input. (Hoey, 2005: 181) All of these experiences play a role in forming the speaker’s linguistic identity.

As a result of these individual variables, the speaker's language sets him apart from others and makes him, to a certain extent, identifiable. In fact, one's personal style can be thought of as something of a linguistic fingerprint, a product of a lifetime of language use which has become ingrained in his identity.

The very existence of the concept of idiolect has been denied by a number of linguists, among them Jakobson (1971:82), who believed idiolect to be "a somewhat perverse fiction". Barlow (2010), in defence of the notion of idiolect, proposes that the problem lies in its original definition: "These particular negative positions may have arisen in response to Bloch's original definition of an idiolect as "the totality of the possible utterances of one speaker at one time in using a language to interact with one other speaker."

However fragile the concept of idiolect may be, the truth remains that its existence is the underlying principle of authorship attribution, a discipline popular in the forensic setting and one which has in countless cases been successful in identifying authors through linguistic analysis.

1.2 Style and Variation

The term 'style', in one of its broader definitions, refers to "a selection of language habits, the occasional linguistic idiosyncrasies which characterise an individual's uniqueness." (Crystal and Davy, 1969: 9) Whenever there are more available linguistic variants from which a language user must select one, we may speak of 'style'. These choices are governed by many factors, including the identity of the addressee, the social situation, the genre and the medium (e.g. text message, letter, e-mail) and the purpose with which the text is produced. This results in a great amount of variation not only among several individuals, but also within the one person's idiolect, especially depending on the circumstances, context and aims with which the text was produced. These factors account for possible stylistic discrepancies between two texts from the same author.

Internal variation, however, is not dependent solely on the register or the text type. The difficulties surrounding intrapersonal variation have been pointed out on a number of occasions. "As Labov showed, when we survey how speech varieties, we find variation 'within the individual speaker' across contexts of talk, as well as between individuals and groups." (Coupland, 2007: ix) Though variability within an idiolect may seem too problematic for successful author identification, it does in fact create the opportunity to

assess the range of variation, which in and of itself can be the distinctive mark of one author. In fact, Olsson claims that “under certain circumstances the range of variation a writer exhibits will itself be distinctive.” (Olsson, 2014: 49) This claim was tested on a corpus of mobile phone text messages, where variants of you (you, u, ya), to (to, 2) and are (are, r) were examined. Although most of the authors did not use one variant exclusively, the combination and range of the variants turned out to be unique enough to facilitate the identification of one author from within the corpus of 53. The conclusion stemming from this experiment was that similarity in the range of variation in two texts could indicate increased chances of shared authorship (Olsson, 2014: 50-51).

This is still a tentative conclusion and Olsson warns that “even when two texts are produced by the same author on the same topic, we should not prejudge the extent to which they may have shared characteristics.”(Olsson, 2014: 63) Nevertheless, there is a number of studies, some of which are outlined below, bearing encouraging results in favour of the possibility of linguistic fingerprinting.

1.3 Methods of Author Identification

Conducting a forensic analysis of the text, there are a number of ways in which it may be approached, some of which are described below. Every facet of a text is relevant when attributing authorship. While idiosyncratic spelling and punctuation habits are very obvious at first sight and a reflection of the identity of the person who produced them, there is arguably not as much opportunity for variation. The possibilities available when making lexical choices make it a crucial component of one’s idiolect. Richness of vocabulary and the preference for certain words, their co-selection and collocational patterns; these all constitute an idiolect. The degree of clause complexity and density, and furthermore the question of how clauses are linked, is also worth noting.

Olsson (2009: 145) provides a list of the basic style features which he considers to be of particular interest in a forensic linguistic investigation. First in order of importance is the appearance of anything “rare, unusual or erroneous” in grammatical structure, for example the apparent notion that one word is in fact two, as in the case of “photo copied” or “back fired”. The significance of this feature depends on the number and type of matches. “Unusual punctuation habits” are next on the list of features recommended for analysis, and can be most helpful in combination with syntax. Furthermore, the use of unlikely

idioms and phrases should be inspected for markedness, as should spelling errors, especially those which are rare. Document layout is another aspect open to evaluation, however one must be especially cautious, as this feature is very closely tied with cultural norms.

The following sections provide a brief summary and several examples of the various methods used in authorship analysis, covering a large scale of style features available for analysis.

1.3.1 Orthography

One of the most noticeable characteristics of a text is its spelling, Olsson (2014:70) mentions the case of an employee who sent hate mail aimed at herself in order to make it appear as though her colleagues were harassing her. The author of the emails seems to have been feigning incompetence in an attempt to anonymize her texts, with what appears to be the inclusion of frequent and conspicuous misspellings, such as *yourselves*, *ritten* and *lightss*. However, the spelling errors were inconsistent; words such as *considering* and *eventually* were spelled correctly. Although the inconsistency in spelling did not play the decisive role in the identification of the text's true author, it was the factor which drew attention to the fact that the misspellings were a deliberate attempt at anonymization.

1.3.2 Lexical Richness

The higher the number of different words used in a text, the richer its lexis. Lexical richness can be, at the most basic level, measured as a type- token ratio, or TTR, which measures the number of different words in the text (*types*) by the total number of words (*tokens*). (Van Gijssel et al., 2005:2)

Honoré devised a simple formula to measure the richness of a text which is $[100 \times \log N / (1 - V_1/V)]$, where N is the total length of the text in tokens, V_1 is the number of hapaxes and V is the number of types. (Coulthard and Johnson, 2007:165) Both the TTR measure and Honoré's formula have been found to be extremely dependent on text length, due to the fact that the repetition of frequent function words such as *has*, *have*, *it*, *a*, and *the* gradually lowers the text's TTR. The attempts at remedying this have been numerous, such as the introduction of a standardized TTR, to facilitate the comparison of texts of various lengths. Alternatively, as an amendment to Honoré's formula, Winter and Woolls suggested that

only the richness of lexical words should be measured, omitting the function words. This method was successfully applied by Coulthard, who was able to measure the stylistic difference between two authors of a co-authored book by examining sentence length alongside textual richness. (Coulthard and Johnson, 2007:165-6)

The TTR has also been found to be dependent on the text's register, quite in line with expectations in that the more informal the text, the lower its TTR. Similarly, word classes also display differences in lexical richness; the TTR of nouns was shown to be most dependent on the register, whereas with adjectives and verbs this was much less pronounced, and the TTR of function words was least affected by register. (Van Gijssel et al., 2005:13)

Many other methods have been tested with varying degrees of success in determining the identity of a text's author. Tweedie and Baayen mention "Sichel's generalised inverse Gauss-Poisson model", the plotting of "developmental profiles" and "comparison techniques using text randomisation." (1998:331-340) However as these methods require an advanced knowledge of mathematics and computational linguistics, they are beyond the scope of this thesis and need not be discussed in detail.

1.3.3 Parts of speech

Allen (1974:911) claims that studying the distribution of certain parts of speech can be helpful with the successful identification of an author, for example a larger proportion of nouns usually indicates a more learned style, associated with more cultivated writing habits of authors. Furthermore, he suggests that the proportion of one word class to another is worth considering, saying that "some authors tend to have a higher ratio of adjectives per noun than others. Perhaps something significant can also be found in the percentage of nouns without adjectives, with one adjective, two, and so forth. Some critics, with good justification, are examining the importance of the ratio between nouns and verbs alone in literary texts."

A higher frequency of long noun phrases and a tendency towards nominalization is typical of academic prose, and a higher style in general, attributable to the discussion of "general (sometimes abstract) patterns and concepts in academic prose" (Biber and Conrad, 2009: 121) There appears to be a generally accepted idea that the ratio of verbs to nouns varies discernibly across registers, i.e. that "activity verbs and time/place adverbs are much more common in the "friends and family" e-mails than in the other categories, reflecting the

primary focus on everyday activities rather than conceptual discussions. In contrast, attributive adjectives and nominalizations are much more common in the professional e-mails, especially those written by “strangers,” reflecting their informational focus (similar to academic prose)” (ibid. 187).

1.3.4 N-grams

One of the most remarkable aspects of a text in terms of its uniqueness is that of strings of words, from rare collocates to entire lexical phrases. Although the use of a number of phrasemes is likely to be shared by the majority of language users and therefore not likely to play a significant role in narrowing down to one person, various combinations of phrasemes may be instrumental in characterizing one’s idiolect. “The associations that people build between words and the ways in which they produce them in combinations is a psycholinguistic phenomenon, and has been analysed in terms of ‘lexical phrases’ (Nattinger and DeCarrico, 1992), ‘formulaic sequences’ (Wray, 2002, 2008), ‘lexical priming’ (Hoey, 2005; Pace-Sigge, 2013), and usage-based theories of lexico-grammar such as exemplar theory (Barlow, 2013). One factor which these different approaches have in common is that they all emphasise the personal or ‘idiolectal’ nature of preferences for certain word combinations and collocational patterns.” (Johnson and Wright, 2014: 40)

Johnson and Wright (2014:38) pose the following question: “Looking at any set of texts by any author, is it possible to identify ‘n-gram textbites’, small textual segments that characterise that author’s writing, providing DNA-like chunks of identifying material?” In their study they attempt to identify one author from a number of email samples with a combination of the n-gram textbite approach and qualitative stylistic analysis. Based on the analysis of recurring n-grams it was possible to successfully isolate the one target author. This conclusion was supported and confirmed by the qualitative analysis. It is concluded that “a word n-gram approach can be used not only to identify and isolate a number of n-gram textbites that distinguish his professional email style from that of other employees in the same company, but also to successfully identify him as the author of text samples, including some as small as 77, 84 and 109 tokens.” (Johnson and Wright, 2014:62)

Hanique et al. (2015) studied phonetic transcriptions in order to determine whether there was any substantial variation in the lexical choices made in casual conversations by

speakers belonging to a homogenous group. They applied a classification algorithm which assigned texts to speakers based on the unigrams and bigrams extracted from their transcribed conversations on the basis of frequency. This method proved to be reasonably accurate, showing that there is considerable variation even among speakers from the same social background.

Barlow (2010) focuses on the presence of high frequency lexical units on a relatively large sample of data, based on his hypothesis that “the variation that distinguishes individual speakers lies in the profile of the central components of lexicogrammar, and not only in some idiosyncratic peripheral phraseology. On the basis of earlier preliminary study (Barlow and Kemmer 2004), it appears to be worthwhile to follow this centrality hypothesis and so we focus on high frequency items rather than search for low frequency (but highly distinctive) markers of the speech of individuals.” (Barlow, 2010: 3)

For his paper he analysed the language habits in the speech of five White House Press Secretaries. The advantages of using this material were firstly the unity of the discourse context, and also the size of the resulting corpora, which were “between 200,000 and 1,200,000 words of running text” for each individual (Barlow, 2010: 3). An analysis of bigrams showed that there is a significant degree of stability intrapersonally, in that one speaker’s six 200 000-word samples proved consistent in their 15 most frequent bigrams. These bigrams were recurrent with all five speakers, nonetheless “the proportion in which they are used varies from one speaker to the next, but remains fairly constant for each speaker,” (Barlow, 2010: 8) indicating a high level of interspeaker variation.

Barlow comments on the significance of bigrams in terms of how they reflect higher levels of language, saying that “bigrams themselves are obviously not the units of grammar. Nevertheless, they are powerful indicators of units of grammar since differences in production of constructions and other grammatical units will surface as differences in bigram distributions.” (Barlow, 2010:7)

Coulthard and Johnson (2007) discuss the effectiveness of single identical lexical strings, claiming that even a string of ten words can be unique. This is demonstrated on a case of disputed police interview records, where the accused maintained that his statement had in fact been written based on a previous dialogue (interview record). Several identical strings found in both texts were searched for in Google, returning zero hits and therefore supporting the claims of the appellant. One should nevertheless avoid coming to hasty

conclusions based on internet searches. Although the authors justify their use of Google over a language corpus, on the basis of its size and accessibility, its reliability and its status of a representative language corpus have been called into question. (Jones, 2010: 64)

1.3.5 Sentence Length

Of the criteria used for author identification in the past, Allen (1974:910) considers sentence length to be one of the most readily accessible methods, nonetheless with the major disadvantage of being “too easily controlled by an author.” The extent to which an author knowingly chooses to control the length of his sentences will also be touched upon in this thesis.

1.3.6 Stylistic Analysis

In their description of the stylistic structure of a text, Crystal and Davy (1969, 83) identify a number of stylistic features which they assign to specific levels of analysis and particular stylistic dimensions. The basic levels of analysis are phonetics, graphetics, phonology, graphology, grammar, vocabulary, and semantics. Each of these components subsumes a number of specific features which should be systematically analysed based on a chosen model and any significant characteristics taken note of. The two basic dimensions of stylistic description are province (legal, religious, etc.) and status (formal, informal, etc.). For example, use of contracted forms such as *we'll*, *they'd*, *he's* is a significant feature on the level of grammar, and indicates that text belongs to the stylistic dimension of *informal*. Crystal and Davy draw attention to a strong tendency for certain categories to co-occur, for example between ‘legal’ and ‘formal’ language, rendering the co-occurrence of ‘legal’ and ‘colloquial’ in one text highly improbable.

This sort of register mixing can be an indication of dual authorship, plagiarism, or attempts at imitation or anonymization. Olsson and Luchjenbroers (2014: 63) provide the following excerpt from an anonymous letter: “Go to Seattle, heck I lived there once, no big deal [...] pick up our gun, and drop it off at a specified location when you are done.” The contrast between the colloquial tone of “heck I lived there once” and the formal phrase “specified location” is evident.

It is recommended that stylistic analyses are carried out in two frameworks simultaneously; linguistic (levels of analysis) and stylistic (dimensions of description). However, the categories of description are in no way firmly established, just as “we cannot say how many categories of status there are, or how far a province may be legitimately sub-classified.”

(ibid. 87) It is to a great extent the responsibility of the linguist to make decisions regarding the classification of stylistic categories.

This method should provide a basis for statements which offer a comparison of two texts in order to determine “exactly what it is that differentiates one author from another” (ibid. 88) However, it is of utmost importance to first determine which linguistic features are idiosyncratic and which are integral to the genre of the text or the social position of the author with regard to his addressees.

1.3.7 Qualitative versus Quantitative Analysis

In current approaches to author identification, we can observe an inclination towards more objective methods. There is a heavy reliance on computing and statistics rather than qualitative analysis, which can be highly dependent on the individual conducting it. Nowadays, it would seem that forensic authorship attribution belongs as much in the realm of computer science as it does in that of linguistics. In the case of alleged plagiarism, for example, there is usually a relatively large corpus of language for the linguist to scrutinise. However, the reality of working with an authentic text in a forensic setting often means that the available linguistic information is extremely limited. On a number of occasions described for example by Olsson (2009:7, 61), as little as a single short text message had to suffice, rendering quantitative computational methods rather impractical and calling instead for a detailed stylistic analysis.

Solan (2013: 562) describes the current trends in authorship attribution, finally asserting that “there is no reason to conclude that intuitive expertise based on experience and insight fares any better or worse than does algorithmic expertise”. Although computational methods are generally more appealing to a jury because of their apparent objectivity and replicability, there is no evidence which would suggest that a more focused stylistic analysis conducted by an experienced linguist is in any way less reliable. Solan points out that these two approaches should supplement each other, calling for an increased collaboration among scholars.

