

UNIVERSITY OF HELSINKI

Metadiscourse in Online Opinion Texts

Exploring variation within a genre

Ylva Biri
MA Thesis
English Philology
Department of Modern Languages
University of Helsinki
October 2017



| | | | |
|--|--|--|--|
| Tiedekunta/Osasto – Fakultet/Sektion – Faculty Humanistiska fakulteten | | Laitos – Institution – Department Institutionen för moderna språk | |
| Tekijä – Författare – Author Ylva Irene Biri | | | |
| Työn nimi – Arbetets titel – Title Metadiscourse in Online Opinion Texts: Exploring variation within a genre | | | |
| Oppiaine – Läroämne – Subject Engelsk filologi | | | |
| Työn laji – Arbetets art – Level Pro gradu-avhandling | | Aika – Datum – Month and year oktober 2017 | Sivumäärä– Sidoantal – Number of pages 74 + 8 sidor |
| <p>Tiivistelmä – Referat – Abstract</p> <p>Syftet för denna avhandling är att analysera användning av metadiskurs i engelskspråkiga blogginlägg och debattartiklar från tidningars webbsajter och nyhetssajter. Metadiskurs är ett teoretiskt ramverk som omfattar språkliga konstruktioner som strukturerar eller kommenterar diskursen i en text. Hur metadiskurs används beror på textens genre och målgrupp. Tidigare forskning, som för det mesta är kvalitativ, har visat att metadiskurs används i argumenterande text, dels för att göra skribentens synpunkt logisk och lättförståelig, dels för att vädja till läsaren på ett mera direkt sätt. I denna avhandling studeras hur olika sorters typer av metadiskurs korrelerar med varandra i argumenterande text, d.v.s. vilka konstruktioner förekommer tillsammans inom samma text.</p> <p>Som metod används multidimensionell analys, en korpus-driven kvantitativ metod som bygger på faktoranalys. Som korpus används en samling av engelskspråkiga blogginlägg och debattartiklar på sammanlagt dryga 285 000 ord som annoterats med 13 metadiskurskategorier. Resultaten utgörs av fyra statistiskt signifikanta faktorer som kan användas för att beskriva vilka retoriska strategier som används i texterna för att kommentera textuella element samt skribentens eller läsarnas närvaro. De fyra faktorerna är nämnda enligt sin funktion: (1) Skribent-orienterad text, (2) Solidaritet, (3) Läsar-orienterad text, och (4) Intertextualitet.</p> <p>Analysen jämför genrerna blogg, nyhetssajt och tidningars webbsajt ur en genreanalytisk synvinkel. Data tyder på variation mellan skribenter inom samma genre, men artiklar på webbsajter för tidningar som också har en pappersversion innehåller i allmänhet endast litet metadiskurs, främst för att syfta till citerade källor. Däremot brukar nyhetssajter utan pappersversion göra referens till läsaren antingen genom pronomenet <i>you</i> eller genom pronomenet <i>we</i> som i kontext kan förstås innefatta också läsaren. Bloggare syftar oftare till sig själva än till sina läsare och använder sällan det inklusiva pronomenet <i>we</i>. Genrerna bloggar och nyhetssajter innehåller också minst explicita källhänvisningar.</p> <p>Denna studie ett exempel på hur metadiskurs kan forskas inom en multidimensionell analys. Även en analys på mindre data tillbringar information som kan användas för att utveckla det teoretiska metadiskurs ramverket.</p> | | | |
| Avainsanat – Nyckelord – Keywords metadiskurs, argumenterande text, nyhetssajt, bloggar, multidimensionell analys, genreanalys | | | |
| Säilytyspaikka – Förvaringställe – Where deposited Helsingfors universitet campusbibliotek, E-thesis http://ethesis.helsinki.fi | | | |
| Muita tietoja – Övriga uppgifter – Additional information | | | |

