Register variation in malicious forensic texts

Andrea Nini, University of Manchester, andrea.nini@manchester.ac.uk

The study reported here examines a corpus of 104 authentic malicious forensic texts for register variation. A malicious forensic text is defined in this paper as a text that is threatening, abusive or defaming and that constitutes evidence for a forensic case. This corpus was firstly tagged with a set of situational parameters and then analysed using the same multidimensional model introduced in Biber (1988; 1989). The results of the study indicate that malicious forensic texts, similarly to non-malicious professional letters, are on average instances of the *Involved Persuasion* text type, which is characterised by linguistic features overtly expressing modality. The results also confirm that threatening texts tend to use more modal verbs than non-threatening texts. Furthermore, the personal knowledge between interactants was found to highly influence the level of information density of the texts, while the narrativity level of malicious texts was found to be affected by whether the text contains harmful content directed to the addressee or to a third party. These findings can inform and improve the authorship analysis of malicious texts and increase our understanding of the creation of language crimes.

Introduction

The present paper defines a malicious forensic text (MFT) as a text that is a piece of written evidence in a forensic case that involves threat, abuse, defamation or a combination of the above. These texts are typically analysed by forensic linguists in real cases of extortion, blackmail, ransom, threat, abuse, stalking and similar criminal activities (Coulthard and Johnson, 2007; Grant, 2008). Notwithstanding their centrality within the field of forensic linguistics, only limited analysis of their register has been carried out so far, with the term *register* defined as a linguistic 'variety associated with a particular situation of use' (Biber and Conrad 2009: 6).

Some research has been carried out for threatening texts, although these studies have until now mostly focused on the pragmatic dimension rather than on register variation. Fraser (1998), for instance, recognises that threats are a particular type of commissive illocutionary speech act (Searle 1979) and proposes that a threat can be further distinguished from other commissives such as promises because: (1) threats intimidate the addressee and (2) the speaker of a threat only expresses an intention and not a commitment to perform an act. Fraser (1998) therefore concludes that the illocutionary force of the speech act of threatening is created when three conditions are met:

C1: the speaker has the intention to commit (or have someone commit) an act;

C2: the speaker believes that the outcome of this act will be unfavourable for the addressee;

C3: the speaker intends to intimidate the addressee.

In terms of the language of threatening texts, Fraser (1998) concludes that even though there are several syntactic patterns common to threatening utterances, it is difficult or probably impossible to determine whether a text is threatening by the language alone. These conclusions are also supported by Solan and Tiersma (2005) and, more recently, Kaplan (2016), who highlight the importance of the context in the determination of the illocutionary/perlocutionary threatening status of an utterance.

The focus on the intention to commit an act and on the intimidation proposed by Fraser (1998) and Solan and Tiersma (2005) was largely confirmed linguistically by Gales' (2010; 2011) exploration of a sample of threatening letters for their lexicogrammatical expressions of linguistic stance (Biber, 2006). As part of her investigations, Gales compared a corpus of 470 threatening texts against a control corpus of 556 non-threatening texts and found that the high incidence of modals is a characteristic of threatening texts. Furthermore, Gales discovered that threatening texts were far more likely than non-threatening texts to express deontic/volitional meaning aimed at controlling the addressee. This use of modals, especially in prototypical syntactic threatening patterns such as 'if you don't X I *will* Y', was also noted by Fraser (1998: 168) as the prototypical realisation of a threatening act. These same features were found by Gales (2015) to be even more frequent in threatening stalking texts.

Some further preliminary work on the classification of threatening texts was also carried out by Napier and Mardigian (2003), who divided threatening texts into three categories:

1. *direct/specific*: the threat specifies clearly all the circumstances in which the harmful action will be carried out;

2. *conditional*: the threat is presented as a condition with a request for the recipient to carry out an action;

3. *indirect*: the threat is not conditional and not precise regarding its circumstances.

Although this preliminary taxonomy resembles an initial register variation analysis, Napier and Mardigian's (2003) results were not based on linguistic criteria but on investigative experience.

Research on other types of MFTs, such as abusive texts, ransom demands, extortion letters, or defaming texts, has been so far restricted mostly to case studies. For example, real cases involving ransom demands and extortion letters have been discussed in several forensic linguistic studies (McMenamin 2002; Kniffka 2007; Olsson 2003). Similarly, texts involved in defamation cases have been extensively explored from a pragmatic perspective (Shuy 2010; Tiersma 1987). In general, however, a register analysis of a corpus of MFTs is still missing.

The virtual absence of research on the register of MFTs can be attributable to at least two factors. Firstly, MFTs are generally difficult to access, thus making the analysis of large corpora of these texts rather difficult. Secondly, it is perhaps possible that MFTs are generally not considered as being regulated in their structure and form by any community and hence unlikely to present common linguistic features that characterise their register. Even though MFTs are not texts that are commonly encountered, these texts do share many extra-linguistic parameters of context, such as being written as opposed to oral, being produced with the potential of being edited, as opposed to having real time production constraints, and so forth, and this collection of parameters is likely to influence the occurrence of certain linguistic features in such a way as to distinguish MFTs from other registers of a language or to create substantial differences in sub-registers within the MFTs themselves. The analysis of the language of MFTs according to these and other parameters of context can help in reaching conclusions regarding their linguistic identity. This knowledge would give forensic linguists the ability to predict which linguistic features to expect in MFTs depending on their situational characteristics.

The present study is aimed at filling the gap in the research on the register identity of MFTs, that is, on which linguistic features are typical and pervasive of the situational characteristics of MFTs. This study explores MFTs in order to establish: (1) how English

MFTs compare to other common registers of the English language; (2) which situational parameters are the most useful to describe MFTs and to predict variation in their language; (3) whether a categorisation of MFTs based on their language can be established.

Methodology

In order to carry out a register analysis, three components are needed: a description of the situational characteristics of the texts, a description of the language of the texts, and the functional interpretation of the connection between linguistic features and situational context (Biber and Conrad 2009). To this end, each text in the corpus was firstly tagged with a set of situational parameters that describe their extra-linguistic characteristics and secondly analysed for a set of linguistic features that describe their language. This section firstly describes the MFT corpus and then both sets of features in more detail.

The Malicious Forensic Texts corpus

The Malicious Forensic Texts corpus is a corpus of authentic forensic texts compiled for the purposes of the present study. In total, the corpus comprises 104 texts, with a total of 39,188 word tokens and an average text length of 357 tokens (min: 103; max: 1610; SD: 280.3). The dates of production of the texts span from 1937 to 2013: 35% from 1937 until 1970; 22% from 1970 until 2000; 16% after 2000; 27% of unknown date. Because of this fragmented distribution, in the present study the diachronic perspective is ignored and the corpus is overall considered as a synchronic sample. Since the present analysis deals with linguistic features such as frequencies of parts of speech or general grammatical constructions, it is unlikely that this choice significantly affected the results. Although, ideally, a corpus should be both large and equally sampled across several parameters, this is virtually impossible for MFTs because of their rarity. For this study, it was thus decided to include as many texts as possible and allow some imbalance.

The majority of the texts in the corpus (63%) was collected from the FBI Vault, the repository of old case files that the FBI has digitised and made public (https://vault.fbi.gov/). A maximum of three texts per case were chosen from this repository so as to avoid the possible influence of a single author's style on the overall corpus. The second largest data source for the corpus (18%) was the private collections of MFTs owned by a number of

American and British forensic linguists who agreed to provide anonymised data for the present research. Web searches were also carried out with the aim of finding publicly available authentic MFTs and this source provided 13% of the texts of the corpus. These sources were mostly online newspapers or websites and online encyclopaedias dealing with high profile cases, such as the Zodiac case (Kelleher and Van Nuys 2002). Finally, the least common category (6%) consisted of those texts published in books, such as Olsson (2003).

The task of data collection was not unproblematic. Even though the MFT corpus has the advantage of being constituted of authentic MFTs, the difficulty of gathering such highly confidential forensic data led to the impossibility of collecting texts by rigorously sampling the population. As such, the distribution of situational characteristics found in the corpus does not necessarily reflect their real distribution. Despite this limitation, most of the findings are relatively unaffected by this potentially skewed distribution because the analyses involve comparisons of sub-sections of the corpus using rigorous statistical tests of significance.

Since the main bulk of the analysis was largely quantitatively oriented, only texts with at least 100 word tokens were included in the corpus. This methodological decision reflects the fact that even though most of the features are normalised by text length, the required text length to estimate the real frequency of a feature from a sample varies depending on the rarity of the feature (Biber, 1993). Most of the very frequent linguistic features are already stable in 100 token samples and previous multidimensional studies have employed the 100 token cut-off with good results (Biber and Jones, 2005).