1.4 Authorship disguise and imitation

Olsson (2009: 138) discusses a case in which a man suspected of sexual assault sent a number of letters to his neighbours asserting his innocence and attempting to dispute the claims made against him. Concurrently, these same neighbours had been receiving threatening letters from an anonymous sender, with strikingly similar content to those sent by the accused man. Olsson observed several marked structures in both texts, for example the inversion of tenses and the disregard for temporality in projected clauses as in “I could not believe this as I am and was impotent” (Known author) and “Neither would have known that Joe [...] is and was impotent” (Unknown author). Although this type of structure cannot be classified as erroneous, it is uncommon and therefore displaying some level of markedness. Searches in the BNC and on the internet showed that the preference for a past-present formulation is indisputable. Another marked feature shared by the known and unknown authors was the occurrence of punctuation at unexpected syntactic boundaries, for example in “I and my partner, have over the past years and months, studied the case”. Another point of interest is the unconventional phrase “I and my partner”. Striking similarities in the disorderly structure of lexis were also observed, specifically words such as “inter course” and other wise,” which were separated by a space in the case of both authors. A list of matches between the known and unknown texts was compiled, on various structural levels. Some matches were considered more vital than others, based on a hierarchy which is described in section 1.3. (Olsson, 2009:145) It is possible that the report based on these observations contributed to subsequent confession of the accused.

1.5 Linguistic self-awareness and competence

The extent to which an attempt at disguising written style can be successful is arguably dependent on the author’s metacognition, or some form of linguistic self-reflection. Perhaps the first question which must be posed is whether people possess an awareness of the fact that they have an idiolect, and if so, to what degree they are able to control it.

When Jakobson challenged the understanding of structural stylistics with the thought that there are more aspects at play than had been previously believed, “his ‘metalingual’ function [...] pointed to reflexive and self-referential processes at work in linguistic style. He therefore opened a perspective on language in some ways referring to itself, and

speakers speaking through some level of awareness of their own stylistic operations and constructed images and identities.” (Coupland, 2007: 11)

It’s not uncommon to hear of a text “not sounding like the author” or that someone “doesn’t write like that” in connection to written language, indicating that people are, at least on some level, aware of the fact that when expressing themselves through language, they do so in a manner which somehow characterizes their distinctive style. This could perhaps point to the fact that the existence of idiolect is generally agreed upon. However, it is unclear whether people are equally as prepared to judge their own idiolect, or whether they are simply too close to see it clearly. Is the majority of people, at least those without extensive linguistic education, in the possession of such a level of introspection that they can objectively identify the differences between their style and that of another?

If it is indeed the case that people do have a deep enough linguistic self-awareness, then it would stand to reason that they should be capable of altering their idiolectal features at will. People who we could expect to have the clearest ideas about their own language habits are probably those with linguistic education, including learning a foreign language. It could be that the level of linguistic competence determines one’s ability to critically assess his own language.

1.6 Similar Experiments

1.6.1 Textual Fingerprints

Baayen et al. (2002) propose the existence of the concept of linguistic or ‘textual’ fingerprinting, and set out to support this hypothesis in their experiment in authorship attribution. The experimental design is similar to that of this thesis, especially concerning the collection of data. The experiment was carried out in Dutch, with the participation of 8 Dutch students, who were native speakers of the target language. Each of the students produced a total of 9 texts, of approximately 1000 words each, in the genres of fiction, argument and description. The texts were strictly controlled for genre and topic, with very specific tasks such as the retelling of Little Red Riding Hood. The ensuing corpus of 72 texts was analysed, with the primary aim of determining the extent to which authors of similar backgrounds could be identified on the basis of genre controlled texts. Especially relevant to this thesis is the method of data collection, i.e. texts artificially produced by native speakers according to very specific instructions, and based on which it was possible to conduct a series of analyses with meaningful results.

Following a number of analyses, it was concluded that even for authors of similar backgrounds there is a significant variation in style, supporting the hypothesis that there is such a thing as “textual fingerprints”. The influence of genre was found to be considerable; a much higher accuracy of author classification was observed when the genre was restricted. Although the role of genre in idiolect is not among the main concerns of this thesis, the knowledge of this is essential for the collection of appropriate material. In order to eliminate the effect of genre, the data in this thesis was strictly controlled for genre and topic. One of the most surprising findings of Baayen et al. (2002) was the impact of including punctuation marks in the analysis, which led to a higher classification accuracy. The authors conclude with the proposition that “punctuation marks may prove to be effective style markers, especially for texts that have not been subjected to editorial normalization.” The question is whether the language users themselves are aware of the role that punctuation plays in forming their written style.

1.6.2 Disguised Voices

Although experiments dealing with anonymization strategies are sparse, a study in the neighbouring field of forensic phonetics by Masthoff (1996) provides a number of valuable insights into the issue of masking one’s identifying characteristics. Masthoff’s endeavour was to identify the most frequent ways in which speakers disguise their voices, which he achieved by recording twenty undergraduate students, all of them native speakers of German. Their task was to read two texts in their own voices, and then proceed to read the first text again, but this time they were instructed to disguise their voices so as to avoid recognition: “Imagine being a blackmailer who is to transmit a spoken message by telephone. Obscure your identity to the best of your knowledge by disguising your voice while still clearly delivering the meaning of the prescribed sentence.” (Masthoff, 1996: 162)

All of the resulting samples were analysed for four categories; *phonation*, *respiration*, *articulation* and *manner of speaking*. *Phonation* covers the alteration of regular vocal fold vibration, whereas *articulation* refers to disguise by means of altering the airflow above the larynx, covering anything from nasality to a simulated speech pathology. *Manner of speaking* subsumes essentially everything above the level of individual segments, for example speech tempo and temporal patterns. The results show that the majority of speakers (55 per cent) selected only one means of disguise from one of the categories, and that no speaker disguised his voice using more than two parameters at one time. In all, the preferred forms of disguise were at phonation level for 65 % of the speakers, most often achieved by

whispering, either by itself or in combination with one other feature. Manner of speaking was altered only in combination with another parameter, suggesting that it was not perceived as a sufficient means of disguise in and of itself.

As a final point, it must be said that only three out of the twenty speakers were judged as successful in disguising their voices, even though they had all deemed their own efforts satisfactory. In fact, it is remarkable that “several speakers who exhibited a strong regional accent in their modal voices never did attempt to obscure this important speaker-specific feature.” (Masthoff, 1996: 164) When a speaker is attempting to disguise his voice and simultaneously produce a meaningful text, the result must be a compromise between intelligibility and anonymity. Based on the results of the experiment, Masthoff concludes that “in the majority of the cases, a sufficient number of parameters can be expected to be available which will permit a reasonable examination and subsequently a successful identification of the speaker.”(Masthoff, 1996:166).

At this point an analogy may be drawn between written and spoken language, where phonation and articulation constitute the basic form of the given utterance, just as morphology does in written texts. Prosody and its written counterparts are perhaps not so simple to grasp and therefore disguise. In that respect, similar results could be expected from a linguistic experiment in which the subject’s style is disguised in the written medium.

1.6.3 Imitation

Zetterholm (2003), although she had no interest in strategies for disguising voices, produced an experiment whose findings could be interpreted in relation to author anonymization or imitation in forensic linguistics.

Working with the premise that voice impersonators select features which are most salient and subsequently imitate them, Zetterholm’s main point of interest was the perception of vocal characteristics by three impersonators (two of whom were professionals) and specifically the question of whether there would be unity in their choice of the target speaker’s most distinct features. Constituting the material were recordings by the 9 target speakers, 22 recordings of imitations produced by the three impersonators, and recordings of the impersonators’ own voices.

On the segmental level, the impersonators all identified the phonemes which were characteristic of the given target speaker and his dialect, and imitated them successfully, albeit exaggeratedly. A thorough acoustic analysis of the data was carried out, during

which trained phoneticians made judgments of pitch, voice quality, speech tempo, dialectal markers and individual phonetic habits. They concluded that all three impersonators had predominantly chosen the same characteristic features of the target speakers and that all three had manipulated their pitch level so that it would correspond to that of the target speaker, although the amateur impersonator was not as successful in achieving the desirable level of variation. As for voice quality, only the two professional impersonators were judged to have copied it successfully, although a solid attempt was made on the part of the amateur. All three impersonators appear to have taken into consideration the tempo, rhythm and pausing of the target speakers, the two professionals doing so with more variation. The professional impersonators also copied the speaker's patterns of intonation, which the amateur did not manage to do.

Zetterholm concludes: "The results of this study indicate that there may be some individual features in a speaker's voice and speech behavior that seem to be more important than others for the recognition of a voice, both considering the features selected by the three impersonators and the comments from the listeners. That may give a clue about individual features useful in a speaker identification task." (2003:2208)

Analogously, in the field of linguistics this could indicate that an author is more likely to exert some control over variation on lower levels of language, for example orthography and morphology. On the other hand it can be expected that variation on higher levels, such as syntax and discourse, will be overlooked. Furthermore, there may be reason to expect those with a linguistic education to be more capable of working with their own idiolect and altering it in order to disguise their style. Generally speaking, the findings of Zetterholm's study suggest that segmental aspects of speech are easier to imitate than the suprasegmental.

1.7 Hypothesis

The central hypothesis of this thesis is that language users have varied degrees of linguistic self-awareness and competence, which is reflected by the depth and the nature of their anonymization strategies. Based on previous research such as that of Masthoff (1996) and Zetterholm (2003), it was expected that speakers would disguise their idiolect only on certain levels of the language, for example that they would make conscious changes to their style on the level of spelling and morphology, but would fail to alter longer strings, such as phrasemes which are typical of their idiolect. One of the basic assumptions was that some aspects of 'anonymous style' could possibly be shared by the majority of language users.

The research was conducted qualitatively, through a detailed stylistic analysis of the available texts produced in known and anonymous modes. The results are supplemented by quantitative corpus based data analysis.

Desired outcomes of this research include an assessment of the speakers' awareness of their own idiolects. Another endeavour is to identify the areas of language with least intrapersonal variability, i.e. on which levels could it still be possible to identify an author even after his attempts to anonymize the text.

It is hoped that it will prove possible to make some generalizations about the anonymizing strategies of language users, which could then prove to be of use in practical applications such as authorship attribution in the field of forensic linguistics.

2 Material and Method

2.1 Material

The material comprises 10 pairs of English texts collected from 10 native speakers of English, whose gender, age and education were recorded; basic variables which, in the case of a larger sample, may lend an additional perspective to the interpretation of results. The participants were asked to write two fictional letters of complaint to their superior, describing an unhealthy workplace environment. Both letters outlined the same situation, however the second text was written in anonymous mode. The task was described in detail in a form (see Appendix I) which the participants received by email. In it they were introduced to the situation which they were to write about and instructed that they should write the second text differently so as not to be found out as its authors. The recommended length was 300 words per text, nonetheless there is substantial variation among authors. The time limit given was 14 days, based on the assumption that in a real life scenario the authors would be able to write the complaint letter at their leisure.

The acquisition of the texts themselves proved to be much more challenging than had been anticipated, and so it was not possible to make a selection of the most interesting cases of anonymization from a wider range of text samples. For this reason, some of the letters included in this study are substantially shorter compared to others and would have otherwise been discarded in favour of those that were nearer the desired 300-word mark. On the other hand, in a forensic setting, there is often very little more to build a textual analysis on and therefore even the examination of extremely short texts can be meaningful.

The 20 texts are available in Appendix II and are named A01, A02, B01, B02, etc., where authors are labelled A, B [...] to F. The “x01” texts were written with the assumption that the author will be known, and “x02” texts were written in the anonymous mode. The two texts by one author were compared and, at the conclusion of the analysis, so were the anonymization strategies of the individual authors. Wherever possible, the participants were contacted with several follow-up questions regarding their motivation and the difficulty of the task, which is discussed in Chapter 4.

2.2 Quantitative analysis

Among quantitative methods are those studying sentence length, lexical richness, n-grams and parts of speech. All of these quantitative approaches to authorship analysis mentioned in the introductory chapter were conducted on texts of considerably greater lengths; usually

they were entire bodies of work by one author, or chapters at the very least. For example Johnson and Wright had at their disposal a corpus of 63,000 emails and 2.5 million words written by 176 people, on which they demonstrated the n-gram textbite approach (2014:37). Considering the length of our texts, it was not anticipated that these quantitative measures would yield very significant results. However, for the sake of completeness, attributes such as sentence length were measured, the results briefly summarized and presented as a supplement to the qualitative stylistic analysis.

Regarding the number of sentences and average sentence length, the text was manually divided based on punctuation signalling the end of a sentence (full stop, exclamation mark, question mark), or if the string following a colon or semicolon was judged to be a completely independent sentence, e.g. in the case of text *A01*, where the colon precedes a list of events described by the author. Sentence length in words was calculated by copying the data into Excel, one sentence per cell and applying the following formula:

=IF(LEN(TRIM(A1))=0,0,LEN(TRIM(A1))-LEN(SUBSTITUTE(A1," ",""))+1)

The average sentence length was then obtained simply by dividing the number of words by the number of sentences.

Lexical richness was measured with the aid of the programme Wordsmith (Scott, 2016), which was able to calculate the text's Standardized type to token ratio (STTR). The STTR is essentially an average of TTR values per n words, which enables us to make comparisons between texts of various lengths. Wordsmith allows for the setting of the n boundary to any number between 50 and 1,000, a function which proved extremely helpful during the course of our work. Should we have wished to compare only the two texts of one author, we would have set the STTR basis according to the shorter of the two and compared each pair individually. However, seeing as we were additionally interested in the comparison of all signed (01-type) and anonymous (02-type) texts, it became necessary to select a common basis for the STTR, which logically had to be based on the shortest of the texts. Unfortunately, this happened to be a text of only 59 words, in contrast with the longest text of 466 words. The difficulty with this is the fact that the shorter the given text samples, the higher their apparent TTR, and the higher the probability that their STTR will be similar. This is due to the fact that repetition generally occurs only after a certain point (usually not within the first several words, as is for example in the case of this sentence). Another factor is the genre, which is in this case a letter, and therefore a number of standard

phrases can be expected, such as ‘To whom it may concern,’ and ‘I am writing this letter regarding...’ etc. Nevertheless, a STTR basis of 59 words appeared to be the best possible solution, and the results of this calculation are presented below.

N-grams were extracted in Antconc (Anthony, 2014) with the condition of at least two occurrences of a given n-gram. Whenever a trigram (or, much less frequently, any other larger n-gram) was present, the results in Antconc also included the bigrams which the given n-gram was made up of. For example, if a text contained two instances of the trigram ‘my colleague’s wellbeing,’ there were also two recorded instances of ‘my colleague’s’ and ‘colleague’s wellbeing.’ Such bigrams were evaluated as redundant and therefore not included in the analysis.

2.3 Qualitative analysis

The greater part of this paper is dedicated to a qualitative stylistic analysis of the texts, which will perhaps yield more significant results with regard to the length of our texts, which rarely exceeds 300w and therefore quantification may not be appropriate. Loosely based on Olsson (2009: 145) a list of suspected style features has been compiled in order to guide the stylistic analysis:

punctuation	sentence complexity, coordination and subordination
spelling	spelling variation and/or errors
nominalization	the appearance of any unusual or erroneous forms
lexical choices	unusual or marked lexical choices, idioms and phrases
style / register	formality, departures from the genre
sentence features	sentence length and complexity

Table 1: Style features for stylistic analysis.