Table of contents

| | | |
|-------|--|----|
| 1 | Introduction..... | 1 |
| 2 | Metadiscourse | 3 |
| 2.1 | The scope of metadiscourse..... | 4 |
| 2.2 | What metadiscourse is not..... | 10 |
| 2.3 | Summary of the concept of metadiscourse..... | 15 |
| 3 | Genre and Register Analysis..... | 16 |
| 3.1 | Concepts of genre and register analysis | 17 |
| 3.2 | Online opinion texts as a genre..... | 18 |
| 4 | Material and Methods | 23 |
| 4.1 | Compiling the corpus | 24 |
| 4.2 | Annotation of metadiscourse | 26 |
| 4.3 | Multi-Dimensional Analysis | 31 |
| 4.3.1 | Multi-dimensional Analysis as an empirical method ... | 32 |
| 4.3.2 | Conducting a Multi-Dimensional Analysis | 35 |
| 5 | Findings..... | 40 |
| 5.1 | Interpretation of dimensions..... | 41 |
| 5.1.1 | Dimension 1: Writer-oriented text..... | 42 |
| 5.1.2 | Dimension 2: Solidarity..... | 43 |
| 5.1.3 | Dimension 3: Reader-oriented text..... | 45 |
| 5.1.4 | Dimension 4: Intertextuality | 46 |
| 5.2 | Other metadiscoursal markers | 48 |
| 5.3 | Characterising Opinion Texts..... | 50 |
| 5.3.1 | Overview of frequencies and validity issues | 50 |
| 5.3.2 | The prototypical Opinion Text | 52 |
| 5.3.3 | Opinion Texts across dimensions..... | 55 |
| 5.4 | Summary of findings | 65 |
| 6 | Conclusion and Future Considerations | 68 |
| | References | 71 |
| | Appendix 1: Data sources of corpus | 75 |
| | Appendix 2: Frequencies and variation of metadiscourse markers | 78 |
| | Appendix 3: List of potential metadiscourse markers..... | 79 |

List of Tables

| | |
|---|----|
| Table 1: Summarising and comparing metadiscourse frameworks by Hyland (2005) and Dafouz-Milne (2008)..... | 8 |
| Table 2: Non-metadiscoursal items..... | 15 |
| Table 3: Metadiscourse marker categories..... | 30 |
| Table 4: Eigenvalues and variance explained in unrotated factor analysis.. | 38 |
| Table 5: Extracted factors and loadings above 0.1 | 41 |

List of Figures

| | |
|---|----|
| Figure 1: Scree plot for Principal Component Analysis | 37 |
| Figure 2: Texts in corpus plotted on Dimensions 1 and 3..... | 56 |
| Figure 3: Source dimension scores plotted on Dimensions 1 and 3 | 60 |
| Figure 4: Texts in corpus plotted on Dimension 2..... | 61 |
| Figure 5: Texts in corpus plotted on Dimension 4..... | 63 |

1 Introduction

Non-fictional persuasive texts are communication events between two parties, the author and the audience, but there are many things the author has to consider for the persuasion to be effective. Besides presenting the author's own opinion, the text should also engage the audience by involving them in the argument, as into a dialogue, even if the writer has never met nor will never meet their audience face to face. The manner of presentation is also important as especially lengthy and complex texts can become difficult to follow without coherent order and explicit signposting. As noted by Mauranen (1993:34), persuasive texts benefit from a *rhetorical strategy* where the author takes into consideration both the presentation and the argumentation. Having chosen a rhetorical strategy, writers use a set of linguistic devices to make their text effective to their purpose and to create an authorial voice that will be heard, understood and accepted by the intended audience. Linguistic devices that writers can use to organize the text and to interact with readers have been the interest of many scholars, and one framework for describing said devices is that of *metadiscourse*.

Metadiscourse is a rhetorical and pragmatic phenomenon not limited in linguistic form (Hyland, 2005:25). Metadiscourse markers frame the propositional content of the text by paving the way for the reader's comprehension: they remind the reader of earlier ideas, explain new concepts, soften a claim, express an opinion and anticipate the reader's reply. The amount of metadiscourse varies depending on the context, the purpose of the text and, consequently, the *genre* of the text. For example, Hyland's (2005) framework for analysing metadiscourse has been developed especially for academic written English, where many complex ideas and concepts need to be made accessible and where social relations must be negotiated (Hyland, 2005:66). On the other hand, metadiscourse is needed to build credibility also in shorter persuasive writing. According to Mauranen (1993:167), readers feel that metadiscourse makes a text feel not only clearer but more convincing and authoritative. Therefore, argues Dafouz-Milne (2003:33–34), all metadiscourse has a persuasive function, even though different metadiscourse markers have different degrees of explicitness. Crismore et al. (1993) have studied metadiscourse in persuasive essays by university students. Dafouz-Milne (2003) and Fu and Hyland

(2014) have shown that metadiscourse is effectively used in newspaper opinion articles, also persuasive by nature. Thus, persuasive writing is a relevant and viable focus of metadiscourse and genre analysis.