The situational parameters

The extra-linguistic context of each text was described using a taxonomy of parameters initially based on Biber's (1994) Situational Parameters of Variation. Biber's (1994) framework was developed as an attempt to arrive at a universal taxonomy of features of the situations that influence the language of a text. The framework used here is an adapted version of Biber's (1994) original taxonomy. Three types of modification were made to Biber's (1994) taxonomy in order to adapt the system to the present study. Firstly, since all of the MFTs in the corpus share the same characteristics of being written texts with no immediate feedback, the parameters in Biber's (1994) system, originally designed to account for these options, were not considered for the present study. An example of such a parameter is Permanence, for which the option *recorded* would apply to all of the texts in the MFT

corpus and which can therefore be excluded. Secondly, some parameters introduced by Biber (1994), such as Topic, were not included because of their open ended and subjective character that does not lend itself to the type of analysis carried out in the present work. Finally, some original parameters were added to the taxonomy to account for the fact that within MFT there is variation of, for example, the kind of maliciousness or the direction of harm or other distinctions that are not present in Biber's (1994) original taxonomy.

A set of 13 parameters was eventually considered to analyse the corpus of MFTs, where each parameter consisted of a set of alternative options that model an aspect of their extra-linguistic context. Table 1 presents the parameters considered for this study with the values for each parameter ordered from the most frequent to the least frequent in the corpus. Not every text could be classified for each parameter due to missing information and therefore the option 'unknown' was also included.

Table 1 – Summary of the 13 parameters of the situation used to analyse the MFT corpus

Parameter	Values	Distribution in MFT corpus
Addressor	*single:* the addressor was a genuine or fictitious single person	65 (63%)
	institutional: the addressor was a genuine or fictitious organisation	13 (13%)
	plural: the addressor was a genuine or fictitious group of people	12 (12%)
	unknown	14 (13%)
Addressee	*high profile person:* a text addressed to a person who has higher social power than the addressor, that is, a person who could in theory control the behaviour of the addressor – e. g. 'Dear Prime Minister', 'Dear Inspector Smith'	24 (23%)
	celebrity: a text addressed to a celebrity - e. g. 'Dear Marilyn Monroe'	16 (15%)
	unacquainted person: a text addressed to a known addressee for whom there is evidence that they are not acquainted with the	13 (13%)

	addressor - e. g. 'Dear Mr Smith'	
	impersonal: a text addressed to no one or to an institution or entity - e. g. 'to whom it may concern', 'FBI' or no addressee	11 (11%)
	unknown person: a text addressed to an unknown person - e. g. 'Dear Sir/Madam', 'Dear Editor'	11 (11%)
	acquaintance: a text addressed to an acquaintance - e. g. 'Hello John'	11 (11%)
	open: an open letter addressed to a group of people – e. g. 'Citizens of New York', 'To the people of the US'	3 (3%)
	unknown	15 (14%)
Audience	*no audience:* the text did not have an audience	80 (77%)
	audience: the text had an audience, that is, one or more non-addressee individuals who nonetheless the addressor knew would have read the text	13 (13%)
	unknown	11 (11%)
Interactiveness	*none*: the text resembled a published prose, as it did not require any action from the addressee apart from reading the text	75 (72%)
	slight: the text required some form of response or interaction from the addressee, for example by confirming a piece of information or by requesting some actions	29 (28%)
Place of Communication	*private:* the communication happened in a private space	94 (90%)
	public: the communication happened in the public domain	4 (4%)
	unknown	6 (6%)
Extent Place is Shared	*removed:* the interactants were not	94 (90%)

	aware of their respective locations when the communication took place	
	familiar: the interactants were familiar with the place in which the communication took place even though they were not sharing it during the time of the communication	4 (4%)
	unknown	6 (6%)
Mode	*handwritten*: the text was handwritten	46 (44%)
	typed: the text was typed using a computer or typewriter	37 (36%)
	unknown	21 (20%)
Embedded in a Larger Text	*not embedded*: the text was not embedded in a larger text	99 (95%)
	embedded: the text was part of a larger text	5 (5%)
Malicious Purpose	*threat:* one of the purposes of the text was to threaten the addressee or a third party, where a threat is defined following the conditions outlined in the section above	51 (49%)
	defame: one of the purposes of the text was to deliver information that is defamatory to the addressee or a third party but that is not abusive	36 (35%)
	abuse: one of the purposes of the text was to overtly and extensively be abusive (insult/offend/express hate) towards the addressee or a third party in more than just one instance	18 (17%)
Threat Type	*violent act:* the threat was a violent act	40 (78%)
	extortion: the threat involved a request of money	6 (12%)
	spread malicious information: the threat involved spreading malicious	3 (6%)

	information regarding the recipient of the threat	
	ransom: the threat involved harming the victim of a kidnapping	2 (4%)
Threatening Communication Type	*indirect*: the threat was not conditional and not precise regarding its modalities	30 (59%)
	conditional: the threat was presented as a condition in which a request is made to the recipient to carry out a real action, that is, an action that the recipient should be able to carry out to stop the threat	19 (37%)
	direct/specific: the threat specified clearly all the modalities in which the harmful action would be carried out	2 (4%)
Direction of Harm	*towards third party:* the threat/abuse/defamation was directed towards a person or people that were not the addressee of the text	40 (38%)
	towards addressee and third party: the threat/abuse/defamation was addressed to both the addressee of the text and a third party. This option included those threats in which the violent action involved the addressee and their family or in cases in which the threat was directed to the addressee and the violent act to a third party - e. g. 'give me the money or I will kill your daughter'	28 (27%)
	towards addressee: the threat/abuse/defamation was directed to the addressee of the text	26 (25%)
	unknown	10 (10%)

Most of the parameters were taken from Biber's (1994) system, with the exception of the original parameters specific to MFTs, that is, Direction of Harm, Malicious Purpose, Threat Type (based on Napier and Mardigian's (2003) classification), and Threatening Communication Type. The last two categories applied only if a text contained a threat. Some other parameters were modified to be adapted to MFTs, such as the parameter Interactiveness, for which the value *extensive* was removed as it never applies to MFTs. The most substantial changes were applied to the parameter Addressee, where instead of a general distinction in *single/plural,* finer categories tailored to the corpus were adopted. Values for all the parameters were assigned using information from the extra-linguistic context of the text. For each parameter only one value could be selected, with the exception of Malicious Purpose, since texts can perform multiple malicious acts at the same time.

The section below describes the linguistic framework used to measure the linguistic variation that the texts present across the situational parameters.

The linguistic features

For the sake of performing a register analysis of the corpus, a multidimensional analysis of linguistic features was applied following the paradigm created by Biber (1988) and further applied in the exploration of several other registers (e.g. Grieve et al. 2011; Titak and Roberson 2013; Gray 2013). In this work, Biber (1988) automatically calculated the frequency of a set of 67 lexical and grammatical features that have been previously found to vary across different registers, such as parts of speech or more complex grammatical constructions like relative clauses, WH clauses, sentence relatives, and so forth. He then performed a factor analysis in order to find dimensions of variation that could account for the latent functional patterns that group and explain the frequency of these features. Biber (1988) found that 23 of the most common registers of the English language can be located in a six dimensional space created by the six dimensions summarised in Table 2.

Table 2 – **Summary of Biber's (1988) dimensions of register variation for English**

Dimension	Description
Dimension 1 **Involved vs Informational discourse**	Low scores on this dimension indicate that a text is informationally dense and that it uses many nominal features, as for example in academic prose, whereas high scores indicate that a text is affective and

	interactional and that it uses many verbal and pronominal features, as for example a casual conversation.
Dimension 2 **Narrative vs Non-Narrative Concerns**	Low scores on this dimension indicate that a text is non-narrative whereas high scores indicate that a text is narrative, and that therefore uses many past tenses and third person pronouns, as for example a novel.
Dimension 3 **Context-Independent Discourse vs Context-Dependent Discourse**	Low scores on this dimension indicate that a text is dependent on the context and that it uses many adverbial features, as in the case of a sport broadcast, whereas a high score indicates that a text is not dependent on the context, as for example an academic text.
Dimension 4 **Overt Expression of Persuasion**	High scores on this dimension indicate that a text explicitly marks the author's point of view as well as their assessment of likelihood and/or certainty by using modals or other features that express modality, as for example in professional letters or prepared speeches.
Dimension 5 **Abstract vs Non-Abstract Information**	High scores on this dimension indicate that a text provides information in a technical, abstract and formal language with the use of conjuncts and passive voice, as for example in scientific discourse.
Dimension 6 **On-line Informational Elaboration**	High scores on this dimension indicate that a text is informational in nature but produced under certain time constraints and therefore using nominal elaboration such as *that* relative clauses, as for example in prepared speeches.

As a follow-up to this study, Biber (1989) used a cluster analysis to automatically group the texts in his corpus according to their dimension scores. Each cluster of texts was defined by Biber (1989) as a *text type*, since it represents a type of language production that reflects shared linguistic identity. The eight text types found by Biber (1989) are summarised in Table 3.