3 Results

The results obtained in the course of the analyses are presented below. Firstly, a brief summary of the quantitative analysis is provided, followed by a more comprehensive stylistic analysis of the individual texts. The methods and approaches used in the quantitative analysis are the following: average sentence length, Standardized type-to-token ratio, parts-of-speech distribution, and n-grams.

3.1 Quantitative analysis

As had been previously suspected due to the length of our texts, the results of quantitative analysis are in dire need of support from the qualitative stylistic analysis. However, certain trends can be observed in the anonymized versions, such as the shortening of sentences, and in fact the shortening of entire texts in some cases.

3.1.1 Sentence length

As can be seen below in Table 2, there appears to be a weak tendency towards shortening the anonymous text, observed in 6 out of 10 authors. In some texts, the difference in length could perhaps be considered of consequence, for example *I01* is, at 466 words, more than twice the length of its anonymous counterpart *I02*, which has only 213 words. The discrepancy is even more pronounced in the texts of author *G*, with 58 words in the anonymous text and 176 words, three times as many, in the signed. Elsewhere, for example in the texts by author *H*, the anonymous text is shorter only by two words. On the other hand, author *D*'s anonymous text is actually longer by 64 words.

text ID	no. of words	no. of sentences	average sentence length
A01	321	12	26.75
A02	250	10	25.00
B01	326	13	25.08
B02	230	11	20.91
C01	281	14	20.07
C02	272	13	20.92
D01	386	13	29.69
D02	450	15	30.00
E01	164	10	16.40
E02	88	6	14.67
F01	339	15	22.60
F02	268	18	14.89
G01	176	11	16.00
G02	58	7	8.29
H01	241	12	20.08

H02	243	17	14.29
I01	466	16	29.13
I02	213	14	15.21
J01	336	10	33.60
J02	351	12	29.25

Table 2: Number of words, sentences and average sentence length for each text.

The actual number of sentences serves as a basis for the calculation of the average sentence length, as seen in Table 2. Variation in sentence length appears to be one of the key features of anonymization; 8 of the 10 anonymous texts contain, on average, shorter sentences than their signed counterparts. Once again, some authors (e.g. author I) have changed the sentence length very markedly, whereas others (e.g. author A) have made only barely observable alterations.

3.1.2 Sentence Complexity

The assessment of sentence complexity was achieved by retrieving the number of sentences, clauses, and the identification of basic sentence structure, i.e. main (coordinate) and subordinate. Table 3 provides the comprehensive results:

text	clauses	main clause	subordinate clause	sentences
A01	29	15	14	12
A02	23	12	11	10
B01	29	18	11	13
B02	18	12	6	11
C01	34	18	16	14
C02	34	17	17	13
D01	40	16	24	13
D02	48	21	27	15
E01	20	14	6	10
E02	11	6	5	6
F01	38	23	15	15
F02	33	24	9	18
G01	23	13	10	11
G02	9	8	1	7
H01	26	17	9	12
H02	31	22	9	17
I01	59	29	30	16
I02	26	18	8	14
J01	33	14	19	10
J02	38	17	21	12

Table 3: Number of clauses, main (coordinate) and subordinate clauses, and number of sentences.

The average number of clauses per sentence for every individual text can be found in Table 4, as well as the difference between the signed and anonymous text. Salutations and signatures were exempt from the count. From the absolute numbers we can clearly perceive which of the authors modified sentence complexity the most, and these findings are confirmed by the stylistic analysis in Chapter 3.2. According to these results, Author I demonstrates the most variability in sentence complexity, with the average number of 3.5 clauses per sentence in the signed text, and only 1.9 in the anonymous text. On the other hand, Authors A and D show very little or no variation.

Another point of interest was the proportion of subordinate clauses to coordinate clauses within the texts, which is recorded in the fourth column of Table 4 as a percentage of subordinate clauses to the total number of clauses in the text, with the neighbouring column containing the difference between texts x01 and x02 as an absolute value. These results show that Author G's proportion of subordinate clauses changed most prominently, followed by that of authors I and E. Author A, on the other hand, preserves the same ratio of subordinate to coordinate clauses in both texts.

text	clause:sentence	ABS 01 - 02	% subordinate	ABS 01-02
A01	2.3	0.0	48.3	0.4
A02	2.3		47.8	
B01	2.2	0.6	37.9	4.6
B02	1.6		33.3	
C01	2.1	0.5	47.1	2.9
C02	2.6		50.0	
D01	3.1	0.1	60.0	3.8
D02	2.9		56.3	
E01	2.0	0.2	30.0	15.5
E02	1.8		45.5	
F01	2.5	0.6	39.5	12.2
F02	1.9		27.3	
G01	2.1	0.8	43.5	32.4
G02	1.3		11.1	
H01	2.2	0.3	34.6	5.6
H02	1.8		29.0	
I01	3.5	1.6	50.8	20.1
I02	1.9		30.8	
J01	3.3	0.1	57.6	2.3
J02	3.2		55.3	

Table 4: The average number of clauses per sentence, the difference between the signed and anonymous text as an absolute value, and percentage of clause relations which are subordinate.

3.1.3 Standardized TTR

Following measures of the standardized type-to-token ratio with a basis of 59 words, the results (Table 5) suggest that there is an inclination towards more repetition in the anonymous texts. This is based on the fact that 7 of the 10 authors have a higher level of lexical density in the first, signed texts. This tendency was displayed most markedly by author *H*, whose lexical richness fell from 80.51 to 73.31. However, this relatively prominent departure from the standard deviation (SD= 2,22) may be attributed to the genre of the author's anonymous text, which he produced as a fairy tale, whereas his signed text was a formal letter.

text ID	STTR 59	(word count)
A01	76.27	315
A02	74.15	247
B01	80	335
B02	77.97	231
C01	76.69	282
C02	79.24	274
D01	81.64	371
D02	81.11	450
E01	81.36	166
E02	81.36	90
F01	82.37	344
F02	80.08	283
G01	84.75	180
G02	79.66	59
H01	80.51	241
H02	73.31	244
I01	78.21	464
I02	81.36	212
J01	81.36	336
J02	81.02	352

Table 5: Standardized type-to-token ratio per 59 words.

3.1.4 Parts of speech

The parts of speech distribution of the texts (Table 6) does not do much to support the initial hypothesis that anonymized texts will display a tendency towards denominalization. Contrary to expectations, the pairs in which the most salient differences were observed actually demonstrate a shift in the opposite direction, where in fact 5 of the 10 authors nominalized their anonymous texts and another 3 authors retained the same distribution of nouns to verbs in both texts.

text ID	noun %	verb %	adj. %	total w.	noun	verb	adj.
A01	30.2	17.1	4.7	321	97	55	15
A02	36.4	17.2	4.4	250	91	43	11
B01	23.3	18.7	6.1	326	76	61	20
B02	25.7	14.3	7.8	230	59	33	18
C01	22.1	22.1	5.3	281	62	62	15
C02	21.7	22.4	5.5	272	59	61	15
D01	16.6	21.2	6.2	386	64	82	24
D02	15.8	22.0	6.4	450	71	99	29
E01	17.1	26.2	6.7	164	28	43	11
E02	29.5	17.0	10.2	88	26	15	9
F01	18.0	24.8	7.1	339	61	84	24
F02	18.3	28.4	5.2	268	49	76	14
G01	21.0	23.9	6.3	176	37	42	11
G02	29.3	20.7	8.6	58	17	12	5
H01	22.8	17.0	7.9	241	55	41	19
H02	22.2	17.3	9.9	243	54	42	24
I01	18.0	23.0	6.2	466	84	107	29
I02	19.7	22.1	4.2	213	42	47	9
J01	25.9	16.7	8.0	336	87	56	27
J02	25.6	16.5	9.4	351	90	58	33

Table 6: The distribution of nouns, verbs and adjectives in the texts.

Author	N:V 01	N:V 02	Difference
A	1.8	2.1	0.3
B	1.2	1.8	0.6
C	1	1	0
D	0.8	0.7	0.1
E	0.7	1.7	1
F	0.7	0.6	0.1
G	0.9	1.4	0.5
H	1.3	1.3	0
I	0.8	0.9	0.1
J	1.6	1.6	0

Table 7: The ratio of nouns to verbs in the text and the difference as an absolute value.

The ratio of nouns to verbs, as seen in Table 7, shows texts with a higher proportion of nouns as those with a ratio higher than 1. The difference between the 01-type and 02-type texts is given in absolute values. From these results it can be said that author *E* has achieved the most noticeable discrepancy between his two texts in terms of parts of speech. Text *E01* has a higher proportion of verbs, whereas *E02* is evidently nominalized.

Although the majority of the authors did alter the proportions of verbs, nouns and adjectives, there is no clear tendency towards either nominalization or denominalization.

3.1.5 N-grams

It was observed that the ratios of unique and shared n-grams (i.e. n-grams that were shared by both the signed and anonymous texts for the given author) differed vastly among the individual authors. An example of a high number of matching n-grams can be found in the texts of author *C*, for whom all but two n-grams were identical. The repetition of essentially grammatical bigrams such as “it is” and “I was” is conceivably a more common occurrence as the recurrence of the 4-gram “promised a workplace that.” The repetition of n-grams can also indicate the author’s preference for certain grammatical structures, as was observed in B02, where there were 4 individual occurrences of the tri-gram “of our team” and not a single instance of the analogous “our team’s.”

Author: C

text 01 (signed)		text 02 (anonymous)	
freq.	n-gram	freq.	n-gram
4	it is	3	it is
3	their work	3	their work
2	as i	2	as i
2	because as	2	because as
2	believe in	2	believe in
2	colleagues to	2	colleagues to
2	i feel it is	2	i feel it is
2	i was	2	i was
2	in a	2	in a
2	of their	2	of their
2	promised a workplace that	2	promised a workplace that
2	their confidence suffers	2	their confidence suffers
2	their work i	2	their work i
2	when he	2	when he
	-----	2	for the
	-----	2	i have

Table 8: N-grams in the texts of author *C*. The frequency is always for the given text.

Author *H*, who wrote the anonymous text as a fairy tale, on the contrary produced two texts with absolutely no n-gram congruence. The high degree of repetition in the second text, documented by the existence of 6 instances of the bigram “storm cloud” and 3

instances of the trigram “the cheerful sun,” is attributable to the fairy tale genre, which generally contains simple and repetitive language.

Author: H

text 01		text 02	
freq.	n-gram	freq.	n-gram
3	i have	6	storm cloud
2	but also	3	the storm
2	for the	2	hopefully the
2	health of	2	in the sky
2	his subordinates	2	it is
2	it may	2	lightning and
2	this letter	2	of the
2	to the	2	of them
2	working environment	2	on the
		2	the cheerful sun
		2	the little
		2	the nice
		2	the storm cloud
		2	those below
		2	ugly storm cloud
		2	viewed from
		2	will stay

Table 9: N-grams in the texts of author *H*. The frequency is always for the given text.

The recurrence of the same n-grams signifies that there are identical strings in the two texts. Furthermore, because our search was set for a minimum of two occurrences of the same n-gram, the results tell us which authors have a tendency to repeat themselves even within the scope of one letter, as seen above in author *C*.

What we cannot determine from these results is the actual percentage of duplicated content, due to the fact that we are working with strings which have a minimum of two occurrences in one text, which means that this method cannot facilitate the detection of two texts which are identical but do not contain any recurring strings. Therefore, the lack of identical strings cannot allow us to automatically assume that the texts are different. On the other hand, the recurrence of n-grams strongly suggests that there is a significant relationship between the two texts. Of course, working with a much larger corpus would allow us to identify the n-grams which are characteristic of an individual’s idiolect, whereas at this stage we can basically only assess the degree of repetition. In general, an n-gram based approach is more suitable for larger datasets, e.g. Johnson and Wright, (2014)

	unique 01	unique 02	identical
A	11	3	12
B	13	9	2
C	0	2	14
D	2	8	7
E	4	5	0
F	10	8	1
G	4	0	1
H	7	17	0
I	28	6	1
J	3	5	8

Table 10: The total number of identical and unique n-grams per author.

Of the ten participants, six had two or fewer identical n-grams, suggesting that these authors had attempted to make not only lexical changes, but also changes in grammatical structure. One such example is author *F*, who in text *F01* writes “it is,” “he is,” and “I had,” whereas in *F02* these n-grams are replaced by the contracted variants “it’s” “he’s” and “I’d.”¹ The remaining four authors very obviously retained large portions of their signed letters in their second, anonymized, texts. The number of n-grams is generally related to the length of the text, as the text’s growing size increases the probability of repetition, a point which should be kept in mind.

¹ Contracted forms were evaluated by Antconc as bigrams, which proved useful in cases where use of contracted forms was one of the anonymization strategies.

3.2 *Qualitative stylistic analysis*

This section deals with each author individually, offering a detailed comparison of the signed and anonymous texts.

3.2.1 Author A

The second (anonymous) text closely resembles the first, signifying that it had been modelled on it, with only minor changes made. No spelling or morphological inconsistencies have been identified in either of the texts, and furthermore both the layout and the structure of the argument remain without changes. The salutation *To whom it may concern* remains in both letters, presumably due to its generic nature. Overall, the anonymous letter is much more impersonal, carrying no indication of the writer's relationship with the reader or any person in the letter, thereby emotionally detaching herself from the situation. This effect is achieved, for example, by the omission of the personal pronoun in [1b], weakening the author's link with his team:

[1a] I am also reporting Mr. X to the Employment Ombudsman (A01)

[1b] ... team members will consider filing an official complaint (A02)

Changes can be observed in specific lexical choices, such as the *details of the events* [2b] replacing the *list of the incidents* [2a]. This particular change could be due to the fact that *event* is a semantically broader concept, which might suggest the author's efforts to communicate a neutral and professional opinion. However, this does not show to be consistent in the remainder of the text. Lexical choices notwithstanding, the structure of the sentence in [2b] remains unchanged, and even the phrase *to which I am referring* is retained verbatim, although the construction is not so frequent as to be considered a set phrase and therefore could be a defining feature of the author's style.

[2a] Below is a list of the incidents to which I am referring. (A01)

[2b] Below are the details of the events to which I am referring: (A02)

[3a] ... contributed to an unprofessional and malicious office environment (A01)

[3b] ... created an unsafe and unproductive workplace environment. (A02)

Other lexical variations are observable in [3b], where *unprofessional* and the strong expression *malicious* has been replaced by much more neutral, although not synonymous, terms *unsafe* and *unproductive*. It may be the case that in the anonymous text the author is

appealing to the needs of the reader, i.e. the need of a productive workplace, rather than the needs of the author herself, i.e. to work in an encouraging (not malicious) environment. However, one change which arguably does not support the impartiality of the author's argument is the shift from *contribute* to *create* in [3b], which places the blame solely on Mr. X. as opposed to only making him seem partly responsible for the problems in the workplace. The paragraphs describing the individual incidents remain unchanged but for the one which reports the actions of Mr. X toward the author herself, where the personal pronouns *I* and *my* are replaced by *Ms. D* and *her*, which is understandable given that the author wishes to remain anonymous.

Regarding the syntactic structure of the two letters, the differences are minimal, as was already testified by the quantifying results for sentence complexity.

3.2.2 Author B

The author of texts B01 and B02 has utilized a wide range of anonymization strategies, most noticeably the use of synonyms and synonymous multiword expressions.