The focus of this thesis is metadiscourse in persuasive writing in written pieces focusing on current affairs in society and politics, but contrary to most previous studies it also uses data from smaller and less formal published sources. To be specific, the present empirical study employs a small-scale corpus that consists of texts uploaded onto personal opinion blogs (truncation of “web-logs”), columns published by columnists in newspapers, and op-eds (from “opposite the editorial page”) written by journalists of a newspaper, who as per the are not affiliated with the editorial board (see Rivers et al., 1988). Here the word *text* refers to the written part, the body of the publication, excluding multimedia. Online publishing is a central way of sharing thoughts on matters ranging from mundane to controversial. Although the chosen texts are similar in mode, topic and aim, blogs especially give freedom to the author’s personal style and format. When the aim is to persuade, writers must make the reader informed and involved in the matter. Beyond the mere general description of metadiscourse in the chosen genre, the exploratory aim is to determine how much variation exists within a single, seemingly homogenous genre: after all, all the chosen genres are opinion texts on politics published online to be read by the mass audience. I am also interested in how the uses of different types of metadiscourse correlate with each other: does the frequency of one type predict the frequency of another and what does this mean for the particular text? The research questions are thus the following:

1. For what purposes is metadiscourse used in online opinion texts?
2. How do different metadiscourse marker categories co-occur with each other in online opinion texts?

Insight in co-occurrence and rhetorical functions of metadiscourse markers can offer deeper understanding in different types of metadiscourse markers and thus also be used to develop existing frameworks of metadiscourse. To analyse the co-occurrence of different metadiscourse marker categories, this study will use Multi-dimensional analysis (MD analysis, see Biber, 1988), a quantitative corpus-driven method to analyse correlation patterns. This method has had little use in previous research on metadiscourse or other pragmatic functions, but it is ideal for the present study as it facilitates the detection of patterns across several variables. A MD analysis groups

variables, such as metadiscourse markers, into clusters of co-occurring variables. Variables belonging to the same cluster, or *dimension*, can be interpreted as sharing a rhetorical function. When multiple dimensions are found, the rhetorical function of the text can be analysed from multiple perspectives. The quantitative statistical analysis is then complemented by a more qualitative reading of the individual metadiscourse markers. By using this method for studying metadiscourse, the present study hopes to look at metadiscourse in opinion writing from a perspective unavailable to more qualitative methods.

The definition of metadiscourse along with theoretical background is given in Chapter 2. The genre of persuasive writing is explored in Chapter 3, and the data for the present study is presented in Chapter 4 along with an overview of MD analysis as a method. The results are examined in Chapter 5, with conclusions and suggestions for further research given in Chapter 6.

2 Metadiscourse

As several definitions of metadiscourse are in existence, it is important to declare what one means by the term. Mauranen (1993) and later Ädel (2006) suggest that models of metadiscourse can be divided into two types of approaches. According to the first approach, metadiscourse is concerned with rhetorical devices used for organising a text for the reader's benefit. The author can facilitate their reader's comprehension by using, e.g. logical connectives, code glosses, or references back to the text in order to connect and explain ideas. As in Halliday's and Hasan's (1976:6) description of *cohesion*, conjunctions are used to point out the relationships between sentences or paragraphs. This kind of metadiscourse has been italicised in the example (1) below¹:

- (1) "It is possible that the powers that be on the Right in the Conservatives wanted Mrs May all along *though* this requires a belief in a conspiracy theory that is improbable. *However* Mrs May, *though* supporting "Remain", had been invisible during the Referendum campaign." (article017)

Here the adjuncts *though* and *however* combine and contrast ideas. Models of metadiscourse concerned with the way language is used to organise a text are referred to as *reflexive models* (Ädel & Mauranen, 2010). As reflexive models limit

¹ The convention of marking metadiscourse in examples with *italics* will be used throughout this study. The code at the end of each example (e.g. "article017") marks from which text in the used corpus the example is taken. A list of codes and sources for the corpus is given in Appendix 1.

the notion of metadiscourse to linguistic text-organising devices, Ädel (2006:177) has termed the definition as *narrow approach* to metadiscourse.