Table 3 - A summary of Biber's (1989) text types

Text type	Characterising registers	Characterising Dimensions	Description
Intimate Interpersonal Interaction	telephone conversations between personal friends	high score on D1, low score on D3, low score on D5, unmarked scores for the other Dimensions	This cluster typically contains interactions that have an interpersonal concern and that happen between close acquaintances.
Informational Interaction	face-to-face interactions, telephone conversations, spontaneous speeches, personal letters	high score on D1, low score on D3, low score on D5, unmarked scores for the other Dimensions	This cluster typically contains personal spoken interactions that are focused on informational concerns.
Scientific Exposition	academic prose, official documents	low score on D1, high score on D3, high score on D5, unmarked scores for the other Dimensions	This cluster typically contains informational expositions that are formal and focused on conveying information and very technical.
Learned Exposition	official documents, press reviews, academic prose	low score on D1, high score on D3, high score on D5, unmarked scores for the other Dimensions	This cluster typically contains informational expositions that are formal and focused on conveying information.
Imaginative Narrative	romance fiction, general fiction, prepared speeches	high score on D2, low score on D3, unmarked scores for the other Dimensions	This cluster typically contains texts that present an extreme narrative concern.
General Narrative Exposition	press reportage, press editorials, biographies, non-sports broadcasts, science fiction	low score on D1, high score on D2, unmarked scores for the other Dimensions	This cluster typically contains texts that use narration to convey information.
Situated Reportage	sports broadcasts	low score on D3, low score on D4,	This cluster typically contains on-line

		unmarked scores for the other Dimensions	commentaries of events that are in progress.
Involved Persuasion	spontaneous speeches, professional letters, interviews	high score on D4, unmarked scores for the other Dimensions	This cluster typically contains persuasive and/or argumentative texts.

In the present study, this same model comprising six Dimensions, eight text types and 67 linguistic features was applied to the MFT corpus using the Multidimensional Analysis Tagger (MAT) (Nini, 2015a). This computer program replicates Biber's (1988) analysis by plotting a new text or corpus on to Biber's (1988; 1989) original dimensions and text types. Using the Stanford Tagger (Toutanova et al., 2003) MAT performs an initial part-of-speech tagging of a text and it then applies the algorithms presented in Biber's (1988) appendix to calculate the frequencies of those same 67 linguistic features used in the study. The program then plots the analysed text or corpus on to the six Dimensions proposed and it assigns one of the eight text types to the analysed text or corpus. By using MAT to plot the MFT corpus on to the multidimensional model, it is possible to obtain a comprehensive linguistic analysis of the texts and, at the same time, a comparison of the MFT corpus with their registers and text types of the English language. This analysis can then shed light on the register identity of the MFTs. To explore the accuracy of the tagger for MFTs, a manual check of a random 20% of the data was performed. For each text, the percentage of correct tags (or true positives) was divided by the number of total tags so as to have a measure of precision. On average, MAT performed well, achieving an average precision of 96% (Min. = 87%, Max. = 100%, SD = 3.4). The next section presents a description of register variation in the MFT corpus as well as a comparison of the MFTs to other registers and text types of the English language.

Analysis

After tagging the MFT corpus with the situational parameters and after running MAT for each text, a contrastive analysis was carried out to compare the language of the texts with different situational parameter values (e. g. how are *threatening* MFTs linguistically different from *non-threatening* MFTs?). This type of comparison involves the statistical testing of significance between two or more samples. Since the majority of the frequencies of linguistic features considered were not normally distributed, non-parametric tests of significance were

chosen. These tests were the Mann-Whitney U test for the analysis of parameters with only two possible values and the Kruskal-Wallis test for parameters with more than two possible values (Dancey and Reidy 2011: 528-565; Baayen 2008: 108).

A series of non-parametric statistical tests of difference for each of the 67 linguistic features considered above plus the six Dimension scores was therefore carried out for each of the situational parameters. As opposed to the classic 0.05 p-value threshold, the stricter p-value of 0.001 was considered as the threshold for statistical significance for this study, meaning that any significant result has a 1 in 1000 chance of being due to sampling error. Using this significance threshold, the parameters Addressor, Audience, Interactiveness, Place of Communication, Extent Place is Shared, Embedded in a Larger Text, Threat Type and Threatening Communication Type did not present any significant effect. The significant effects observed for the remaining parameters are discussed in the sections below.

Malicious Purpose

For this parameter, only two of the malicious purposes considered showed significant effects: *threat* and *defame,* while the contrast between abusive and non-abusive texts did not present significant differences.

The analysis of threatening texts showed that previous findings related to threatening language are confirmed in the MFT corpus. Threatening texts presented a higher frequency of prediction modals (e. g. *will* and *shall*), second person pronouns, and first person pronouns than non-threatening texts. Threatening texts also had higher scores on Dimension 4, Overt Expression of Persuasion, and a shorter average word length. Finally, threatening texts presented a lower frequency of third person pronouns, and of attributive adjectives. The distribution of these features are visualised in the boxplots in Figure 1.

These findings suggest that the frequency of all personal pronouns is an important feature of threatening texts. However, although threatening texts present many personal pronouns overall, threatening texts are more likely to directly address the recipient and are therefore more likely to use first and second person pronouns rather than third person pronouns. The shorter average word length itself is partially explained by the frequent use of pronouns, as a correlation test confirms (*Spearman rho* = -0.573, p < 0.000001).

Figure 1 - Boxplots describing the distribution of first person pronouns (top left), second person pronouns (top right), third person pronouns (bottom left) and average word length (bottom right) for threatening and non-threatening texts
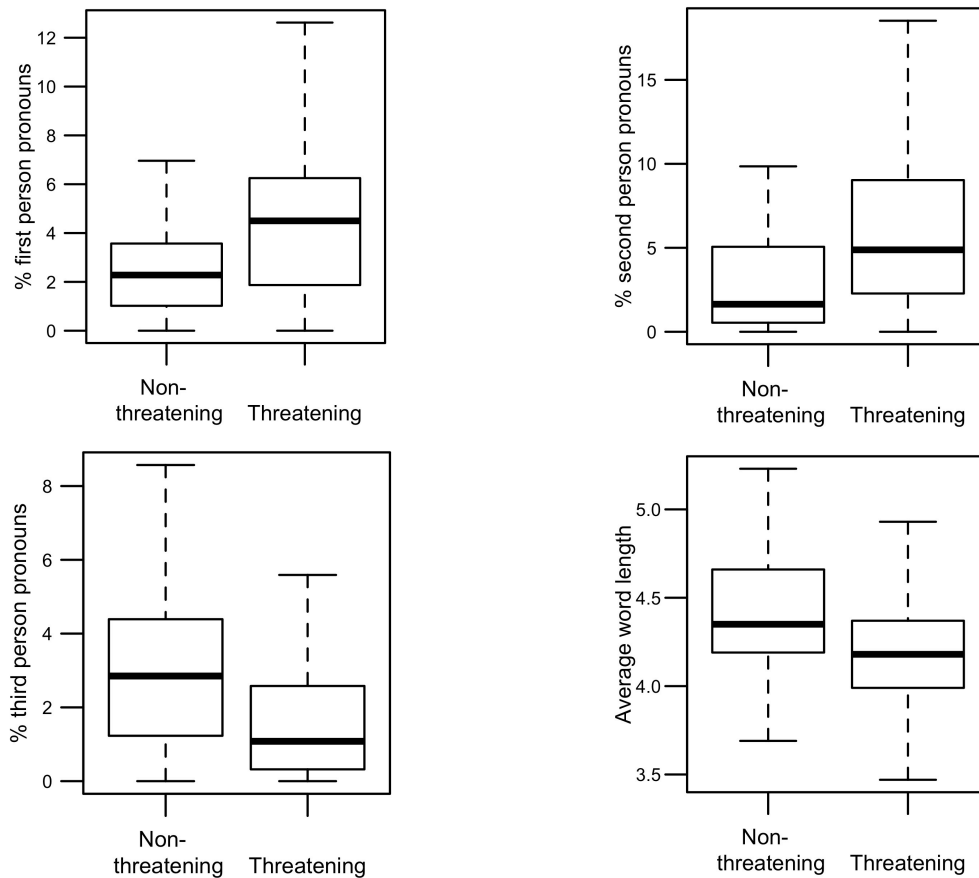
Overall, the more frequent use of pronouns and less frequent use of attributive adjectives indicate that the threatening texts in the corpus are less likely to use complex noun phrases and, therefore, less likely to be informationally dense than non-threatening texts.

A further significant pattern noted for threatening and non-threatening texts is related to the different distribution of Dimension 4, the Dimension of Overt Expression of Persuasion, characterised by high incidence of modality. As the boxplots in Figure 2 show, threatening texts tend to have on average higher scores on Dimension 4 and, consequently, to be on average more explicit in their use of persuasive means than non-threatening texts.