The texts contain grammatically and semantically similar constructions which document the author's attempts to communicate similar content while making different lexical choices which are possibly uncharacteristic of his idiolect. This concerns for example the words used to describe the escalation of negative emotions, as recorded below, where the transitive verb *mounted* [4a] is replaced by a synonymous term *raised* [4b]. The objects following the verb are also different in the anonymous text, although they all convey similar notions. Sentence 4a has the adverb *immensely*, which is not included in the anonymized version, possibly because it is secondary information and therefore not considered to be particularly significant. Additionally, it is an emotionally charged adverb, which could be the reason for its omission.

[4a] demands [...] have mounted the stress and tension among our team members immensely (B01)

[4b] Recent events [...] have raised the level of fear, discord, and distrust within our team (B02)

One thought-provoking aspect of the B01 text is that it contains two counts of the conjunction *but*, whereas the more formal *however* is used in the anonymous B02. Another

adjustment in the anonymous text is that it frequently refers to *team members* and only once to *colleagues*, whereas the first text contains only one mention of *team members*, but three instances of *colleague*. The motivation could be to obscure the specific relationship of the author to the other employees, or else it could be another case of anonymization with the help of synonymous expressions. Similar cases include the replacement of *in person* (B01) by *in a private setting* (B02), and that of *get in touch* (B01) by *reach out* (B02), expressions which use very similar imagery to describe the same action or concept.

Although author *B* has taken many pains to anonymize the second text, she appears to have overlooked her own frequent repetition of a number of adjectives. The expressions *deeply regretful*, *deeply upsetting*, *deeply ingrained*, *deeply personal* and *deeply saddened* are all present in the first text. Objectively speaking, this is a very high frequency of occurrence for a text of 326 words. Although the anonymized text contains only two occurrences – *deeply hurtful* and *deeply destructive*, even these two instances could contribute to the author's positive identification, especially if this adjective is one which she uses frequently, which we assume is the case. Another such adjective is *enormous* and its derived adverb *enormously*, which occur twice in B01 and once in B02. Although one occurrence in the anonymous text does not amount to much, if the author used it frequently outside the context of the two letters, it could still be an idiolect marker and a means of identifying the author.

No attempts have been made at changing the spelling, although there were several opportunities, such as *behavior*, which could easily have become the British variant *behaviour* in the second text. Text B02 contains an erroneous article in *an guarded manner*, however this is not repeated in the text and therefore it is unclear whether or not the author made the error deliberately. The author makes occasional use of the dash as punctuation; although in B01 there is only one single dash, whereas in B02 there are two pairs of dashes fulfilling a bracketing function. Once again, we can only speculate as to whether or not there was any deliberation behind this.

The anonymous letter is much more straightforward, dispensing with any niceties that are integral to the first letter. For example, the first paragraph in B01 contains a sentence which only foreshadows the content of the letter [5], and the actual topic is dealt with from the second paragraph onwards. Conversely the anonymous text begins to describe the problem immediately following the greeting, as seen in [4b]. This is also related to the word count, which is lower in the anonymous text.

[5] I'm deeply regretful to be writing you an email of this nature (B01)

As regards the structure of the text as a whole, the author has changed the order in which the information is presented. While B01 first briefly outlines the resultant situation and only then names the culprit, B02 goes straight to the point by stating that the letter concerns the behaviour of the boss, Jeremy.

3.2.3 Author C

The two texts are basically the same, with absolutely no alterations made to the anonymized text, save for the omission of two sentences at the conclusion of C01 and the addition of one sentence at the beginning of C02, which explains why the letter has been written anonymously. Semantically, both of these portions of texts contain mentions of the truth value of the letter's content; *if you feel I am being dishonest* and *information [...] is true*. Furthermore, both texts contain spatial references alluding to the information contained in the letters, i.e. *my above concerns* [6b] and *provided beneath* [6c], and the noun phrase *my employment* is repeated in the anonymous text, rather than one of the possible alternatives such as *my job*.

[6a] I will accept all responsibility, including termination of my employment, if you feel I am being dishonest. (C01)

[6b] It is in your best interest to witness my above concerns. (C01)

[6c] For the sake of my employment I wish to stay anonymous but the following information I have provided beneath is true. (C02)

Another feature of the texts is that the nominal clauses tend to be connected without *that*, as in *if you feel I am being dishonest* [6a] and *information I have provided* in [6c]; changing this would have been another possible means of anonymizing the text. This author has clearly made no attempts to anonymize the text, which is clear because the three paragraphs are identical in their entirety, and the only aspect of the text C02 which could make it anonymous is the lack of signature.

3.2.4 Author D

Author D has, in a way very similar to author C, made only minimal changes to the second text. In fact, there are once again entire paragraphs left without any alterations. The majority of lexical variation lies in the choice of pronouns thanks to which the author can assert his viewpoint, as witnessed previously, for example with author A. Another instance

of this is provided below, where *my work* in [7a] is changed to *the work of a colleague* [7b], and by the same token *in front of colleagues* [7a] is replaced by *in front to me* [7b]. Similarly, the author writes of the boss throwing papers *back in my face* in D01, but *back in peoples faces* (sic) in D02. This continues throughout the letter, for example when *my colleague* (D01) logically becomes *another colleague* (D02) after already having written about herself in the third person.

[7a] On several occasions, he has undermined my work in front of colleagues during a meeting (D01)

[7b] On several occasions, I have witnessed him undermine the work of a colleague in front of me during meetings (D02)

In [7b], there is the addition of the verb *to witness*, which could be a means of strengthening the credibility of the statement, especially due to its legal connotations, or it could simply be a way of varying the sentence. A change from the singular in D01 to plural in D02 is seen in [7b], where *a meeting* suddenly becomes *meetings*. However, the change in number does not seem to have any effect on the statement's meaning, and so it is likely to be another example of attempts at variation.

The majority of the original letter remains in the anonymized version. For instance, the verb *to accost*, although it is much less frequent than some of its synonyms, such as *confront*, is retained in the anonymized text, arguably because the author did not consider this feature to be characteristic of her idiolect. Regarding spelling errors, there are only two instances; the omission of the apostrophe in *peoples faces* (D02), and the omission of 't' in the noun *complains* (D02). However, as there are other occasions in the same text with the correct spelling of *complaint*, there is no reason to suspect that the error was intentional.

Even when there are differences between the two letters, the structure had a tendency to remain very similar, as in [8a] and [8b]. Both sentences begin with the conjunction *as*, in this case introducing clauses of reason, followed by the subjunctive *I would like to (VERB)*. It is conceivable that the author did not realize the similarity between the two sentences, however it is more likely that she simply did not consider the connection to be of any significance.

[8a] As my time with this company is finishing, I would like to state [...] (D01)

[8b] As I am still employed here, [...] I would like to remain anonymous (D02)

A final point must be made regarding the politeness detectable in these two texts. The signed letter is more forceful and direct, as can be seen in [9a]; there is a clear warning and an assertion of the writer's intentions. On the other hand, the anonymized text contains the politeness device *I fear*, which suggests that further action is not the author's preferred option. However, it must be noted that both D01 and D02 variants are in the passive, i.e. *is ignored* and *is done*, which serves to create distance between the author and the intended reader, thus having the effect of being more polite. Furthermore, the subject *we as a group* rather than the singular pronoun *I* adds to the gravity of the situation and could be interpreted as an indirect threat.

[9a] If this letter is ignored, I will be taking this further. (D01)

[9b] I hope something is done about this or I fear that we as a group will take this further. (D02)

3.2.5 Author E

Relative shortness notwithstanding, texts E01 and E02 document a number of anonymization strategies employed by the author. On the most fundamental level, there are several differences in lexical choices, such as *office environment* (E02) replacing *workplace* (E01), and *his employees* (E02) rather than *the people working under him* (E01). These two cases could suggest a tendency towards a more corporate vocabulary in the anonymized text, although a larger text sample would be necessary in order for us to draw any meaningful conclusions.

The anonymous text contains more references to the opinions and beliefs of other people, whereas the signed text mostly consists of the author's own observations. Below in [10a] there is a neutral statement of the facts, whereas in [10b] the passive voice suggests that the author was not necessarily a witness to the incidents. A similar inclination can be observed in [11]; the excerpts from the signed text contain clear statements of the author's own opinion, which is communicated with the aid of the attitude verbs *believe* [11a] and *find* [11b]. Conversely, the anonymous text communicates the opinions of *many people* [11c].

[10a] He has called people "stupid" [...] (E01)

[10b] Some of the phrases that he is reported to have said are "stupid," [...] (E02)

[11a] I also believe that he has been spreading stories [...] (E01)

[11b] I find his behaviour unprofessional and very counterproductive. (E01)

[11c] Many people in the department have issues with his management style. (E02)

The signed text is significantly more polite, which can be clearly seen in the opening lines; *I wanted to write* [12a] as opposed to the more direct *I am writing* [12b]. A similar discrepancy is documented in the final sentences; the signed text contains an indirect request, *I'm writing to ask you*, its politeness further strengthened by *please* [13a]. The anonymous text contains a sentence with an identical discourse function, but it is much more straightforward thanks to its imperative mood, though it still contains the word *please* [13b]. This shift towards more direct expressions is also related to sentence length and complexity (see Chapters 3.1.1 and 3.1.2) Changes have also been made to the sentence structure of [12b], in which *aware* is followed by a nominal content clause, whereas in the signed text *aware* is followed by a prepositional object, and the subsequent relative clause modifies the noun *problems* [12a].

[12a] I wanted to write you to make you aware of some serious problems that I and my coworkers are having with my manager A___ B___. (E01)

[12b] I am writing you this letter anonymously to make you aware that many people in the financial department have issues with A___ B___, a manager in charge of financial statements [...] (E02)

[13a] I'm writing you to [...] ask you to please do something to address his behaviour. (E01)

[13b] Please do something about A___'s behaviour. (E02)

In general, the anonymous text contains fewer elaborations, for example the entire sentence in [14] is reduced to "*lazy cow*" (E02), an item on the list of insults.

[14] Recently, he called a woman working on our team a "lazy cow." (E01)

3.2.6 Author F

In this study of anonymizing strategies, the two texts by author F proved to be among the most valuable for our analysis. There are no perceived attempts at spelling variation, but other than that, there are examples of anonymization on every level. Although it is not entirely consistent, there is a marked tendency for the use of contracted forms in the second text, for example *looks I had been getting* (F01) becomes *looks I'd been getting* (F02). Likewise,

we have all heard (F01) is rewritten as *we've all heard* (F02). The use of contracted forms in F02 has also been documented in the n-grams chapter, and is linked to the overall informality of the anonymized text, when compared to its signed counterpart.

A number of lexical choices are open to discussion, for example in the first text, there are two occurrences of *coworkers*, whereas in the anonymous text there are only the plural pronouns *we* and *us*. Another is the replacement of *I overheard Alice telling Jean* (F01) by *I heard someone tell someone else* (F02), where not only has the verb been changed, the identity of the participants has been concealed.

There are several instances when a lexical item was substituted by a phrase or even an entire clause, such as *publicly* (F01), replaced by the more colloquial *in front of everybody else* (F02). In [15b], the meaning of the adverbs *subtly and insidiously* [15a] is expressed in simple language and by an entire clause: *I don't know how he does it*. Conversely, the phrase *On the other hand* (F01) is replaced by *But* (F02). However, even here the process is actually that of simplification. The shift appears to be conclusively from a formal style towards the colloquial, which is backed by both lexical and grammatical evidence. For example, *we have to devote so much mental energy* (F01) becomes *we have to focus so much* (F02), and *difficult to function* (F01) becomes the more colloquial *hard to function* (F02).

[15a] When he does talk to us in private, it is to spread rumors, which he often does subtly and insidiously. (F01)

[15b] He also spreads rumors. I don't know how he does it, everyone knows they're coming from him, but he does. (F02)

The simplification and shortening of sentences is evident for instance in *He also spreads rumors*. [15b], a simple sentence in the anonymous text, but one which was originally part of a much more complex structure in [15a]. Another instance of a shift towards informal language which can be observed in the syntax of the text, lies in the change from *how horrible a boss he is* [16a] to *how horrible he is* [16b], as the construction “*how + adjective + a + noun + pronoun + verb*” is generally considered to be more formal, much less common than the alternative “*what + a + adjective + noun + pronoun + verb*” (37 as opposed to 234 hits in the BNC) and, according to its distribution the corpus search, slightly outdated².

² The BNC allocates the texts in the corpus to periods, e.g. 1958-1993, based on the publication dates.

[16a] I get the sense that he is gaslighting us to keep us from trusting our shared instincts about how horrible a boss he is, and keeping us scared that we would face retaliation for reporting him. (F01)

[16b] He's making us feel like we can't trust ourselves about how horrible he is and scaring us into not complaining about it. (F02)

Furthermore, the construction *gaslighting us to keep us from trusting our shared instincts* [16a] is drastically simplified to *making us feel like we can't trust ourselves* [16b] in the anonymous text. The verb *gaslighting* denotes a form of psychological manipulation, and has returned zero hits in the BNC, suggesting that it is quite uncommon and that the author correctly replaced it with the paraphrase [16b] in the anonymous text. The phrase *face retaliation* [16a] is found in the first text, whereas the anonymous text contains the less formal *get retaliation* (F02), although not in the corresponding sentence, but in another section of the text.

Even the opening lines indicate dissimilar levels of formality among the two texts, when *bring to your attention* [17a] is replaced by the less formal *tell you* [17b]. Equally interesting are the similarities of portrayal in the two texts; in both cases, the boss is presented as an impersonal entity. In the first text, he is equate to a *reason*, and in the second he is presented as a *problem*, possibly owing to the authors efforts to dehumanize his superior. Furthermore, the boss is separated from the rest of the sentence by a colon or comma, which could be a rhetorical device serving to add dramatic effect.

[17a] I wanted to bring to your attention one of the main reasons for my departure: [boss's name]. (F01)

[17b] I am writing to tell you about a problem we have, [boss's name]. (F02)

Another interesting rhetorical device, serving to intensify the argument and simultaneously to change the style of the letter, is present in the anonymous text. Firstly, it is the isolated adverbial of time, separated from the rest of the sentence with a colon and providing an almost theatrical introduction. Secondly, the subsequent declarative *Alice is "a stupid bitch"* [18b] does not explicitly state the originator of this claim, but nonetheless it is evident from the surrounding context that it was the author's boss. The resultant tone is almost ironic, emphasizing how ridiculous, unjustified and excessive the boss's insults are.

[18a] Today he called Alice "a stupid bitch" for watering the plants with cold water instead of lukewarm water [...] (F01)

[18b] Today: Alice is “a stupid bitch” for watering the plants with cold water. (F02)

Overall, the anonymized text is shorter, more concise, less formal and more direct. This has been documented in Chapter 3.1.1, and is partly due to the fact that the author generally tends to omit additional information which is not indispensable; this is the case of the absence of *lukewarm water* [18a] in the second text [18b].

3.2.7 Author G

Although the texts by Author G are extremely short compared to the majority of those produced by the ten respondents, there are a number of changes made to the second text which can be attributed to the author’s attempts at anonymization. The most striking aspect of the anonymous text is the sentence length and complexity (or lack thereof), as was already touched upon in the quantitative results (Chapter 3.1.1); text G01 has an average sentence length of 14.67 words, which is more than twice that of G02 with its average of 6.44.