While reflexive models are useful for describing textual organisation and cohesion, the objectives of this study are to describe both the textual organisation and the writer-reader interaction. For this another approach of metadiscourse is required, namely an *interactive model* (Ädel, 2010). Following a *broad approach* (coined by Ädel, 2006:168)², interactive models encompass the text-organising features, but beyond this they include as equally important the features used by the writer or speaker to persuade or otherwise interact with the reader. Such features include direct addresses towards the reader or attitudinal commentary by the author, both of which are used to engage the reader in a dialogue. Consider example (2):

- (2) “If *you* want to close the gap between the super rich and the rest of *us*, *you need to* consider a very different tax bracket structure. And *check out* that marriage penalty for the upper classes!” (article362)

Here personal pronouns make the reader a participant in the hypothetical event. The semi-modal *need to* and the directive *check out* also engage the reader to become involved in the processing of ideas and information.

As this study focuses also on the writer/reader-relationship, it follows the broad approach, which will be described more closely below. The following sections will first consider metadiscourse in relation to other fields of analysing reader-writer interaction, and then introduce frameworks and taxonomies for interactive models of metadiscourse relevant for the present study.

2.1 The scope of metadiscourse

Metadiscourse has often been described as “text about text” or “discourse about discourse” (Hyland & Tse, 2004:156; Ädel & Mauranen, 2010:1). It can certainly be thought of as simply discourse that discusses text or discourse on a meta-level, but in reality, metadiscourse is a more complex phenomenon. Hyland and Tse (2004:157) define the scope of their broad, interactive model as “the author’s linguistic and rhetorical manifestation in the text,” as a way of creating a reader-friendly discourse that conveys the author’s opinions and credibility. Features of metadiscourse that seek to express the author’s own attitude or to persuade the reader do not merely add

² Mauranen (1993), on whose work Ädel (2006) bases her own division, refers to the narrow and the broad approach as *non-integrative approach* and *integrative approach* respectively, the integrative approach integrating both the text organising and the attitudinal aspects of metadiscourse. In this study, Ädel’s terminology is used.

non-propositional information to the text: they function as writer-reader interactions. The definition of metadiscourse as “text about text” becomes problematic, as not all features that seek to engage the reader refer back to the text, but instead directly address the reader and the reader’s assumed knowledge, background or preconceived opinions on the topic of the text. Or as Hyland and Tse put it, rather than representing writer’s self-awareness of *text*, metadiscourse “represents the writer’s awareness of the unfolding text as *discourse*”, involving also the reader (Hyland & Tse, 2004:167; italics original). Metadiscourse can thus be seen as an aspect of authorial voice. The concept of authorial voice, when considered social or dialogic, is concerned with writer’s presence in relation to the audience (Tardy, 2012).

Metadiscourse has been studied since the late 1970’s (Boggel, 2009:11), but the linguistic features and rhetorical concepts it encompasses have been studied much earlier. Already in his 1956 model of the functions of language, Roman Jakobson (1985) wrote of the concept of *metalanguage* in linguistics, borrowing the term from the study of logic and mathematics. In linguistics, the function of metalanguage is to talk about language itself, about linguistic elements including sound, structure and meaning. Another function Jacobson (in Ädel, 2006:164) suggests is the expressive function in language where the writer’s or the speaker’s presence can be noticed, e.g. in first- and second-person pronouns or imperative clauses. The expressive function exists also for phrases such as *to summarize*, where the author is only implicitly recognized as the agent. In another early study of metadiscourse, Schiffrin (1980) writes about *meta-talk* having a twofold purpose: it is both for organizing the text flow and for evaluating the content of it. It thus applies to both the informational and the expressive plane of language. The features of Schiffrin’s framework include verbs of saying, operators that modify propositions, and noun phrases that refer to different sections of text.

A framework of metadiscourse that has been highly influential and widely adapted in later studies is one proposed by Vande Kopple (1985). The model follows Joseph Williams (Williams, 1981; in Vande Kopple, 1985) in treating metadiscourse as text about text and suggests that it does not expand propositional information but is concerned with organising the text and the presence of the author and the reader in the text (1985:83). Vande Kopple (1985:83–84) shows how a writer can use linguistic devices to connect or explain ideas, to remind the reader of previous sections of text, or to spell out discourse goals. The writer can also guide a reader’s

understanding and assessment using validity markers which estimate the trustworthiness of a proposition, e.g. *clearly* or *perhaps* (ibid.:84).