Figure 2 - Boxplots representing the distribution of Dimension 4 (left) and prediction modals (right) for threatening and non-threatening texts
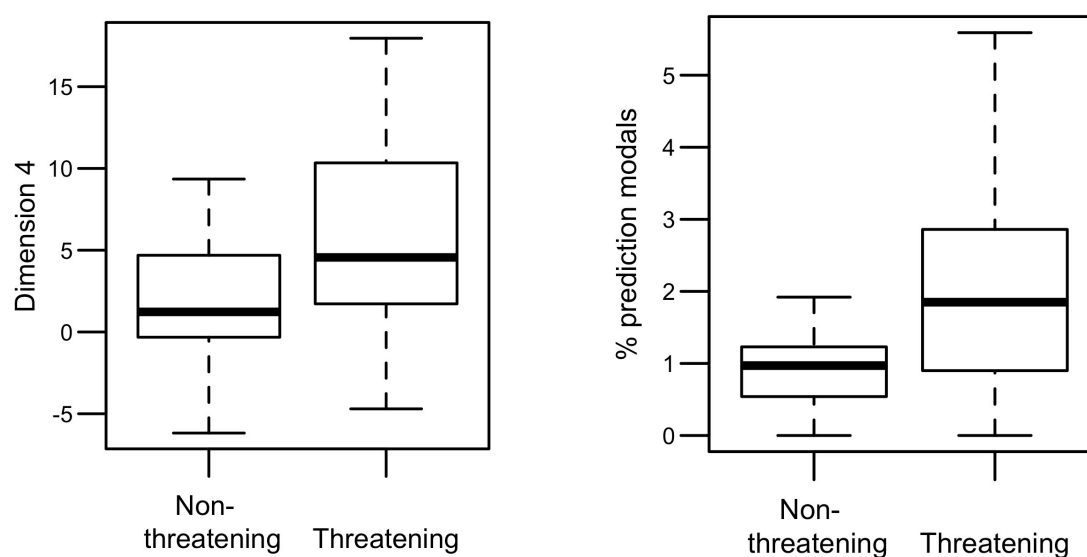
These findings thus strongly support Gales' (2010; 2015) conclusions, since a high score on Dimension 4 corresponds to a more frequent use of means of expressing modality, such as suasive verbs, infinitives or modal verbs. Indeed, the prototypical example of a threatening utterance given by Fraser (1998: 168), '*if* you don't X I *will* Y', is exemplary in showing two Dimension 4 features: conditionals (*if*) and predictive modals (*will*). These features that express persuasive discourse are sprinkled across threatening texts, as shown in Table 4, where a letter collected from the FBI Vault and addressed to Senator Helms is displayed with its Dimension 4 features highlighted.

Table 4 - A sample text marked for its Dimension 4 features (prediction modals, infinitives, suasive verbs and split auxiliaries). The features are in bold and underlined. The language of the text is reproduced as in the original, including its non-standard features.

Letter collected from the FBI Vault and addressed to Senator Helms

Dimension 4 score = 15.79

Dear Senator Helms,
or to whoever reads this letter first.

Inform everyone who are national senators and representatives and say or show this letter amonge them.

*A desolation has been **<u>decreed</u>** upon the nation and upon those in the nation. **<u>If</u>** I do not receive $25,000 dollars in cashiers checks at $25 each these things **<u>shall</u>** take place.*

*The two wings of the capital building **<u>shall</u>** crumble to the ground. It **<u>will</u>** happen in this manner.*

*Twenty four hours after this letter is read by anyone the walls of the two wings **<u>shall</u>** being **<u>to shake</u>** and **<u>will</u>** shake for 30 minutes. At the end of thirty minutes the walls **<u>will</u>** **<u>slowly give way</u>** **<u>allowing</u>** any and all who are in there **<u>to flee</u>**. The crumbling **<u>will</u>** last for 30 minutes and at the end of the 30 minutes period all of the two wings of the capital **<u>shall</u>** be crumbled to the ground.*

*The part that holds the dome **<u>shall</u>** not be harmed, just the outer wings.*

*I **<u>will</u>** deal with the SBI official here in Murfreesboro N.C.*

As the example shows, the persuasive features of Dimension 4 are used to express the conditional threat (*if I do not receive $25,000 [...] these things shall take place*). However, when the threats are made explicit, a series of Dimension 4 features such as infinitive clauses and certainty modals are used in order to emphasise the certainty of the outcomes of the threat.

The other malicious purpose that presented significant effects after the analysis was *defame*. The analysis of this parameter showed that defaming texts contained more third

person pronouns, perfect aspects, *that* as verb complement, and past tenses than non-defaming texts. Defaming texts presented significantly fewer predictive modals, first person pronouns, and second person pronouns than non-defaming texts. Finally, defaming texts presented a higher average word length and a lower Dimension 4 score than non-defaming texts. In sum, the results of the analysis of defaming *vs.* non-defaming texts seemed to mirror the results of the analysis of threatening *vs.* non-threatening texts, in the sense that defaming texts seemed to behave as non-threatening texts. This opposition comes from the fact that these two categories rarely overlapped in the corpus, since only three texts were both threatening and spread malicious information at the same time. This bias comes from the main disadvantage of the corpus of not being a balanced sample, which was difficult to achieve given the lack of availability of MFTs.

Another characteristic of defaming texts was the fact that they also tended to be more narrative than non-defaming texts, as highlighted by the high incidence of past tenses, perfect aspects and third person pronouns. However, this characteristic is again linked to another parameter of the context, the direction of harm. Eighty per cent of the defaming texts were indeed written to the addressee with the purpose of spreading malicious information about a third party and the analysis of the direction of harm parameters showed that texts with harm directed to a third party were found to be more narrative than other kinds of texts (see section below).

Addressee

The categories of addressee showed differences for second person pronouns, present tenses, and Dimension 1, Involved *vs* Informational discourse. These three variables consist of the same underlying pattern, since both second person pronouns and present tenses are used in the calculation of Dimension 1. In order to explore the underlying pattern of variation between Dimension 1 and the types of addressee, the distribution of Dimension 1 was plotted in Figure 3, rank-ordered by medians of Dimension 1, starting from the group with the lowest median to the group with the highest median.

Figure 3 - Boxplots showing the distribution of Dimension 1 for the different addressee types ranked by medians from lowest to highest.
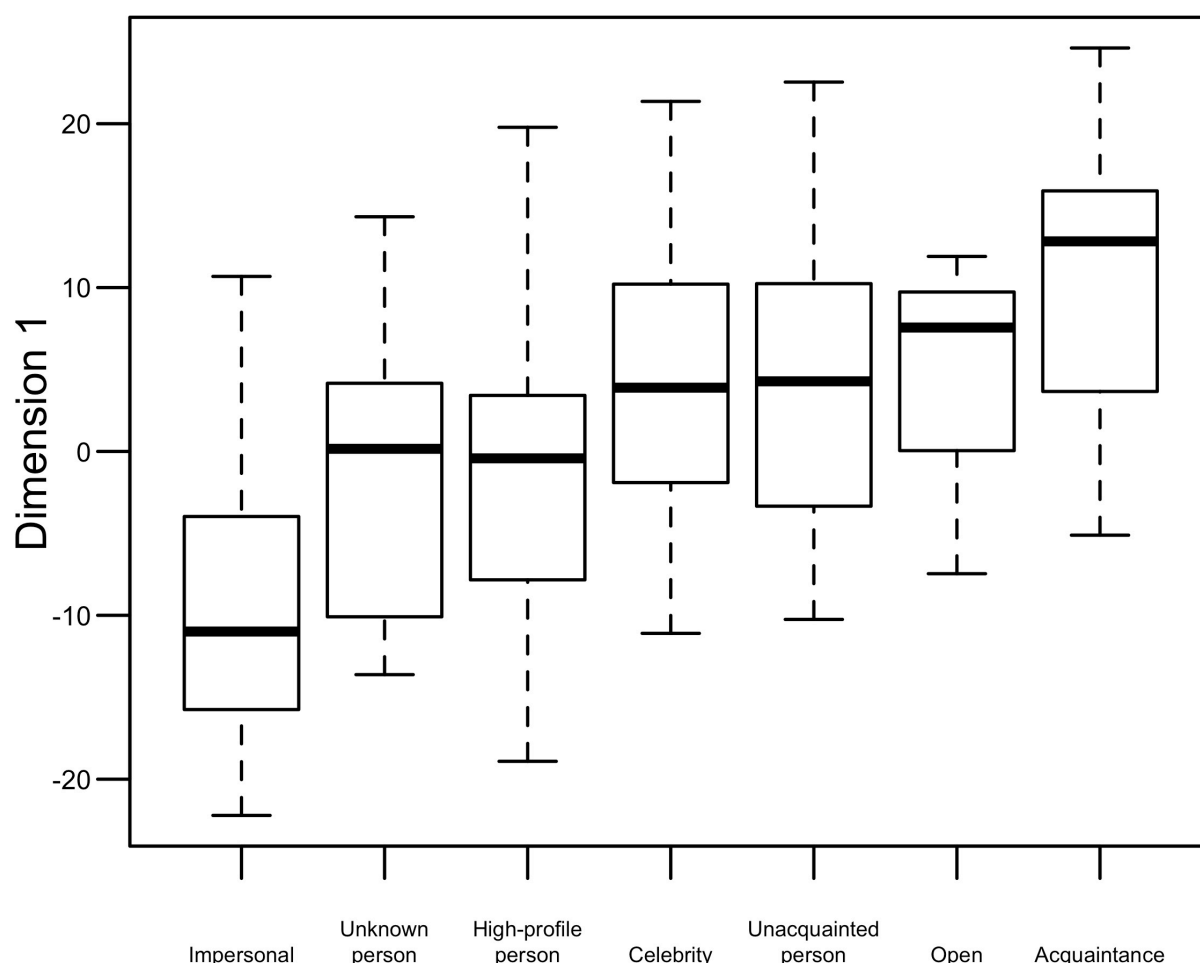
The graph indicates that the addressees of MFTs are arranged by Dimension 1 in terms of interpersonal distance. These findings thus support Bell's (1984) *audience design* theory, which maintains that the adjustment of a speaker/writer's style according to their addressee is the one of the most fundamental predictors of linguistic variation. The present findings support the hypothesis that this holds true even for MFTs. Texts with an impersonal addressee are the texts that score the lowest on Dimension 1, thus showing a highly Informational character. As the distance between addressor and addressee decreases, the degree of Involved discourse starts to increase. Therefore, an unknown person and a high-profile person are at the same level of Dimension 1, whereas texts addressed to celebrities or other unacquainted persons are slightly more Involved. Texts addressed to acquaintances are the most Involved of all the texts. Open letters also showed a high degree of Involvement. However, since this category included only three cases, it is not possible to understand how this category would fit into the more general pattern of interpersonal distance. By looking at