The second letter is very impersonal, beginning with the salutation *To the Management Team* (G02) as opposed to *Alex* (G01), disguising any links of the author to the intended reader. As part of the same strategy, the anonymous text contains no pronouns indicating the author’s connections to the people on the team.

There is considerable simplification and exclusion of redundant information in the anonymous text, for example the bluntly put statement in [19b] fulfils the same function as the slightly more elaborate *his actions have convinced me too resign* in [19b]. Furthermore, the anonymous letter does not contain any politeness markers such as in the first text *I would like to give you feedback* [19a]. As for the author’s leaving out presumably superfluous information, one such case is that of [20b]. In the signed text, the author describes rumours about a colleague’s unsatisfactory contribution and subsequently explains the actual situation, leading his explanation with [20a]. The anonymous letter contains nothing on the subject save for the statement in [20b].

[19a] I want to give you feedback on Sean’s interactions with the team, as his actions have convinced me to resign. (G01)

[19b] Sean has pushed his team too far. (G02)

[20a] The truth is quite the opposite, [...] (G01)

[20b] The team performs better than you've been told. (G02)

There is a surprisingly large number of differences between the two letters, and yet both texts contain frequent mentions of *the team*, or *his team*, missing the opportunity to create variation with the use of a synonymous expression, for example *subordinates*, *employees*, *colleagues*, *co-workers* etc.

3.2.8 Author H

The anonymous text produced by author H is possibly the most imaginative in its use of language and in how it works with genre. Most noticeably, it is written as a fairy tale and, as a result of this, the contrast between the two texts is enhanced much more than it could have been had the author written two letters like the other participants. The fact that the second text is a completely different genre allows the author to select alternative means of expression much more freely. More or less every problem or situation is described with the help of imagery and metaphor, for example *junior employees* or *subordinates* (H01) are described quaintly as *little men and women*, *little people* or *the nice people* (H02) *The cheerful sun* (H02) is often mentioned in the text, and is meant to refer to the intended reader of the text, i.e. the boss's superior.

Below is an example of how the author communicates the same message with entirely different linguistic means. The signed letter states that the boss acts very differently in the presence of his superiors, whereas the anonymous text accuses the *ugly storm cloud* of having *two fronts* (H02). How the boss is being perceived by his superiors is very literally transformed into *viewed from above* [21b], while *viewed from below* is a description of the author's perspective, i.e. *when you (the reader) are out of earshot* [21a].

[21a] his actions around you (his superiors) starkly contrast those when you are out of earshot. (H01)

[21b] On one side, viewed from above, it is a nice fluffy white cloud. But viewed from below it is an ugly storm cloud indeed.

The author writes that the storm cloud *rains down humiliation on all those below* (H02), which is expressed in the signed texts as *regularly humiliates them* (his subordinates) (H01). The symbol of *raining* is in keeping with the storm cloud metaphor, but it also denotes that there have been multiple instances of this action, corresponding to the regularity conveyed in the first text. Another example of the use of symbols and metaphors is the observation in the

first text that *employee retention suffers* (H01), paralleled in the anonymous text by statement that people are *looking for places with better weather* (H02), better weather symbolizing a healthier office environment. When the anonymous text communicates that the boss feels his *position is threatened* and describes his efforts to *impede their* (the subordinates') *progress* (H01), it states that the *cloud* wants the people to *stay on the ground*, whilst itself *staying safely high up in the sky* (H02). Finally, the author urges the reader to solve the problem; *remedial action* [22a] is symbolized by a *wind* [22b] which will carry the problem away.

[22a] I implore you to take remedial action swiftly and effectively. (H01)

[22b] The cheerful sun can cause wind that will blow away the bad storm cloud.
(H02)

As with the use of metaphor, the vocabulary in the anonymous text is characteristic of the genre; it is simple and straightforward, as a story for children would be. For example, the warning against *losing key employees* (H01) is paraphrased and becomes the prediction that *the once cheerful people of XXX might look for another place where they are treated nicely* (H02) The otherwise stylistically undesirable adjective *nice*, the related adverb *nicely*, as well as a range of bland adjectives such as *ugly*, *very*, and *bad*, all find their place in the anonymous text on several counts, and the acceptability of this is once again due to the stylistic conventions of the genre. Furthermore, the second letter contains a substantial amount of repetition, which is evident from the n-gram results, and is also one of the features characteristic of the fairy tale genre. The most frequent phrases are *the ugly storm cloud*, *cheerful sun*, and *little people*.

One question which understandably springs to mind is whether such an anonymous submission to the human resources department would actually be taken seriously. A follow-up inquiry revealed that the author would be more than happy to make this an official submission should his workplace situation call for it, and he asserted that he had no misgivings regarding the weight of the letter's argument.

3.2.9 Author I

Author I displays a level of linguistic self-awareness which surpasses that of the majority of our other participants. Just as it was the case with the other texts, those of author I also fail to introduce variation to spelling. The texts contain a number of instances where the author replaces words and phrases with synonymous expressions, for example *I did not give*

him information (I01) is replaced in the anonymous text by *she didn't email him* (I02). These two fragments are interesting for a number of other reasons. Firstly, the pronoun *I* is replaced in the anonymization process by *she*, a logical step which serves to protect the author's identity, and one which most of the other authors have also made. Secondly, the anonymous text contains the contracted form of the verb *did not*, which indicates a lower register. The anonymous text contains one more contraction, in the sentence *I can't work with Jeremy anymore* (I02). However, there is a number of uncontracted forms in the same text, an inconsistency which can most probably be attributed to the author's inattentiveness. It is a fact, however, that the signed text contains no contracted forms, making it much more constant in its formality.

Another example of differences in register can be observed in the lexicon, namely the shift from *modify his behaviour* (I01) to *things will change* (I02) in the second text. What is referred to as *unprofessional conduct* (I01) in the first text becomes *high school gossip* (I02) in an effort to simplify the vocabulary. Comment should be made on the author's choice of prepositions; the first text contains the sentence *I work for Jeremy [...]* (I01), whereas in the second text there is a preference for the preposition *with*, as in *I can't work with Jeremy anymore* (I02). It is a possibility that the author uses this subtle change to cover up his direct connection to Jeremy, thus increasing his chances of remaining in anonymity.

Most noticeably, however, the anonymous text is considerably shorter than the signed text, which is also reflected by the length and complexity of the individual sentences. Take for example sentences [23a] and [23b], which have the same function in the text and express essentially the same meaning. The first letter first provides an explanation as to why the author is writing it; it is *necessary to inform* the reader of the boss's behaviour. The anonymous text, on the other hand, goes straight to the point and describes the source of the problem in direct terms. The vocabulary in [23a] is in keeping with the formal nature of the letter, principally the phrase *unprofessional conduct asserted by [...]*. The use of the word *horror* in the anonymous text is somewhat melodramatic, and combined with the short sentence creates a powerful effect.

[23a] I feel that it is necessary to inform you of several instances of unprofessional conduct asserted by Jeremy which I have noticed as a common occurrence during my time here. (I01)

[23b] He is a horror to work with. (I02)

Another example of one simplification is documented below in [24], where the information from one sentence in the signed text is distributed among two sentences. The *frequent verbal abuse* is dealt with in [24b], where the adverb *frequently* seems to be slightly redundant, due to the fact that the temporal information is repeated in [24c], where it is additionally stated that this happens *at least once a week*. Another simplification lies in the change from *myself or one of my colleagues* [24a] to *our team* [24b].

[24a] Once a week, Jeremy finds a reason to verbally abuse myself or one of my colleagues.

[24b] He verbally abuses our team frequently.

[24c] He finds a reason to do this to each of us at least once a week.

The anonymous text is generally more straightforward, stripped of most politeness devices; its main objective is to accurately and effectively communicate his concerns, and the author's anonymity permits him to address the reader more directly.

3.2.10 Author J

Although the two texts written by author J are among the longest, with word counts of 336 and 351, there is not much to be said about the anonymizing strategies, which are somewhat lacking. The anonymous text is, for the most part, identical to its signed counterpart, and whatever differences there are between the two texts, they generally exist due to the slightly different scenarios under which the author is operating. However, even these differences are minimalized due to the author's inclination towards certain phrases. An example is provided below, through the comparison of one sentence from the anonymous text and two from the signed text. Although the content of two sentences has been condensed into one sentence, all of the basic information and, more significantly, the singular phrases, remain the same; the *unbearable* conditions, the company's *values, vision and character*, and author's belief that those *higher up the chain of command* should be informed of the situation. The author uses very specific idioms, for example *pack it in* is present in both texts, although with an i.p.m. of 0.56 it is not very frequent in the BNC. Furthermore, its distribution suggests that it is colloquial, and therefore not appropriate in official letters such as those written by author J.

[25a] While the conditions of working under Mr. Manboss have grown unbearable to the degree that I have elected to leave this job, I feel I cannot in good conscience pack it in without communicating some of my concerns up the chain of command. (J01)

[25b] Given the values, vision, and character embodied in this company at the corporate level, I imagine that those in leadership would want to be aware when these same are being directly contradicted by a member of middle management. (J01)

[25c] The conditions of working under Mr. Manboss have grown unbearable, and many would pack it in and leave without a moment's hesitation; but as I believe that the values, vision, and character of this company are worth fighting for, I feel that those higher up the chain of command should be aware that these same are being directly contradicted by a member of middle management. (J02)

One stylistic aspect which sets the two letters apart to some degree is the use of punctuation, namely the semicolon. It can be found in [25c], and on two counts in [26b], all three being present in the anonymous text. In contrast to this, the signed letter contains none. Another aspect worth noting the change in tone following the insertion of the semicolons; every semicolon invites a dramatic pause, making the text read almost like an motivational speech. Additionally, the anonymous text contains a summary of the previous statements, introduced by *in short* [26b].

[26a] a place where managers lead by example, not by denigration, and where success is conceived of in positive-sum terms (as opposed to zero-sum). (J01)

[26b] a place where managers lead by example, not by denigration; where success is conceived of in positive-sum terms (as opposed to zero-sum); in short, a place where [...] (J02)

Upon inquiry following the analysis of this text, the author reported that his primary aim was to communicate his concerns as clearly as possible, even should it be at the expense of his own career. For this reason, he did not deem it appropriate to make any significant attempts at disguising his written style.

3.3 Summary

Following the quantitative and stylistic analysis, the results were inserted into Table 11, which we have partly based on that devised by Olsson (2009: 145). The category *lexical choices* subsumes not only individual lexical items, but also phrases and idioms. Under *sentence features* we have included evaluations of complexity, sentence length and clause count. Some of the categories overlap, for example lexical choices may well be motivated by change in style or register. When an anonymous letter displayed a shift towards informal style with the help of colloquial expressions, the value “1” was added to both the *lexical choices* and the *style/register* cells.

The content of Table 11 reveals several key outcomes: firstly, we can clearly see which authors employed the largest amount of anonymization strategies in their texts. Secondly, we can see the levels of language which were most often worked with, and by the same token those which the authors did not, we presume, consider to be relevant to their idiolect.

	A	B	C	D	E	F	G	H	I	J	Total
punctuation	0	1	0	0	0	0	0	0	0	1	2
spelling	0	0	0	0	0	0	0	0	0	0	0
nominalization	0	1	0	0	1	0	1	0	0	0	3
lexical choices	1	1	0	0	1	1	1	1	1	0	7
style/register	0	1	0	1	1	1	1	1	1	0	7
sentence features	1	0	0	0	1	1	1	0	1	0	5
Total	2	4	0	1	4	3	4	2	3	1	

Table 11: Categories active in anonymization by author, with a grand total per category and author.

The most common means of anonymization proved to be manipulation with register, which was attempted, to some extent, by 7 out of 10 participants. Equally popular was the category of lexical items, which usually entailed the selection of synonymous expressions associated with a lower register. Observable differences in sentence length and complexity were found only in 5 out of 10 cases. A tendency towards nominalization (or denominalization), was observed only in 3 of the pairs of texts, which could be due to the fact that the participants generally do not believe the proportion of nouns to verbs to be a significant idiolect marker. However, it was expected that the degree of denominalization in anonymous texts would be greater, due to their tendency towards a less formal register, in comparison to that of the signed texts. Finally, a surprising finding was that the option of changing the text’s spelling was dismissed by all of our ten authors, while punctuation was opted for only in two cases.

Furthermore, the results show that some authors were much more meticulous and creative in their use of anonymization strategies; Authors B, E and G all made changes in 4 of the 6 categories whereas the efforts of others (Author C in particular) were very slight. However, these numbers alone are no indication of how successful the given author was in disguising his idiolect. A case in point is author H, whose attempts at anonymization were subjectively judged to be among the most successful, and yet objectively the two texts differed only in genre (register) and lexical choices.

4 Discussion

At the beginning of the experiment, it was assumed that, in the anonymous texts, the authors would employ less elaborate style features and display a tendency towards informality; strategies which would include the simplification of syntactic structures, the selection of less formal lexis, and the omission of politeness devices. The reasoning behind this was that when producing a letter to his employer, it is likely to be a presentation of what the author considers to be his best writing – learned vocabulary, prescribed phrases, complex structures. It is then logical that when he wishes to disguise his style, i.e. write differently, the option of simplifying and, to various extents, vulgarizing, will be one to come most naturally. In most cases, these expectations were met by the ten participants in this study, namely authors E, F, G, H, I and, to a lesser extent, author B. Apart from turning to the anticipated strategies, some authors chose to be more creative; Author H proved to be particularly innovative, crossing the boundaries of genre with his fairy tale.

The anonymous texts are generally shorter and less elaborate than their signed counterparts, sometimes lacking large pieces of information. This omission of details could be part of the author's anonymization process. Alternatively, it is possible that the given information is not deemed essential by the author, whose primary concern is to communicate clearly and effectively. It stands to reason that when the reader knows the identity of the author, he will be willing to spend more time reading it, and more likely to tolerate superfluous information. For this reason, the author may feel that his anonymous letter should be as concise as possible. This notion links to the matter of politeness, or lack thereof, in the anonymous texts in contrast to the signed ones. It appears to be a tendency to omit politeness and hedging strategies, which are not indispensable from the point of view of the actual content of the message, but only serve to assert the relationship between the author and reader. The straightforward nature of the anonymous letters is evident, sometimes bordering on the dramatic.

It was anticipated that we would also detect the presence of intentional spelling/morphology errors, which did not prove to be the case. This finding was surprising, as we consider spelling to be among one of the most basic and manifest aspects of language, easily accessible to language users. A possible explanation would be that the authors felt the need to be taken seriously, and that spelling errors would impede this goal,

and therefore they may have been aware of the option after all, but they simply chose not to pursue it.

There were several authors whose anonymous texts were virtually identical to their original, signed counterparts, with only minor alterations made. This was observed to be the case in the texts of authors A, C, D and J. The anonymization strategies employed in such texts were only of the most basic kind; the names of the people in the letter were replaced with less definite expressions, e.g. *someone* / *someone else*, or conversely the pronouns representing the author were replaced with names, e.g. *Ms. D*. Naturally, the signature was omitted, as was, in some letters, the salutation, but otherwise there was very little to distinguish the two letters from one another. There are a number of possible explanations for this, the motivation and the language aptitude of the authors the most prominent among them. Motivation was one of the reasons behind the letter of author J, who, upon inquiry, asserted that he considered the clarity and veracity of his statement to be of greater importance than his anonymity, which is why he made only minimal changes to his second, anonymous text. Judging by the lack of anonymizing strategies in some of the letters, it appears that a number of participants do not have an active awareness of their idiolect, a suspicion which was confirmed by participant C, who said in a follow-up interview said that she was not conscious of any aspect of her written style that would be recognizable by anyone who knew her. Similar opinions were expressed by a number of others.

The results of the analysis have shown that the degree of anonymization varies dramatically from one author to another, and there are a number of factors to which this may be attributed. Firstly, let us contemplate the motivation of the participants; generally, those who later claimed to have enjoyed the task (namely authors H and I) were more thorough, resourceful and more creative in the anonymization of their letter. It is likely that had there been some sort of reward for the most successful attempt at anonymization, the performance of the authors could have been better. However, because the texts were produced owing to the generosity of the participants, with no promise of financial retribution, it must be assumed that the motivation of some participants could have been greater. Another aspect is the time available to the authors for the completion of the task. It stands to reason that the more successful individuals will also have spent more time with the task, seeking to make improvements and alterations wherever possible.

Furthermore, it was expected that the participants' level and field of education would somehow reflect their language proficiency and metacognition, and would correlate with how successful they were at disguising their idiolect. The assumption was that someone with a university degree in language teaching would display much higher levels of linguistic self-awareness than, for example, someone with a degree in engineering. This proved to be a completely erroneous preconception. As suggested by the results of this study, one's field of education has little or no effect on metacognition. Among our most resourceful participants was author H, with his engineering degree, who produced what was arguably the most inventive text. A complete overview of the level of education of the participants is available in Table 12.

	level	field
Author A	Bachelor	Social Sciences
Author B	Bachelor	Theatre
Author C	Secondary Education	N/A
Author D	Masters	Education
Author E	Bachelor	Mathematics
Author F	Bachelor	Linguistics
Author G	Bachelor	Information Technology
Author H	Masters	Engineering
Author I	Bachelors	Humanities
Author J	Bachelors	Humanities and Languages

Table 12: The education of the authors.

The extent of the author's own experience with reading and writing is another variable worth assessing, because it can be expected that a person who reads a large amount and a great variety of literature, or a person who writes regularly, could have clearer ideas about language and style than someone who reads only rarely. Even more problematic for objective assessment is the question of creativity and imagination, which undoubtedly plays a role in the selection and implementation of anonymization strategies, nonetheless to our knowledge there are no objective means of measuring it.

As previously stated, there appears to be a wide array of factors which it would have been advisable to take into account, all of which have an influence on the individual's linguistic self-awareness and introspection (Hoey, 2005: 181). At the beginning of the experiment,

we collected only the most basic variables; age, gender and education. However, it appears that none of these have any direct connection with the participant's performance. More interesting for the interpretation of results would be variables such as current mood, mental and physical state, motivation, creativity and experience working with the written word, all of which have the potential to significantly influence the outcome of the experiment and would therefore be worthy of a more complex assessment. Unfortunately, these are much more difficult to monitor and measure. Another problematic area related to the collection of data represents the actual selection of the texts for further analysis. Owing to the fact that there was a limited number of resources available, we decided to include all ten authors who provided the data, even though several of the texts were extremely short; the anonymous letter of author G was as short as 58 words, which departs from the set minimum of 300 words considerably. Fortunately, even the 58-word text contained a number of examples of anonymization strategies. Nevertheless, had there been a greater quantity to choose from, texts departing from the designated word count would have, in all probability, been discarded.

If the present experiment were to be repeated, there is a number of points which would be wise to reconsider. Although the assessment of the 20 texts showed certain trends and allowed us to make some tentative generalizations, arguably the final results would be more telling if the analyzed sample were larger. Ideally, the group should contain a more varied group of English native speakers, as the participants in the present study are all between the age of 20 and 30. Regarding the design of the task itself, several possible alterations come to mind; for example, it could prove useful to have a control group of participants who would write only one anonymous letter each, and it would be interesting to search for the presence of common denominators in the style of anonymous letters. Another variation could be a design in which the participants would be asked to describe a video, as opposed to following written instructions, which should minimize the interference of lexical items from the task description.

Another possible expansion of the present study would include the addition of learners of English as a second group of participants, which could show if there are any differences in how English is understood by its learners and native speakers. The task itself should remain very much the same as in the present study; the fact that the two texts are written to describe exactly the same situation eliminates the effect of genre and allows us to make bolder conclusions; generally, language users have a specific vocabulary for dealing with

particular situations, and therefore it is with greater certainty that we can make judgements as to which aspects of language the author has altered. The method for analysis would remain the same as in the present study, although it would be advisable to base the results of the stylistic analysis on the individual analyses of several linguists, to minimize the effect of subjective judgements. Furthermore, in the case of a larger quantity of data, a more detailed investigation of n-grams would become possible.

Another point of interest is the collection of data about the participants themselves. As discussed earlier, there are several factors which could correlate with the success rates of the participants and ultimately influence the outcome of the experiment, among the most interesting there is motivation (further discussed below), current mental/physical state, and creativity. Information concerning the number of languages actively used by the participant, as well as his reading and writing habits, could also be of value, because we intuitively believe that an individual's experience with and active use of language is positively correlated with his ability to make conscious linguistic choices. Further studies could determine the truth value of this notion. Were this experiment to be repeated, it would be advisable to conduct a thorough examination of the variables mentioned, a step which would no doubt facilitate a much more complex interpretation of the results.

Regarding the motivation of the participants, the conductor of any future experiments should ideally provide some form of remuneration in order to ensure that the anonymized text produced are genuinely the best of which the authors are capable. It is conceivable that the author whose anonymous text cannot be traced back to him would be promised some sort of reward. Ideally, if the experiment was funded by a research grant, a sum of money for the most successful author could be sufficiently motivating. Furthermore, the conditions under which the participants write the text should be unified, although it is not clear to what extent the resulting anonymous texts would be influenced by such changes. Arguably, it is more natural to write an anonymous text at home, rather than in an artificial, controlled setting, and so the results are actually more true to life when the data is obtained in the same manner as it was in this experiment. The time lapse between when the first and second letters are written should also be subject to consideration; in the present study, there was no time limit for the completion of the task, and the participants were able to place such a time interval between the writing of the two texts as they saw fit. Were the time interval to be set to a week, there would be a much lesser chance of the first text interfering with the second one. This also brings about the question of whether the

participants should have access to their first text when writing the anonymous version. In the present study they did, and indeed this resulted in many of the anonymous letters being simply a revised version of their first text. Although this setup is less authentic, it has allowed for an easier identification of the differences between the two texts.

5 Conclusion

The purpose of this thesis was to identify those specific aspects of written style which native speakers of English modify when attempting to anonymize their texts. The conclusions are based on the analysis of 20 texts by 10 authors, all of whom are native speakers of English. Two texts dealing with the same topic were produced by each participant; one was written as an official letter of complaint, and the other was written as an anonymous letter. The methods used are the same as those used in the investigations of forensic linguists, with the bulk of the results grounded on a qualitative stylistic analysis of the individual texts, and with merely a brief supplementation by quantitative methods, owing to the size of the samples analyzed.

A comparison of the signed and anonymous texts, done separately for each author, demonstrated the complexity of the process of anonymization, in that the authors modified their texts to various degrees, and resorted to different means. One of the key causes of this disparity is the fact that each of the participants possesses a particular level of linguistic self-awareness. This fact was confirmed by the participants themselves, and supports our initial hypothesis. Nevertheless, based on the outcomes of both the qualitative and quantitative analyses, it is possible to make a number of generalizations.

The findings were in line with our expectations that there would be levels of language which do not come into play in the process of anonymization. However, it was anticipated that spelling and punctuation would be the areas most prone to modification, which did not prove to be the case with any of the 10 authors. On the other hand, the majority of authors made different lexical choices in their anonymous texts, most of which corresponded with a modification of the register, the shift invariably being towards less formal means of expression. Another popular strategy was the shortening and overall simplification of sentences, although this was observed only in 5 of the 10 authors.

Among the principle limitations of this experiment we may count the inadequate quantity of data, which does not facilitate further exploration of quantitative methods, and more importantly does not allow for overly conclusive generalizations. Another concern is with the reliability of the available samples, as one can never be entirely certain that the authors anonymized the texts to the maximum of their abilities, and that they were as thorough as they would have been in a real life situation. Any further studies would be advised to

provide sufficient motivation for the participants in order to ensure the best possible quality of texts.

We have briefly discussed the merits of taking into account the effects of variables such as creativity, reading and writing habits, current mood, and motivation. This is a subject worth studying in more detail, ideally with a significantly larger corpus. In further studies, it is recommended that the minimum length of texts is strictly adhered to, a measure which would call for a larger quantity of data from which the selection can be made. Overall, the results suggest that a more careful selection of texts would be required for obtaining conclusive results. It is hoped that the results of the analysis and the questions raised regarding author variables could contribute to further discussion about factors influencing the selection and modification of specific areas of language in the anonymization process.

6 References

- Allen, J.R. (1974) Methods of Author Identification through Stylistic Analysis. *The French Review*, Vol. 47, No. 5. 904-916
- Anthony, L. (2014) AntConc (Version 3.4.3) [Computer Software]. Tokyo, Japan: Waseda University. Available from <http://www.laurenceanthony.net/>
- Barlow, M. (2010) Individual usage: a corpus-based study of idiolects. *Proceedings of LAUD Conference*. Landau, Germany, 1-22
- Biber, D., and S. Conrad (2009) *Register, genre, and style*. Cambridge: Cambridge University Press.
- Bourdieu, P. (1991) *Language and Symbolic Power*. Cambridge: Polity Press.
- Coulthard, M. (2004) Author identification, idiolect, and linguistic uniqueness. *Applied Linguistics*, 24 (4), 431–447.
- Coulthard, M., and A. Johnson.(2007) *An Introduction to Forensic Linguistics: Language in Evidence*. London: Routledge.
- Coupland, N. (2007) *Style: Language Variation and Identity*. Cambridge: Cambridge University Press.
- Crystal, D and D. Davy (1969) *Investigating English Style*. London: Longmans, Green and Co Ltd.
- Davies, G., and A. R. Beech (2012) *Forensic psychology: Crime, justice, law, interventions*. Chichester, West Sussex: John Wiley and Sons.
- Hanique, I, M. Ernestus, and L. Boves (2015) Choice and pronunciation of words: Individual differences within a homogeneous group of speakers. *De Gruyter*, 1-25.
- Hoey, M. (2005) *Lexical Priming: A new theory of words and language*. London: Routledge.
- Johnson, A and D. Wright (2014) Identifying idiolect in forensic authorship attribution: an n-gram textbite approach. *Language and Law / Linguagem e Direito*, Vol. 1(1), 37-69
- Jones, S. (2010) “Using Web Data to Explore Lexico-Semantic Relations” In *Lexical-Semantic Relations: Theoretical and practical perspectives*, Vol. 28, 49-67.

- Love, H. (2010) *Attributing authorship: An introduction*. Milton Keynes, UK: Lightning Source UK Ltd.
- Masthoff, H. (1996) A report on a voice disguise experiment. *International Journal of Speech Language and the Law*, Vol 3, No 1, 160-167
- Olsson, J., and J. Luchjenbroers (2014) *Forensic linguistics*. London: Bloomsbury.
- Olsson, J. (2009) *Wordcrime: Solving crime through forensic linguistics*. London: Continuum International Publishing Group.
- Scott, M. (2016) *WordSmith Tools version 7*. Stroud: Lexical Analysis Software.
- Solan, L. (2013) Intuition versus Algorithm: The Case of Forensic Authorship Attribution. *21 J. L. and Pol'y*, 551-576.
- The British National Corpus, version 2 (BNC World). Distributed by Oxford University Computing Services on behalf of the BNC Consortium. Ústav Českého národního korpusu FF UK, Praha 2001. Available from <http://www.korpus.cz>
- Tweedie, F. J. and R. H. Baayen (1998) How Variable May a Constant be? Measures of Lexical Richness in Perspective. *Computers and the Humanities* 32. 323–352
- Van Gijssel, S. et al. (2005) A variationist, corpus linguistic analysis of lexical richness. *Quantitative Lexicology and Variational Linguistics*.
- Zetterholm, E. (2003) The same but different—three impersonators imitate the same target voices. *Proceedings of the 15th International Congress of Phonetic Sciences: Barcelona, 3-9 August 2003*. Vol. 1/3. 2205-2208.

7 Resumé

Výchozím bodem forenzní lingvistiky je teoretické stanovisko, že každý rodilý mluvčí disponuje svou vlastní jedinečnou podobou jazyka. (Coulthard, 2004: 31). Idiosynkratický osobní dialekt, nebo-li způsob vyjádřování (styl) jednoho daného mluvčího, se označuje za idiolekt, a představuje způsob, jakým je jazyk používán v interakcích. Výsledkem je styl, který je pro daného mluvčího jedinečný. Zatímco sociální či regionální dialekty jsou sdíleny větší skupinou uživatelů jazyka, idiolekt je stejně tak unikátní, jako je jazyková zkušenost jednotlivce, a v závislosti na velkém množství faktorů se neustále vyvíjí a mění. Texty vyprodukované různými lidmi nikdy nebudou naprosto shodné, což nelze připisovat pouze velkému množství všech možných kombinací slov, ale i vlivu jazykové zkušenosti mluvčího.

Naprosto zjevné a snadno pozorovatelné rozdíly se s největší pravděpodobností budou vyskytovat na úrovni pravopisu a interpunkce, avšak největší variabilitu očekáváme na úrovni lexikálních jednotek. Lze předpokládat, že každý text obsahuje jedinečnou kombinaci lexikálních jednotek a frazémů, jejichž volba není náhodná. Dokonce na úrovni vět a v otázce jejich členění má každý autor jisté návyky, které ho mohou odlišit od ostatních. Každý z těchto aspektů dodává jakémukoli projevu, ať psanému či mluvenému, ojedinělou podobu.

Při provádění forenzní analýzy textu existuje řada způsobů, kterými lze k analýze přistupovat. Zejména u delších textů se často provádí analýza na základě kvantifikačních metod, ze kterých byly v této práci použity měření délky slov a vět, poměry slovních druhů, TTR, neboli Type to Token Ratio (poměr unikátních výskytů slov ku celkovému počtu slov), popřípadě standardizovaná hodnota STTR. Dále bylo provedeno vyhodnocení n-gramů, tedy opakujících se víceslovních spojení. Stěžejní částí této práce byla podrobná stylistická analýza, ve které byly zohledněny všechny jazykové roviny; pravopis, interpunkce, slovní zásoba, větná skladba, stylistické aspekty (spisovnost, formálnost) atd.

Každý jednotlivý aspekt textu může mít funkci při určování autorství. Přestože idiosynkratické zvyky v oblastech pravopisu a interpunkce jsou na první pohled velmi zřetelné a mohou být nápomocné při odhalení identity autora, možnost variace zde není nijak výrazná. Množství dostupných možností při lexikální volbě z ní činí významnou složku idiolektu. Míra slovní zásoby, tendence k užívání určitých slov a jejich kolokace jsou součástí idiolektu. Dále stojí za povšimnutí složitost vět a způsob, kterým jsou spojovány.