19

the boxplots it is also possible to notice that the largest differences are present between texts with an impersonal addressee and texts with an acquaintance addressee, with other texts being in the middle of the cline. Thus, excluding the three open letters, the addressee types can be grouped in three main categories of personal knowledge: *no personal knowledge*, including only texts with an impersonal addressee; *medium personal knowledge*, including texts with unknown addressees, high-profile addressees, celebrity addressees and unacquainted addressees; *high personal knowledge*, including only texts with acquaintance addressees. The boxplots in Figure 4 show the distribution of Dimension 1 for these three categories of personal knowledge.

Figure 4 - Boxplots showing the distribution of Dimension 1 for the different degrees of personal knowledge.
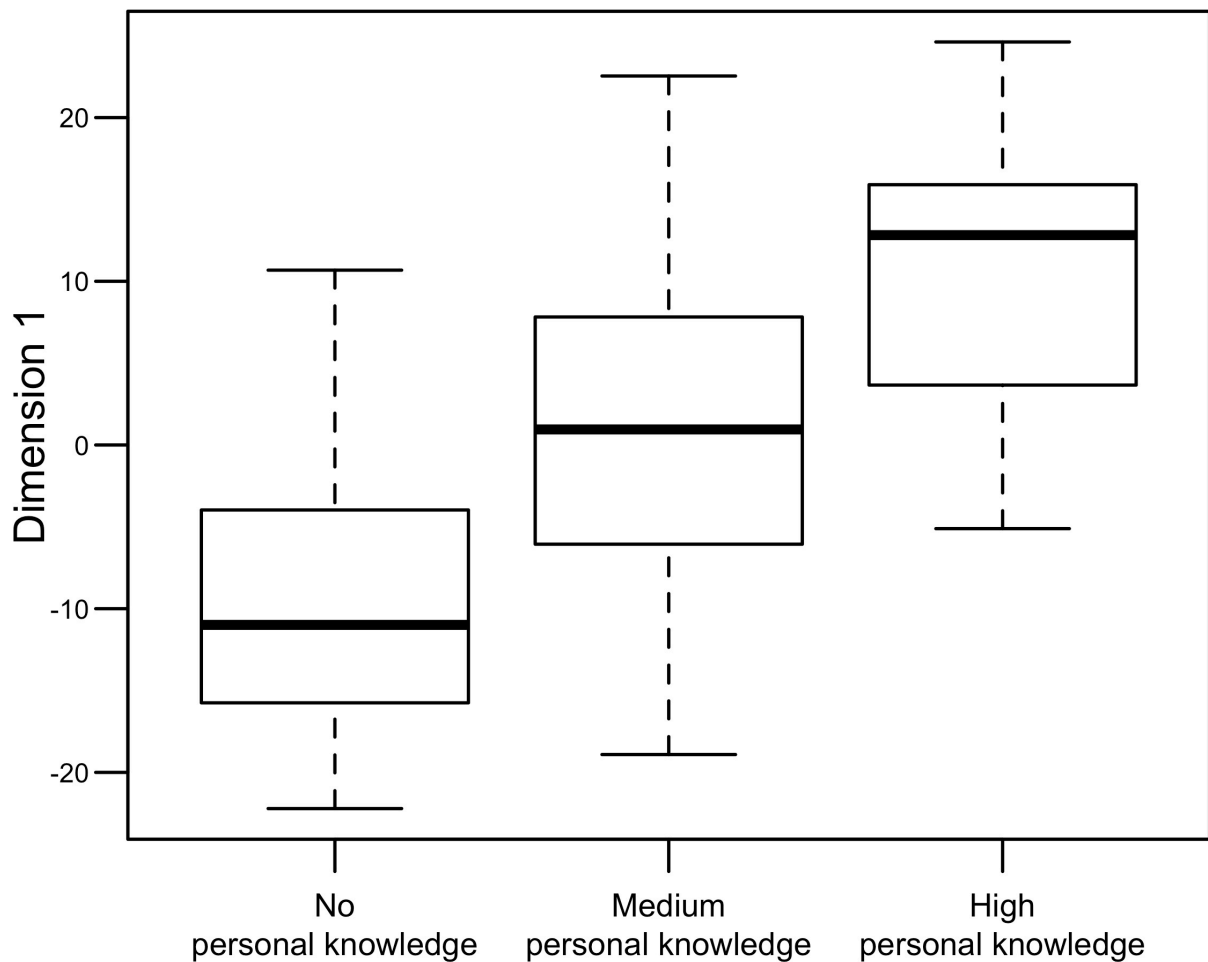
Samples from the data can be seen in Table 5, where one text with an impersonal addressee, one text with a high-profile addressee and one text with an acquaintance addressee are shown with their Dimension 1 features marked in the text.

Table 5 - Three samples of texts marked for their Dimension 1 Involved features (bold) and Dimension 1 Informational features (underlined). The language of the texts is reproduced as in the original, including the non-standard features.

Impersonal letter collected from the FBI Vault and addressed to the FBI

Dimension 1 score = -18

FBI-

*The Progressive Labor Party of Chicago, Illinois **is** going to kill U.S. Senator Jesse Helms of North carolina. The reason **we are** going to do **this is because** the U.S. government **has** been committing atrocities against the members of the New Jewell Movement in Grenada ever since the United States invasion of Grenada in October, 1983. **We are** going to wait until Senator Helms **visits** Boston, Massachusetts. **We are** going to catch him in the subway station in Boston. A Hispanic member of the Progressive Labor Party will sneak up behind Senator Helms while he **is** standing on the platform of the subway station waiting for the subway train to arrive. The Hispanic person will suddenly grasp Senator Helms' ankles with his hands and pull backwards. Helms will drop and smash his face hard onto the floor of the subway station. A white man will then quickly quietly kick Senator Helms in the side of his head like **it's** a football. Then three Chinese members of the Progressive Labor Party will grab Senator Helms and toss him head first off of the platform onto the railway tracks in the subway station **just** before the subway train **rolls** past. When the front of the train **runs** over Jesse Helms **it** will smash him and squash him and slice him up so that he **dies**. After the subway train **leaves** the subway station, then a Black person who **is** a friend of the Progressive Labor Party will take photographs of Helms's remains to publish in the newspaper, "challenge". In the meantime the five heros who knocked "scum of the earth" Helms onto the railway tracks will be riding on the train that ran over him and getting off at various stops to make good their escape. They will then go celebrate Helms' death.*

High-profile addressee letter collected from the FBI Vault and addressed to the Vice President

Dimension 1 score = 1.88

My Dear Vice President --

***I have** become **very** alarmed about the disunity that **has** become **more** than **just** serious --*

I am for *President Johnson,* **you** *and* Dean Rusk *hundred* percent --

You have *no* idea *how this* effects **me** -- *Bobby* Kennedy **is** *taking* advantage *of the* youth *about the* Veitnam War *-- This* borders on treason *-- He* **is** *using* Politicis, **and** *to* **me** *he* **is** *dangerous and a* traitor *using his* high *Position to cover* **this** *-- he* **wants** Power *& he* **is** *selling the* U.S. out *to get* **it** *--*

I could *go* on *&* on *about* men *like him -- William* Fulbright **is** *a* **more** frightful *enemy* than *Ho* Chi Minh *--*

Mr. Vice President **I have** *a* Son *going* to *the* war zone for *the* 2nd time *--* **I have** *another* Son *who will enter* service *soon --* **I have** *a* Son in Law *in the* Service *--* **I am not** *going to let* men *like* McCarty, Kennedy, Morse, Martin King, Stokley Carmichael *divide this* country *-- To keep* encouraging Hanio *-- The* idea of Rep. Lester Woff *-- bringing up a resolution to repeal the* Tonkin Gulf *resolution* **is** *insane --* **this has** *got to be stopped --* **Our** boys *over there* **have** *got to have* **our** full support *--*