Olsson (2009: 145) uvádí výčet několika základních stylistických rysů, které považuje z hlediska forenzní lingvistiky za nejzásadnější. V první řadě zmiňuje výskyt jakýchkoliv

neobvyklých, vzácných či chybných jevů v gramatické struktuře; jako příklad uvádí mylné představy mluvčího o členění slov v anglických výrazech “photo copied” a “back fired.” Významnost takovýchto odchylek závisí na frekvenci jejich výskytu jak v daném textu, tak i v jazyce obecně. Dále Olsson doporučuje do stylistické analýzy zahrnout neobvyklé zvyklosti autora v oblasti interpunkce. Užití neobvyklých idiomů a frází může mít při analýze klíčovou úlohu, stejně tak i výskyty pravopisných chyb, a to především těch méně rozšířených. Rozložení textu či dokumentu je dalším možným předmětem zkoumání, nicméně nelze očekávat variabilitu především vzhledem k faktu, že tento aspekt je často úzce vázán na kulturní normy.

Jedním z cílů této práce bylo zhodnotit, do jaké míry může být psaný projev ovlivněn povědomím mluvčího o jeho vlastním idiolektu. Dalším cílem bylo identifikovat jazykové roviny, ve kterých se vyskytuje nejmenší míra variability; tedy ty jazykové roviny, ve kterých zůstanou znaky autorova idiolektu i poté, co podrobil text anonymizaci.

Očekávaným výstupem byla řada zobecňujících závěrů ohledně anonymizačních strategií uživatelů jazyka, což by mohlo být následně využito v praktických aplikacích lingvistických poznatků, například při určování autorství v oblasti forenzní lingvistiky.

Pro tuto práci bylo pořízeno 20 textů od 10 autorů, z nichž každý byl rodilým mluvčím angličtiny. Účastníci byly nejprve seznámeni se smyšlenou situací na pracovišti, kterou měli poté popsat ve dvou dopisech svým nadřízeným. První z dopisů psali coby oficiální stížnost, pod kterou se na závěr měli podepsat. Druhý dopis měl být anonymní, a účastníkům bylo v instrukcích zdůrazněno, že dopis bude číst někdo, kdo zná styl jejich psaného projevu, a že se z toho důvodu musí snažit svůj styl zamaskovat. Výsledné texty, z nichž každý měl přibližně 0,5 až 1 normostranu, byly následně podrobeny důkladné stylistické analýze, na základě které bylo možné učinit řadu obecnějších závěrů. Na základě Olssonových (2009) stylistických rysů bylo vytvořeno 6 kategorií pro zaznamenávání výsledků; interpunkce, pravopis, nominalizace, lexikální volby, styl/registr a vlastnosti vět (a souvětí). Kvantitativní analýza, která je vhodnější pro práci s rozsáhlejšími texty, byla provedena pouze okrajově.

Nejčastěji využívanou anonymizační strategií byla podle výsledků manipulace s mírou formálnosti, o kterou se do jisté míry pokoušelo sedm z deseti respondentů. Stejně často volili respondenti úpravy na lexikální úrovni, což zahrnovalo převážně volbu synonymních výrazů, které se pojí s méně formálním způsobem vyjadřování. Pozorovatelné rozdíly v délce vět a celkové složitosti konstrukcí se vyskytovaly pouze v pěti případech z deseti. Dále tendence k nominalizaci (popřípadě denominalizaci) byly zaznamenány pouze u tří z deseti respondentů,

což mohlo být díky tomu, že uživatelé jazyka nepřikládají příliš velkou důležitost poměru mezi slovními druhy, zde konkrétně mezi slovesy a podstatnými jmény, pokud se hovoří o znacích idiolektu. Oproti pozorovaným výsledkům se očekávala výrazně vyšší míra denominalizace v anonymních textech, a to především vzhledem k pozorované tendenci mluvčích volit méně formální vyjadřovací prostředky.

Výsledky ukazují, že někteří z autorů předvedli výzarně vyšší míru kreativity a byli ve svých anonymizačních strategiích důslednější, než jiní; autoři B, E a G provedli úpravy ve čtyřech z celkových šesti možných kategorií, zatímco snahy ostatních, zejména pak autora C, byly jen marginální. Nicméně tato čísla sama o sobě nehovoří o míře úspěšnosti, se kterou daný autor maskoval svůj idiolekt. Příkladem je autor H, jehož anonymizační snahy byly hodnoceny jako jedny z nejuspěšnějších, přestože se jeho dva texty podle tabulky šesti kategorií liší pouze v žánru (styl/registr) a v lexikálních volbách.

Prvotní hypotéze odpovídá zjištění, že existují jazykové roviny, které jsou při anonymizaci mluvčím opomíjeny, avšak nepředpokládalo se, že se bude jednat o rovinu pravopisnou. Nicméně ani jeden z deseti autorů se nepokoušel měnit svůj pravopis, a pouze dva z respondentů provedli změny ve svém užívání interpunkce. Většina autorů při anonymizaci pracovala se slovní zásobou, a tyto změny byly často propojeny s posunem registru směrem k neformálnímu způsobu vyjadřování. Další oblíbenou strategií bylo zkracování vět a zjednodušení souvětých celků, nicméně pouze u pěti autorů z deseti.

Za jeden z hlavních nedostatků tohoto výzkumu se považuje nedostačující množství dat, které neumožňuje důkladnější využití kvantitativních metod, a především kvůli kterému nelze učinit jednoznačné závěry. Dále lze polemizovat o spolehlivosti a vypovídací hodnotě dostupných vzorků, neboť není jisté, že každý z autorů skutečně anonymizoval text tak, jak nejlépe uměl, a že by v reálné situaci došel ke stejnému výsledku. Pro budoucí výzkum z tohoto důvodu doporučuje poskytnutí dostatečné motivace pro respondenty, aby byl zajištěna co nejlepší možná kvalita textů.

V závěru práce bylo poukázáno na možné vlivy proměnných, kterými jsou kreativita, zkušenosti autora se čtením a psaním, aktuální nálada, a motivace autora. Vliv těchto proměnných si zaslouží podrobnější zkoumání, a to ideálně za pomoci podstatně většího korpusu. Očekáváním je, že výsledky analýzy spolu s otázkou proměnných povedou k další diskusi o faktorech, které ovlivňují výběr konkrétních jevů a jazykových rovin pro zapojení do procesu anonymizace.

8 Appendices

8.1 Appendix I

The task as it was presented to the participants

Age	
Gender	
Education	
Email Address	

Task 1

Your boss regularly humiliates his subordinates in front of their colleagues, undermines their work and has been known to verbally abuse them. He has been trying to impede their success by spreading rumours. You find working in such an environment unacceptable. Your boss' superiors know nothing about this.

You want to leave your job anyway, and so have decided to write an open letter to your boss' superiors and let them know what is happening. Be sure to provide enough detail (try to write approx. 300 words)

Task 2

You still want to keep your job, and so it is necessary to remain in anonymity. Because your boss may see your letter and he knows how you write, it is important that you write in such a way that the letter doesn't sound like you wrote it. (approx. 300 words).

(The situation is the same as for the previous task: Your boss regularly humiliates his subordinates in front of their colleagues, undermines their work and has been known to verbally abuse them. He has been trying to impede their success by spreading rumours. You find working in such an environment unacceptable. Your boss' superiors know nothing about this.)

8.2 Appendix II

Author A, task 1

To whom it may concern,

As an employee of this firm, I feel it is necessary for me to alert you to a situation that has arisen in the workplace, which I believe is tantamount to bullying, harassment and abuse of power. My superior, Mr. X, has, for several months, verbally abused other team members and contributed to an unprofessional and malicious office environment by spreading rumours about his colleagues in order to undermine them. Below is a list of the incidents to which I am referring. I hope this letter will be given due consideration. Please note that I am also reporting Mr. X to the Employment Ombudsman in order to ensure that his behaviour will be dealt with accordingly.

- On August 14, 2015, Mr. X called Ms. F a “bitch” when she arrived late to a meeting, despite the fact that her tardiness was due to a meeting with a client, of which she advised us in advance.
- On August 31, 2015, Mr. X told Mr. G that Ms. Y had not contributed to the team’s advertising pitch to the national board. Ms. Y, as head of the Marketing Department, had, of course, designed and distributed the advertising material for the pitch meeting.
- On September 2, 2015, Mr. X called Mr. S a “faggot” and told the Sales Team that Mr. S was a homosexual. This is a serious breach of the company’s anti-harassment code, which states that an employee’s sexuality is entirely confidential and that homophobia will not be tolerated.
- On September 13, 2015, Mr. X told Human Resources that I was pregnant and that they should be looking for a replacement to fill my role.

I am available for comment on this matter via email at janedoe@company.com or via 0400 333 666.

I look forward to seeing this matter resolved for the benefit of my colleagues and for the posterity of the company.

Yours sincerely,

[Name]

Author A, task 2

To whom it may concern,

Please consider this an anonymous submission to the Human Resources Department regarding the behaviour of Mr. X, Manager and Director at Inc. Company.

Mr. X has made inappropriate comments in the workplace and verbally abused his colleagues, which I believe has created an unsafe and unproductive workplace environment. Below are the details of the events to which I am referring:

- On August 14, 2015, Mr. X called Ms. F a “bitch” when she arrived late to a meeting, despite the fact that her tardiness was due to a meeting with a client, of which she advised us in advance.
- On August 31, 2015, Mr. X told Mr. G that Ms. Y had not contributed to the team’s advertising pitch to the national board. Ms. Y, as head of the Marketing Department, had, of course, designed and distributed the advertising material for the pitch meeting.
- On September 2, 2015, Mr. X called Mr. S a “faggot” and told the Sales Team that Mr. S was a homosexual. This is a serious breach of the company’s anti-harassment code, which states that an employee’s sexuality is entirely confidential and that homophobia will not be tolerated.
- On September 13, 2015, Mr. X told Human Resources that Ms.D was pregnant and that they should be looking for a replacement to fill her role.

Should Mr. X’s behaviour continue to create discomfort and lack of cohesion within Inc. Company, team members will consider filing an official complaint with the Employment Ombudsman.

Author B, task 1

Dear Frank and Laura,

I hope this finds you well. I'm deeply regretful to be writing you an email of this nature—my time at Porjarová Innovations has been something of a challenge, but I have greatly appreciated the opportunities the two of you have afforded me. It pains me to leave the company under circumstances such as these.

I'm writing in regards to Jeremy, whose behavior in recent months has been deeply upsetting to my colleagues and to me. I'd like you to know a little bit about the environment that he has created within our team at PI. While my experience in glass manufacturing is limited, I have never worked in an office in which belittlement and abuse are such a deeply ingrained facet of the workplace.

There is no doubt in my mind that Jeremy is under enormous pressure, and certainly the demands placed on our team by the Okkerville clients in the past several month have mounted the stress and tension among our team members immensely. But it's enormously disheartening to have the real concerns of my colleagues met with ire and mockery, nearly each day we come to the office. Jeremy's behavior has taken me very much by surprise; his argumentative nature has grown provocative and deeply personal, and his unwillingness to listen to and collaborate with his own employees creates a stalemate in the office, ultimately acting against the interests of PI.

I would appreciate the opportunity to discuss this with you in person and hope that you won't hesitate to get in touch. You all and PI have been very good to me, and I'm deeply saddened to be leaving the company at a moment of so much internal unrest. If it will be of any help to the colleagues I'm leaving behind, I'd very much like to continue this conversation with you in person.

Thank you very much for your support these past two years.

Regards,

[Name]

[telephone]

Author B, task 2

Dear Frank and Laura,

Recent events in our workplace have raised the level of fear, discord, and distrust within our team, and I believe they should come to your attention. I am particularly concerned about the behavior of Jeremy Blackman, our team leader.

Following the loss of the Okkerville client, all members of our team faced enormous disappointment. However, Jeremy's verbal abuse towards many of our team members—and two in particular, who were responsible for liaising directly with the client—came as a surprise to me and to the group as a whole.

Furthermore, Jeremy has been responsible for spreading personal information about the individuals on our team throughout the office that is in no way related to our work. Much of this information is untrue and deeply hurtful. My colleagues and I have found this to be deeply destructive, as many members of our team approach group meetings reluctantly and in an guarded manner.

This behavior is not only unprofessional: it is also distracting. I truly believe that a serious discussion with Jeremy would be in the best interests of PI as a company, as our productivity has faltered with our morale.

I hope you will reach out to the members of our team to speak more directly—and in a private setting—about the events of the past two months.

Thank you very much for your attention.

Author C, task 1

To whom it may concern,

I applied for the job because as advertised, I was promised a workplace that encourages ALL colleagues to respect one another, including management. I was also promised a workplace that allows colleagues to learn from their mistakes. To fulfil my duties as a current member of this company, I feel it is my responsibility to inform you of the behaviour that is being carried out by our boss John Smith.

Although I have not yet been on the receiving end of John's verbal abuse, I do not appreciate his attitude towards my colleagues. It makes me extremely uncomfortable as I feel it is unjust. John always comments that their work is of poor sort. I can see how hard they work to impress him and when he disregards their work, I see their confidence suffers which then effects the quality of their work.

I also believe in decency. There are times when staff members should be pulled aside by management to discuss their behaviour or work ethic which should take place in a private area and spoken in a constructive manor. I don't believe in publicly humiliating staff in front of their colleagues because as I previously said, their confidence suffers. On a daily basis John is continuing to humiliate staff when he is wanting to discipline them.

I ask of you to review *Boss* his behaviour as it is not only affecting the company but also having a negative impact on all staff. I will accept all responsibility, including termination of my employment, if you feel I am being dishonest. It is in your best interest to witness my above concerns.

Kind regards, [Name].

Author C, task 2

To whom it may concern,

For the sake of my employment I wish to stay anonymous but the following information I have provided beneath is true.

I applied for the job because as advertised, I was promised a workplace that encourages ALL colleagues to respect one another, including management. I was also promised a workplace that allows colleagues to learn from their mistakes. To fulfil my duties as a current member of this company, I feel it is my responsibility to inform you of the behaviour that is being carried out by our boss John Smith.

Although I have not yet been on the receiving end of John's verbal abuse, I do not appreciate his attitude towards my colleagues. It makes me extremely uncomfortable as I feel it is unjust. John always comments that their work is of poor sort. I can see how hard they work to impress him and when he disregards their work, I see their confidence suffers which then effects the quality of their work.

I also believe in decency. There are times when staff members should be pulled aside by management to discuss their behaviour or work ethic which should take place in a private area and spoken in a constructive manor. I don't believe in publicly humiliating staff in front of their colleagues because as I previously said, their confidence suffers. On a daily basis John is continuing to humiliate staff when he is wanting to discipline them.

I ask of you to review John Smiths's behaviour as it is not only affecting the company but also having a negative impact on all staff.

Kind regards.

Author D, task 1

To whom it may concern,

I am writing to discuss with you the behaviour of my superior which I believe to be inappropriate, demeaning and abusive. On several occasions, he has undermined my work in front of colleagues during a meeting, in a manner that was neither constructive nor conducive to a healthy workplace environment. He yells and screams, throws papers back in my face and verbally accosts me almost to the point of tears.

On one occasion, my colleague was verbally abused via an email sent to the whole office, outlining how incompetent they were at the task presented to them and a warning was issued via that same email that indicated if we were to make the same mistakes, we could 'kiss our jobs goodbye.'

He has also spread illicit and potentially career ending rumours about several of my other female colleagues, suggesting that the only way they could possibly afford the clothes they wear, or the car they drive, is through prostitution or 'sleeping' their way to the top. These rumours have resulted in two of them being demoted and one missing out on a promotion, which he later received.

As my time with this company is finishing, I would like to state openly how appalling it is, that in this day in age a man is continually allowed to behave in such an unprofessional and bullying manner and still not only be a leader, but be employed at all. If this letter is ignored, I will be taking this further. I have several written samples of evidence, emails and witnesses who would be happy to verify my complaint.

While it is my understanding that you may not have any indications of his behaviour, I am hoping that you will see us as victims who are finally brave enough to step forward and be honest about our work conditions. I hope that you can resolve this matter quickly and harshly and provide the support a company of this magnitude should. I hope you will also put parameters in place where your future supervisors are concerned, so that this never happens again.

If you wish to see any forms of the evidence I have aforementioned, please email me.

Regards,

Author D, task 2

To whom it may concern,

I am writing to discuss with you the behaviour of my superior which I believe to be inappropriate, demeaning and abusive. On several occasions, I have witnessed him undermine the work of a colleague in front of me during meetings, in a manner that was neither constructive nor conducive to a healthy workplace environment. He yells and screams, throws papers back in peoples faces and verbally accosted my colleague almost to the point of tears.

On one occasion, another colleague was verbally abused via an email sent to the whole office, outlining how incompetent they were at the task presented to them and a warning was issued via that same email that indicated if we were to make the same mistakes, we could 'kiss our jobs goodbye.'

He has also spread illicit and potentially career ending rumours about several of my other female colleagues, suggesting that the only way they could possibly afford the clothes they wear, or the car they drive, is through prostitution or 'sleeping' their way to the top. These rumours have resulted in two of them being demoted and one missing out on a promotion, which he later received.

As I am still employed here, and do enjoy my position despite the toxic environment in which I am forced to work in, I would like to remain anonymous. I would also like to state openly how appalling it is that in this day in age a man is continually allowed to behave in such an unprofessional and bullying manner and still not only be a leader, but be employed at all.

While it is my understanding that you may not have any indications of his behaviour, I am hoping that you will see us as victims who are finally brave enough to step forward and be honest about our work conditions. I hope that you can resolve this matter quickly and harshly and provide the support a company of this magnitude should. I hope you will also put parameters in place where your future supervisors are concerned, so that this never happens again. I am certain that if you can protect our identities, we can gather and provide several written samples of evidence, emails and witnesses who would be happy to verify this complaint and lodge many others.

I cannot suggest a way in which you could address this situation, but perhaps it would be best to employ a contact officer or someone who we can lodge these complains with anonymously. I hope something is done about this or I fear that we as a group will take this further. How it is handled, is not up to you.

Regards,

Anonymous.

Author E, task 1

Hello Mr. X___,

I wanted to write you to make you aware of some serious problems that I and my coworkers are having with my manager A___ B___. A___'s actions have made our workplace an uncomfortable place to work, and we are now completely fed up with his behaviour. He has verbally abused the people working under him and humiliates them in front of others. He has called people "stupid" or "an idiot" on many occasions. He also becomes angry at berates us about our work. Recently, he called a woman working on our team a "lazy cow."

I also believe that he has been spreading stories about our personal lives that are meant to embarrass us. I find his behaviour unprofessional and very counterproductive.

I'm writing you to let you know that this is going on, and to ask you to please do something to address his behaviour. Please feel free to contact me if you wish to discuss this issue in person.

Author E, task 2

Hello Mr. X___,

I am writing you this letter anonymously to make you aware that many people in the financial department have issues with A___ B___, a manager in charge of financial statements who reports directly to you.

A___ B___'s behaviour is unprofessional and inappropriate for an office environment. He regularly humiliates his employees. Some of the phrases that he is reported to have said are "stupid," "idiot," and "lazy cow." Many people in the department have issues with his management style.

Please do something about A___'s behaviour.

Author F, task 1

Dear [boss's superiors],

I am going to be leaving soon, and I wanted to bring to your attention one of the main reasons for my departure: [boss's name]. Working under him has become unacceptable, and while I have made my decision, it is important to me that I at least try and make the work environment better for my remaining coworkers.

Almost every single day, sitting at my desk, I suddenly hear screaming from a desk nearby. [Boss] is standing over a coworker's chair, yelling at them publicly, for minor or nonexistent infractions. Today he called Alice "a stupid bitch" for watering the plants with cold water instead of lukewarm water, and yesterday, he asked Rajiv, "How fucking stupid can you be, you Paki fuck?" for taking a long time to find a certain file. He's an abusive sexist and racist and we all have to devote so much mental energy towards futile efforts to avoid his wrath that our work is suffering. It is difficult not to take personally, and makes it very difficult to function over the course of a day.

When he does talk to us in private, it is to spread rumors, which he often does subtly and insidiously. Last week I overheard Alice telling Jean that I had supposedly said mean things about Roberto. I was horrified, I would never do something like that! On the other hand, the dirty looks I had been getting in the coffee room suddenly made sense. While we have all been the victims of these rumors that mysteriously pop up, and have all heard them straight from the boss's mouth, they still manage to spread. I get the sense that he is gaslighting us to keep us from trusting our shared instincts about how horrible a boss he is, and keeping us scared that we would face retaliation for reporting him. Since I am no longer worried about that, I must do what I can to help my coworkers.

Thank you for taking the time to consider this.

Author F, task 2

Dear [boss's superiors],

I am writing to tell you about a problem we have, [boss's name]. It's gotten to the point that he's impossible to work for, and I have to say something about it.

Every single day sitting at my desk you can hear screaming from somewhere nearby. It's [Boss] yelling at someone for no reason at all, in front of everybody else. Today: Alice is "a stupid bitch" for watering the plants with cold water. Yesterday: he asked Rajiv "How fucking stupid can you be, you Paki fuck?" He was taking a long time to find a file on his computer. I hear all this sexist and racist and abusive stuff, and our work is suffering because we have to focus so much on not getting him to yell at us. It's hard not to take personally and makes it really hard to function.

He also spreads rumors. I don't know how he does it, everyone knows they're coming from him, but he does. Last week I heard someone tell someone else about mean things I had said to a third person. I would never do something like that! But now I understood all the dirty looks I'd been getting in the coffee room. We've all heard him spread these rumors, but they keep spreading. He's making us feel like we can't trust ourselves about how horrible he is and scaring us into not complaining about it. I just hope this doesn't get traced back to me, or else I'd get retaliation.

Thank you for reading, I hope you'll do something to fix this awful problem.

Author G, task 1

Alex, I want to give you feedback on Sean's interactions with the team, as his actions have convinced me to resign. He often calls each of us out in front of the team about the work we do. Disguising his dissatisfaction behind jokes only makes it more uncomfortable for us.

He also refuses to let us talk to any other managers, particularly those above him like yourself, hence this letter. I suspect he has been taking credit for the work we do. I heard one comment in the kitchen last week suggesting Kathy was lucky to have Sean's mentorship, without which she would never have finished the project on time. The truth is quite the opposite, that Sean's input delayed the project timeline, and caused much stress to all of us. I believe Kathy could have done far better and quicker work had Sean not been involved.

I'm hoping you can step in and fix this situation. Removing Sean from the team would give the team a much needed boost.

Thank you for your consideration,

[Name].

Author G, task 2

To the Management Team,

Sean has pushed his team too far. He jokes about weaknesses in front of others. He has taken credit for some of their work. He makes the team uncomfortable and says abusive things.

The team performs better than you've been told. They would do well under different leadership. Under actual leadership.

Kind Regards,

Anon.

Author H, task 1

To whom it may concern,

As an employee of XXX for the last X years, I have had the chance to work with and for many great people. In the previous months however, the behaviour of Mr. Smith has become intolerable, and has poisoned what was once a great working environment. I feel compelled to write this letter, because his actions around you (his superiors) starkly contrast those when you are out of earshot.

It is detrimental not only to the mental health of his subordinates when he regularly humiliates them in front of their team members, but also to the health of our company, as employee retention suffers too. Unfortunately the public berations are limited not just to humiliations, but also in numerous cases, to verbal abuse.

It may be that he feels his position is threatened by promising junior employees. I have come to this conclusion personally, as I recently observed him spreading false rumours about his subordinates, as if he wishes to impede their progress. I have not spoken about this with my colleagues, but I find his behaviour repugnant.

The working environment in our department is unacceptable. This letter marks my official resignation, and I wish to terminate my employment with the company effective July 31st. For the sake of my colleages, I implore you to take remedial action swiftly and effectively. Doing so may mean the department will not lose other key employees.

Sincerely,

[Name]

Author H, task 2

A fairytale

There was once a wonderful company called XXX where all the little men and women were happy. Then a ugly storm cloud came and ruined mood of the company. The storm, let's call it Mr. Smith, is unusual because it has two fronts. On one side, viewed from above, it is a nice fluffy white cloud. But viewed from below it is an ugly storm cloud indeed. It strikes those below with bolts of verbal-abuse-lightning, and rains down humiliation on all those below.

The nice people living below are struggling. Many of them talk of leaving, some are already actively looking for places with better weather. Some of them are not able to bear the rain and lightning. It never stops.

The irony is, the storm cloud is not evil. Just stupid. It thinks that by raining on the people below, by striking them down with lightning, and by starting rumour-tornados, he will stay safely high up in the sky, and the little people will stay on the ground. But people want to fly too.

Hopefully the cheerful sun, who only sees the nice side of the storm cloud, might see the true face of this storm cloud. The cheerful sun can cause wind that will blow away the bad storm cloud. Hopefully the sun will do this, otherwise the once cheerful people of XXX might look for another place where they are treated nicely by those floating in the sky.

Author I, task 1

To whom it may concern,

My name is Maria Tonato and I work for Jeremy as the Marketing Officer. I have decided to leave Forewards Inc due to another opportunity that has been offered to me recently which will allow me greater scope to pursue other areas of interest to me in my career. I also feel that is time for me to move on, despite many happy years at Forewards.

I feel that it is necessary to inform you of several instances of unprofessional conduct asserted by Jeremy which I have noticed as a common occurrence during my time here. I do not wish to cause trouble for Jeremy, but I believe that it is my responsibility to let you know of his conduct, as it has affected me at times, as well as causing low morale in the team. Once a week, Jeremy finds a reason to verbally abuse myself or one of my colleagues. On one occasion, he called Jenny a 'fat bitch' and suggested that she ate too much and that this was affecting her her ability to type quick enough for her to complete tasks on time. On another occasion I heard him speaking to Sam about Sarah's correspondence with him via email and suggested that she was not cc'ing him into emails so that she could try and take his position by receiving more information from another team about particular matters and purposefully not sharing this information with him. Jeremy has also called me a 'lazy bitch' a few weeks ago because I did not give him information I promised him by a deadline that he himself had set and I was not aware of. He did not say this in jest and even if he did, I do not accept anyone calling me that in the workplace. Jeremy has called me names quite often in the last 5 years I have worked with him. I have not spoken out as I was afraid to cause conflict and did not believe that he would be able to modify his behaviour.

As I said, I do not wish to badmouth Jeremy or say things that are untrue. I understand that these are quite serious claims and I think you will find that if you interview each team member individually, you will find that other situations have occurred in the last 5 years of a similar nature. Jenny and Sam are happy to continue working here and may not wish to speak badly of Jeremy, though I have told them that I am writing this letter and detailing the two occurrences I have just described.

Thank you for taking the time to read this letter and please feel free to get in touch should you require more information.

Kind Regards,

[Name]

Author I, task 2

To whom it may concern,

Jeremy is my boss. He is a horror to work with. He verbally abuses our team frequently. He finds a reason to do this to each of us at least once a week. He once called Jenny a 'fat bitch' and suggested that her 'large size' affected her ability to type. He once called Maria a 'lazy bitch' because she didn't email him some information in time for a deadline that he himself had set. Sarah was the victim of his gossip when he went and told Sam that she was trying to sabotage him by purposefully not cc'ing him in emails to another team.

I can't work with Jeremy anymore. He creates an environment where everyone is scared to communicate with him and one another for fear that he will hear what we are saying and somehow use this against each other.

This kind of high school gossip, verbal abuse and sabotage happens often. I feel it is my responsibility to let you know so that things will change.

Thank you for taking the time to read my letter. I am not sending this to be malicious or to cause trouble in our team. I am being 100% honest and want things to improve in my workplace.

Author J, task 1

To Whom It May Concern,

For the past several years I have been in the employ of ThisCompany, Inc. working at the Townsvilleshire location, where at present Mr. Token Manboss holds the position of general manager. While the conditions of working under Mr. Manboss have grown unbearable to the degree that I have elected to leave this job, I feel I cannot in good conscience pack it in without communicating some of my concerns up the chain of command. Given the values, vision, and character embodied in this company at the corporate level, I imagine that those in leadership would want to be aware when these same are being directly contradicted by a member of middle management.

Mr. Manboss's managerial style (as I have witnessed and experienced it firsthand) is not only abusive, but highly unproductive. He actively seeks to undermine his employees' work, impeding the success of the individuals under his supervision and thereby hindering the growth and vitality of ThisCompany, Inc. as a whole. If the cliché of chains and weak links is brought to bear, Mr. Manboss's strategy capitalizes on the weakening of other links so that his may appear stronger by comparison -- an approach that hardly serves the interests of the larger organism. Even disregarding the possible legal ramifications his conduct might bring about for ThisCompany, Inc. (several employees would be well-justified in filing formal charges of verbal assault), I think there is due cause for concern.

I have submitted my two weeks notice, and am looking forward to working in a healthier environment in the future -- a place where managers lead by example, not by denigration, and where success is conceived of in positive-sum terms (as opposed to zero-sum). However, I have enclosed my contact information below in the event that you may wish to contact me further about any of this. I am grateful for all that ThisCompany, Inc. represents, and for the opportunity to contribute to the pursuit of its vision and mission.

Respectfully,

[Name]

Author J, task 2

To Whom It May Concern,

For the past several years I have been an employee of ThisCompany, Inc. working at the Townsvilleshire location, where at present Token Manboss holds the position of general manager. The conditions of working under Mr. Manboss have grown unbearable, and many would pack it in and leave without a moment's hesitation; but as I believe that the values, vision, and character of this company are worth fighting for, I feel that those higher up the chain of command should be aware that these same are being directly contradicted by a member of middle management.

Mr. Manboss's managerial style (as I have witnessed and experienced it firsthand) is not only abusive, but highly unproductive. He actively seeks to undermine his employees' work, impeding the success of the individuals under his supervision and thereby hindering the growth and vitality of ThisCompany, Inc. as a whole. If the cliché of chains and weak links is brought to bear, Mr. Manboss's strategy capitalizes on the weakening of other links so that his may appear stronger by comparison -- an approach that hardly serves the interests of the larger organism. Even disregarding the possible legal ramifications his conduct might bring about for ThisCompany, Inc. (several employees would be well-justified in filing formal charges of verbal assault), I think there is due cause for concern.

I love my job, and believe that the Townsvilleshire location can be a healthier environment in the future -- a place where managers lead by example, not by denigration; where success is conceived of in positive-sum terms (as opposed to zero-sum); in short, a place where the tone of the office matches the tenor and tradition set forth at the corporate level. It is this hope which motivates my writing, and which further advises me to remain anonymous: should Mr. Manboss read this letter, I fear I would not be around long enough to see the aforementioned changes brought about. I am grateful for all that ThisCompany, Inc. represents, and for the opportunity to contribute to the pursuit of its vision and mission thus far.

Respectfully,

[Name] (Anonymous)