Mr. Vice President **have you** *any* idea *what it is like to play the* game of Russian Roulette? **Well** *that what the* Marines at K.S. **are** *doing ever* hour of *the* day *--* **and** *they* **know** *what* **[WH-clause]** *they* **are** *fighting* for *--* **We need** complete victory in Vietnam *-- even* to the point of *invading* N. Vietnam *--* Fulbright *--* Kennedy *--* King *--* Morse **have** *delayed this* victory *--*

Mr. Vice President **I am** *begging* **you** *to help* **me** *-- some* way *legally to shut these* men *up --* **Now**! *--* **I am** *a* ex-Marine *--* **my** passions *& emotions* **are very** *high* in *this* matter *--*

I would **really feel it** *a* duty *to eliminate these* men *or as many as* **I could** *before* **I** *was killed --* **I believe** in *this* enough *to die* for **my** Country -

Dear Bill,

I suppose **[that]** *you* **thought** *I* would **forget** *but* **you are** *wrong.* **[WH question** *How*

could I forget a rat like you]. I have sent a letter with all your past details to the chief executive. All your debts and past missdemeanors. If you dont resign from the senate immediately the press will print a list of all your debts both locally and nationally. I also know of the bribes paid to the undesirables to vote for you. Your behavior in Mexico also leaves me wondering why you bothered to go. It certainly was,nt to pay homage to the dead as you were never sober. [WH question *What would the locals think of that]. I also know of the money you have pocketed from the money raised for the celebrations you might be able to fool some people but not me. You forget [that] I have known you for all of your life. Go back into your rat hole where you belong and stay there.*

The examples show how Dimension 1 features like pronouns, verbs, contractions or questions increase as the personal knowledge between addressor and addressee increases from one letter to the next in descending order. This increase of Involved features could be caused by the increased likelihood that two interactants that personally know each other use language to deal with interpersonal matters rather than with the transmission of abstract information. On the other hand, Informational features that convey more complex information, such as nouns and adjectives, are far more common in texts in which there is no relationship between addressor and addressee, such as the letter addressed to the FBI above, than in texts in which addressor and addressee are acquainted because shared background knowledge between interactants allows a more frequent use of deictic features, such as in the letter above addressed to *Bill*.

Direction of Harm

Texts in which the harm was directed to a third party had a higher score on Dimension 2, the Dimension of Narrative vs Non-Narrative Concerns and, consequently, a higher frequency of third person pronouns and a lower frequency of second person pronouns. Figure 5 shows these relationships graphically.

Figure 5 - Boxplots describing the distribution of Dimension 2 (top left), second person pronouns (top right) and third person pronouns (bottom left) for the direction of harm categories
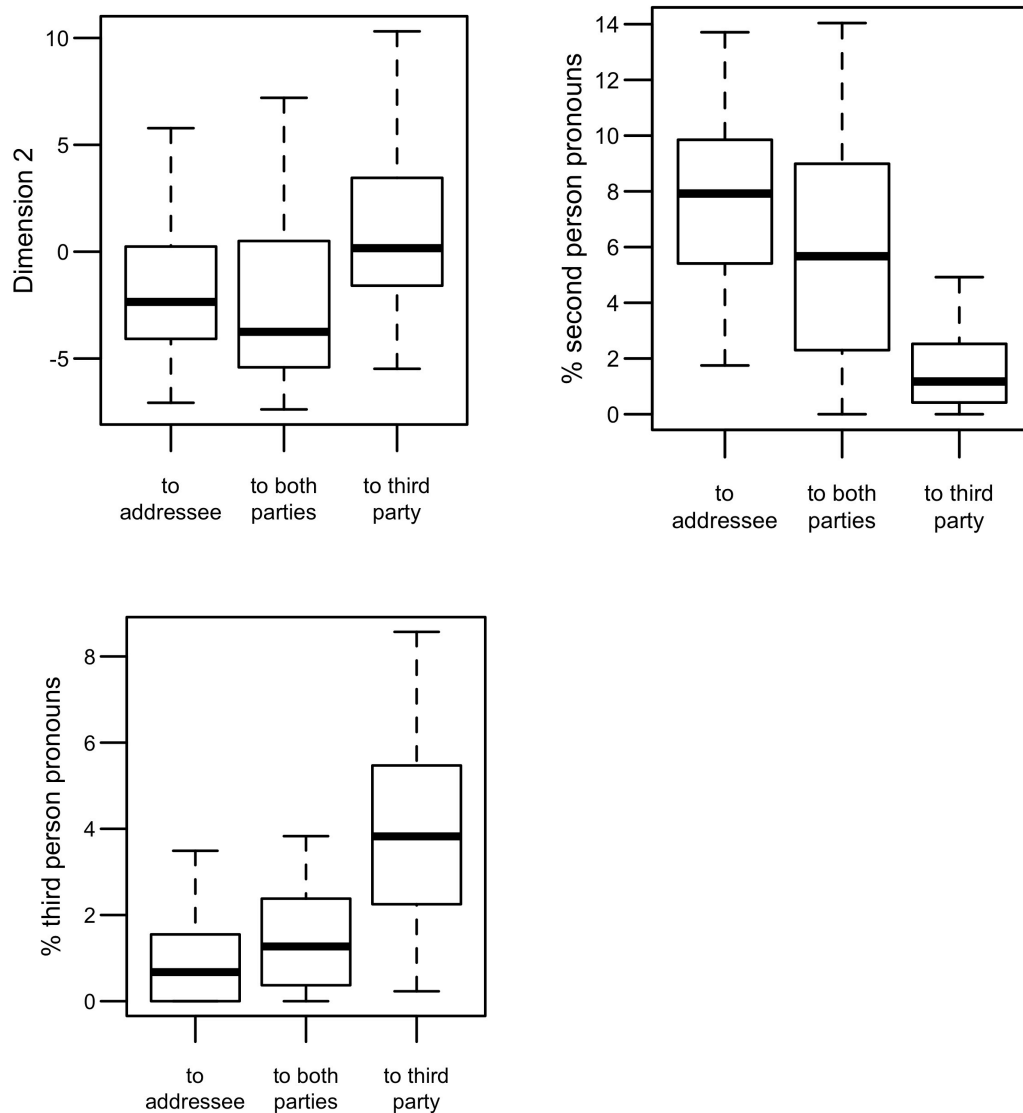
The visual inspection of the boxplots confirms that the differences are in the predicted directions, with 'harm-to-addressee' texts having more second person pronouns and few third person pronouns than 'harm-to-third party' texts whereas the category 'both' is situated in the middle. The distribution of Dimension 2 indicates that the texts in which the threat or abuse was directed towards a third party other than the addressee were more narrative than other MFTs. This pattern is linked to the pattern of pronouns, as the frequency of third person pronouns is a feature included in the calculation of Dimension 2. This pattern can be explained by the fact that authors of the texts in which the harmful content is directed towards a third party are likely to recount events to the addressee regarding the third party to whom the threat or abuse is addressed and to use more third person pronouns in general when talking about the third party. In Table 6 one text with harm directed to an addressee and one

text with harm directed to a third party are reproduced with their language marked for Dimension 2 features.

Table 6 - Two samples of text marked for their Dimension 2 features (third person pronouns, past tenses, public verbs, synthetic negations). The features are in bold. The language of the texts is reproduced as in the original, including the non-standard features.

Letter collected from the FBI Vault containing harm directed to the addressee

Dimension 2 = -5

---

*Sir: I **grew** up in N.C., still have many friends there. It is, however, most unfortunate that you **were** elected as a Senator from N.C. You are, unquestionably, the worst Senator ever elected by N.C.*

*Thank God, however, the tide is turning. With the help of distingueshed Senators like Bob Packwood of Oregon or Goldwater of Arizona, not to **mention** Lowell Weicky of Comm., You are being **revealed** as the #1 scoundrel of all times.*

I wish someone could could push you down some stairs and break your neck.

PS – Real Time magazine of Sept. 82, pages 12 and 13. You will find your pictures on page 13, the unlucky page (because it has your ugly face on it)

To hell with Helms!

Letter collected from Pennebaker (2011: 255) containing harm directed to a third party

Dimension 2 = 5.04

Ms. Livington:

*I think you should know that David Simpson **has perpetuated** the idea that you have **no** credibility among your colleagues. **He says** you **altered** depositions and **falsified** expense reports at your last job in New York. **He says** this is the reason you **left** so abruptly.*

***He has spread** these stories to people in various departments, including Billing, Personnel, Public Relations and to those at the executive level. It is uncertain how and when our senior partners will deal with this. But if you start getting the cold shoulder, you will know why.*

*When I first **heard** of this, I **was** surprised, but **took** what **he said** at face value. Of course, this **was** before I **learned** of **his** voracious appetite for propagating half-truths, gossip, and outright lies, all in the name of somehow making **himself** look knowledgeable and "better."*

*Such a pity. **He** obviously has talent, but it is all negated by **his** vile, malicious tongue. All I can think of is a tremendous sense of insecurity. But I digress. I just **thought** you would like to know.*

A friend

The samples above show how texts in which the harmful content, that is, the threat, the abuse or the malicious information, is addressed to a third party are more narrative and refer to the third party more often. For a text to address the harm to a third party, the reference to a third person is necessary. However, the analysis reveals that texts with harm directed to a third party also tend to provide further details regarding the third party in the form of a narrative.

Mode

The only variable differently distributed across mode types was average word length, as shown in Figure 6.

Figure 6 - Boxplots describing the distribution of average word length for typed and handwritten texts
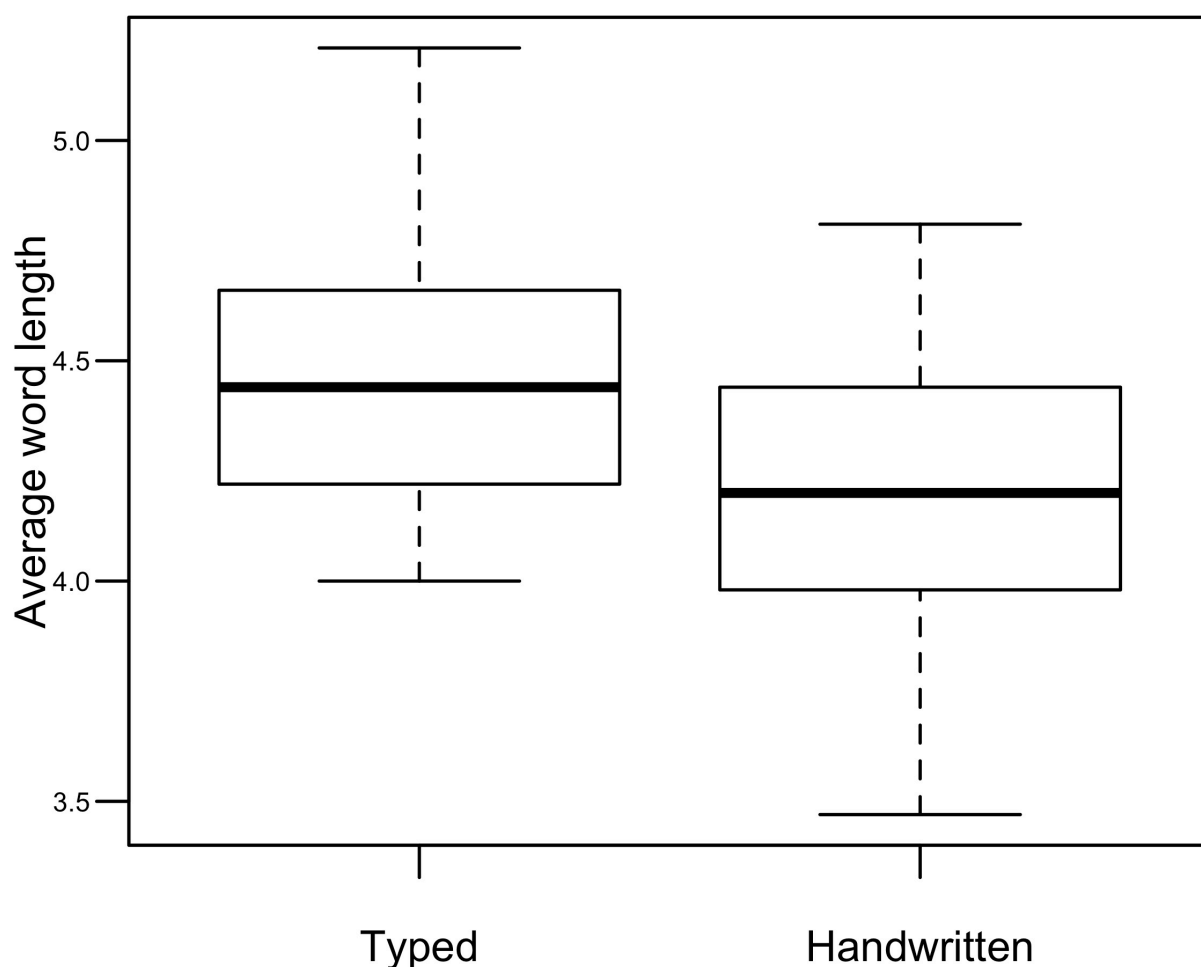
Handwritten texts presented a shorter average word length than typed texts and this could be due to the highest level of strain experienced by the writers as well as to the fact that writers of handwritten texts are less likely to carefully edit their text. This difference, however, could be the result of other effects as there is evidence that average word length is affected by education level (Bromley, 1991; Berman, 2008; Nini, 2015b). Another explanation could therefore be that only relatively well-educated individuals had access to a typewriter or a computer or that educated individuals felt the need to use typewriters or computers to write letters. More experimental research is needed to explain the effect of Mode on average word length.

MFTs and other registers of the English language

Since Biber's (1988) Professional and Personal Letters are the closest registers in terms of communicative situation to MFTs, a comparison can be made between the average

scores of the six Dimensions for these two registers and the average scores of the six Dimensions for the MFT corpus. These comparisons are presented in Table 7.

Table 7 - A comparison of Dimension scores between MFTs and Biber's (1988) Professional and Personal Letters

	Dimension 1 Involved vs Informational discourse	Dimension 2 Narrative vs Non-Narrative Concerns	Dimension 3 Context-Independent Discourse vs Context-Dependent Discourse	Dimension 4 Overt Expression of Persuasion	Dimension 5 Abstract vs Non-Abstract Information	Dimension 6 On-line Informational Elaboration
Personal Letters	19.5	0.3	-3.6	1.5	-2.8	-1.4
MFTs	1.50	-0.68	2.63	3.94	0.30	0.14
Professional Letters	-3.9	-2.2	6.5	3.5	0.4	1.5

As expected, in all the Dimensions MFTs are located in between Personal and Professional Letters except for Dimension 4, where MFTs score higher than either of the two letter registers examined by Biber (1988). Given that the analysis above has shown that threatening letters are distinguished by a higher score on Dimension 4, it is likely that the presence of threatening texts, absent from Biber's (1988) corpus heavily contributes to this score. Even though this finding was somewhat expected, these results quantify exactly the linguistic closeness of MFTs to non-malicious letters.

The distribution of text types produced by the analysis carried out with MAT was: 47% Involved persuasion; 25% General narrative exposition; 9% Informational interaction; 8% Imaginative narrative; 7% Scientific exposition; 4% Learned exposition; 1% Situated reportage. The most common text type was therefore the Involved Persuasion text type, which was found by Biber (1989) to be the most distinctive and common text type for professional letters. This text type is characterised by a high score on Dimension 4, the Dimension of modality and of expression of persuasion.

Table 8 compares the distribution of text types for the MFT corpus and the distributions for Personal and Professional Letters found by Biber (1989) and confirms that MFT on average exhibit a register that is linguistically similar to professional letters.

Table 8 - Distribution of text types for the MFT corpus and for Biber's (1989) registers Personal Letters and Professional Letters.

Rank	MFTs	Professional Letters	Personal Letters
Informational Interaction	9%	10%	50%
Involved Persuasion	47%	40%	33%
Imaginative narrative	8%	0%	17%
General Narrative Exposition	25%	20%	0%
Learned Exposition	4%	30%	0%
Scientific Exposition	7%	0%	0%
Situated Reportage	1%	0%	0%

In conclusion, these findings suggest that MFTs are a special type of involved persuasive letter characterised by high scores on Dimension 4, the dimension of Overt Expression of Persuasion. Malicious texts are, on average, texts that are more Involved than many other written registers while at the same time focused on conveying information to or making a transaction with the addressee, whether this is abuse, a threat or an attempt to distribute malicious information. This transaction is, however, realised by strong expressions of overt persuasion which are typically manifested linguistically by high scores on Dimension 4, especially for those texts that contain a threat.

Conclusions

The aim of the present paper was to shed light on the register identity of MFTs which involve threat, abuse or spread of malicious information. The findings of the study presented in this paper are two-fold, as conclusions regarding the linguistic variability of MFTs can be drawn both at a general level of comparison with other English registers and at an internal level in terms of situational parameters.

On one hand, the analyses above have concluded that the register of MFTs is characterised by an involved and persuasive yet transactional character and that English MFTs can be located within the typical text types of professional letters: Involved Persuasion. Furthermore, this study has found that Dimension 4, Overt Expression of Persuasion, is a useful Dimension of variation that can distinguish threatening from non-threatening texts, thus confirming Gales' (2010) findings. Even though, as Fraser (1998) admits, it might be impossible to determine from the language alone whether an utterance is threatening, the

research presented in this paper suggests that it is indeed possible to understand with some degree of certainty whether a particular text is a malicious forensic threatening text based on an assessment of how modality and persuasive means are used.

In terms of MFTs' internal linguistic variation, the analyses above have suggested that the most relevant Dimensions of variation for the MFTs are Dimension 1, Dimension 4, and, to some extent, Dimension 2. Dimension 1, the Dimension of Involved vs Informational features, presented a strong pattern that was shown to be linked to the interpersonal distance or personal knowledge between interactants and seems, therefore, to account for the *personal vs impersonal* continuum of MFTs. Dimension 4, the Dimension of expression of persuasion and modality, was found to be a good discriminator of threatening and defaming texts and seems therefore to account for the *threatening vs non-threatening* continuum of MFTs. Finally, Dimension 2, the Dimension of narrativity, was found to distinguish texts in which the harmful content is solely directed to the addressee from texts in which the harmful content is directed towards a third party. This Dimension would thus correspond to the degree of *direction of harm to third party* of a MFT. It follows that the most important situational parameters that can be used to describe MFTs are the parameters that correspond to these three clines of variation: personal knowledge, presence or absence of a threat and direction of harm.

These three clines thus form a space in which any MFT can be located similar to the graph presented in Figure 7. This model not only represents an abstract summary of the findings of the present study but also constitutes a preliminary model that explains the most common patterns of linguistic and situational variation in MFTs.

Figure 7 - Graph representing a schematic a model that represents the variation of MFTs according to situational parameters and Biber's (1988) Dimensions of linguistic variation
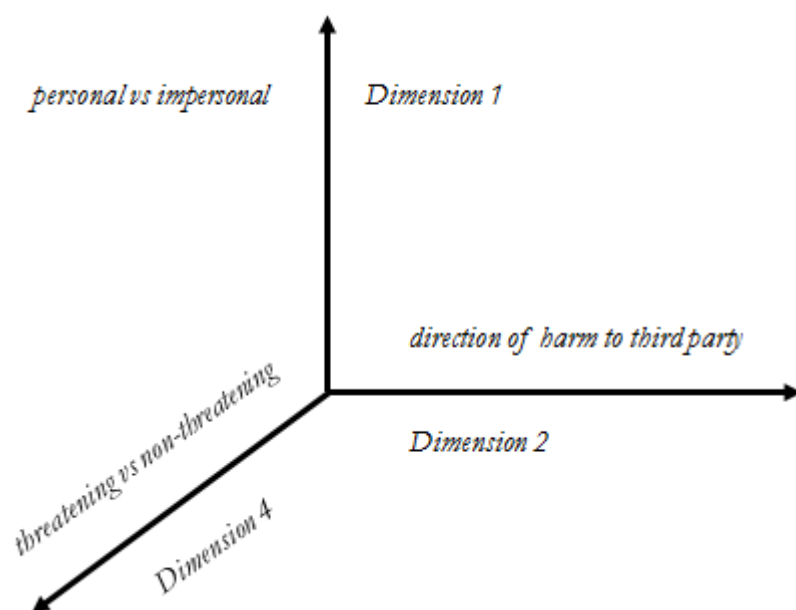
In conclusion, this study shows that it is possible to establish clines of variations across Dimensions that distinguish different types of MFTs and that correspond to variation in terms of their situational parameters. The importance of these findings for forensic linguistics lies in its possible applications for various domains. For forensic authorship analysis, this work could represent a start towards creating a model in which forensic linguists can operate when dealing with questioned MFTs, as knowledge of expected variation for frequencies of features can help to establish population-level distinctiveness (Grant, 2010; Nini and Grant, 2013). For forensic linguists, judges, or legal professionals working on cases involving the question of whether a malicious act was or was not carried out, knowledge of the register identity of MFTs can help informing such decisions. This research also makes a step forward in bridging the gap between forensic linguists and professionals working on threat assessment. Finally, research of this type can generally increase our understanding of crimes committed using language, which can in turn affect law and policy makers.

In general, however, although a replication of this study using more data is needed, a better confirmation and expansion of the present study would be an experiment in which the conclusions proposed by this paper are empirically tested on samples that are controlled for other types of variation. These types of experiments are needed in order to reach safer

conclusions and to develop a better working model that can predict and explain the language of malicious communication.

References

Baayen, H. (2008) *Analyzing Linguistic Data: A Practical Introduction to Statistics using R*. Cambridge: Cambridge University Press.

Bell, A. (1984) Language style as audience design. *Language in Society* vol. 13: 145–204.

Berman, R. (2008) The psycholinguistics of developing text construction. *Journal of Child Language* vol. 35: 735–771.

Biber, D. (1989) A typology of English texts. *Linguistics*, vol. 27: 3–43.

Biber, D. (1994) Register and social dialect variation: An integrated approach. In D. Biber and E. Finegan (eds) *Sociolinguistic Perspectives on Register* 315–347. Oxford: Oxford University Press.

Biber, D. (1993) Representativeness in Corpus Design. *Literary and Linguistic Computing* vol. 8: 243–257.

Biber, D. (2006) *University Language: A Corpus-Based Study of Spoken and Written Registers*. Amsterdam: John Benjamins.

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

Biber, D. and Conrad, S. (2009) *Register, Genre, and Style*. Cambridge; New York: Cambridge University Press.

Biber, D. and Jones, J. (2005) Merging corpus linguistic and discourse analytic research goals: Discourse units in biology research articles. *Corpus Linguistics and Linguistic Theory* vol. 1: 151–182.

Bromley, D.B. (1991) Aspects of written language production over adult life. *Psychology and Aging* vol. 6: 296–308.

Coulthard, M. and Johnson, A. (2007) *An Introduction to Forensic Linguistics*. London: Routledge.

Dancey, C. P. and Reidy, J. (2011) *Statistics without Maths for Psychology*. Harlow: Prentice Hall.

FBI, FBI Vault. Available at: http://vault.fbi.gov/ [Accessed January 1, 2012].

Fraser, B. (1998) Threatening revisited. *Forensic Linguistics* vol. 5: 159–173.

Gales, T. (2011) Identifying interpersonal stance in threatening discourse: An appraisal analysis. *Discourse Studies* vol. 13: 27–46.

Gales, T. (2010) Ideologies of Violence: a Corpus and Discourse Analytic Approach to Stance in Threatening Communications. Unpublished PhD thesis.

Gales, T. (2015) The stance of stalking: a corpus-based analysis of grammatical markers of stance in threatening communications. *Corpora* vol. 10: 171–200.

Grant, T. (2008) Approaching questions in forensic authorship analysis. In J. Gibbons and M. T. Turell (eds) *Dimensions of Forensic Linguistics* 215–231. Amsterdam: John Benjamins Publishing Company.

Grant, T. (2010) Txt 4n6: Idiolect free authorship analysis. In Coulthard, M. (ed.) *Routledge Handbook of Forensic Linguistics* 508-523. London: Routledge.

Gray, B. (2013) More than discipline: Uncovering multi-dimensional patterns of variation in academic research articles. *Corpora* vol. 8: 153–181.

Grieve, J., Biber, D. and Friginal, E. (2011) Variation among blogs: A multi-dimensional analysis. *Genres on the Web* vol. 42: 303–322.

Kaplan, J. P. (2016) Case report: Elonis v. United States. *The International Journal of Speech, Language and the Law* vol. 23(2): 275–292.

Kelleher, M. D. and Van Nuys, D. (2002) *'This is the Zodiac Speaking': Into the Mind of a Serial Killer*. Westport, Conn.; London: Praeger.

Kniffka, H. (2007) *Working in Language and Law: A German Perspective*. Basingstoke, New York: Palgrave Macmillan.

McMenamin, G.R. (2002) *Forensic Linguistics: Advances in Forensic Stylistics*. Boca Raton, Fla: CRC Press.

Nini, A. (2015a) *Multidimensional Analysis Tagger* 1.3 - Manual. Available at: https://sites.google.com/site/multidimensionaltagger/.

Nini, A. (2015b) *Authorship Profiling in a Forensic Context*. Unpublished PhD dissertation. Aston University, UK.

Nini, A. and Grant, T. (2013) Bridging the gap between stylistic and cognitive approaches to authorship analysis using Systemic Functional Linguistics and multidimensional analysis. *International Journal of Speech Language and the Law*, vol. 20(2): 173–202.

Napier, M. and Mardigian, S. (2003) Threatening messages: The essence of analyzing communicated threats. *Public Venue Security*. September/October: 16-19.

Olsson, J. (2003) *Forensic Linguistics: An Introduction to Language, Crime and the Law*, London: Continuum.

Pennebaker, J. W. (2011) *The Secret Life of Pronouns: What our Words Say about us*, New York; London: Bloomsbury Press.

Searle, J. (1979) *Expression and Meaning: Studies in the Theory of Speech Acts*, Cambridge: Cambridge University Press.

Shuy, R. (2010) *The Language of Defamation Cases*, Oxford: Oxford University Press.

Solan, L. and Tiersma, P. (2005) *Speaking of Crime: The Language of Criminal Justice*, Chicago: University of Chicago Press.

Tiersma, P. (1987) The language of defamation. *Texas Law Review* vol. 66: 303–350.

Titak, A. and Roberson, A. (2013) Dimensions of web registers: An exploratory multi-dimensional comparison. *Corpora* vol. 8: 235–260.

Toutanova, K., Klein, D., Manning, C., and Singer, Y. (2003) Feature-rich part-of-speech tagging with a cyclic dependency network. *Proceedings of HLT-NAACL* 2003. 252–259.