As metadiscourse encompasses both devices for organising and evaluating, most frameworks draw a line between these two functions. Vande Kopple. (1985:86) categorises the devices as marking either *textual* or *interpersonal metadiscourse*. His division echoes the *textual* and *interpersonal metafunctions* presented in M.A.K. Halliday's Systemic Functional Grammar (SFG) (Halliday & Matthiessen, 2014:30). In SFG, the textual metafunction of a clause is concerned with making the discourse flow coherent and continuous; the interpersonal metafunction is concerned with the ways language is used for enacting relationships, encompassing the author's attitudes towards both the reader and the topic at hand. In their study of metadiscourse in persuasive writing by students, Crismore et al. (1993) adapt and reorganize Vande Kopple's (1985) framework but retain the distinction between textual and interpersonal. Hyland (1998) and Dafouz-Milne (2003), who base their models on that of Crismore et al. (1993), also follow the textual/interpersonal distinction. However, Hyland and Tse (2004) emphasise that both textual and interpersonal metadiscourse are concerned with the author's self-awareness of the discourse, that is, with awareness of not only the text but of the text as effective communication, too. Thompson and Thetela (1995) distinguish between reader-friendly (*interactive*) and reader-managing (*interactional*) rhetoric choices. Building on this and on previous models of stance and metadiscourse, Hyland and Tse propose new names for the two dimensions of metadiscourse: the *interactive* and the *interactional* respectively. With interactive metadiscourse the author guides the reader through the content of text and with interactional resources they persuade and "involve the reader in the argument" (Hyland & Tse, 2004:169). Dafouz-Milne (2003:33) points out that textual (i.e. interactive) markers are also concerned with persuasion, only less explicitly than interpersonal (i.e. interactional) markers. Turning to the other half of the interaction, the writer, interactional metadiscourse markers such as attitude and modality markers also form the perceived persona of the writer (Hyland, 2005:67–71). Ädel (2006:88) argues that the writer can choose their writer persona to display parameters such as friendliness, didacticism or professionalism. Writer persona is an important factor in how persuasive the text is to the audience, although the needs of different potential audiences differ. Connecting metadiscourse to Aristotle's rhetoric theory, Crismore and Farnsworth (1989) illustrate how authors use interpersonal

metadiscourse markers to establish and re-establish their *ethos*, i.e. the character or disposition of the speaker in front of the audience. In their study, Crismore and Farnsworth describe the use of metadiscourse by Charles Darwin in his *On the Origin of Species*. They note that Darwin's *ethos* is different from chapter to chapter, but that he often uses metadiscourse to establish a cautious and tactful scientist writer persona in order to retain credibility. A similar tentativeness is retained in academic writing until this day.

The model presented in Hyland and Tse (2004) is a model for metadiscourse in academic discourse especially. Dafouz-Milne (2003, 2008), who has worked with newspaper discourse, has proposed a taxonomy that is largely similar in content: the main differences are in the division of the categories. Table 1 compares the two taxonomies: note that while only macro-categories are listed, both taxonomies also detail subcategories. Some interactional markers are mentioned in both taxonomies but belong to macro-categories with different purposes. Note also how Dafouz-Milne has categories for Sequencers, Topicalisers, Illocutionary markers, Reminders, and Announcements; Hyland and Tse group them as subcategories of Frame markers and Endophoric markers respectively. Meanwhile the function of Hyland and Tse's macro-category of Self-mention is a subcategory of Commentary in Dafouz-Milne's model. Hyland and Tse's Evidentials have about the same function as Dafouz-Milne's Attributors, but Hyland and Tse categorise Evidentials as part of reader-friendly interactive metadiscourse whereas Dafouz-Milne categorises Attributors as part of reader-managing interpersonal metadiscourse, as do Crismore et al. (1993).

| Hyland and Tse (2004) | <i>Metadiscoursal function</i> | Dafouz-Milne (2008) |
|----------------------------|---|------------------------|
| Interactive markers | organize information for reader's benefit | Textual markers |
| Transition markers | signal additive, comparing, or causative relations between propositions (and, furthermore; similarly, in contrast; thus, anyway). | Logical markers |
| Frame markers | segment text by referring to order ("first off", "secondly") | Sequencers |
| | mark topic shifts ("let's return to", "as for") | Topicalisers |
| | explicit discourse goal ("to summarise") | Illocutionary markers |
| Endophoric markers | refer to previous text ("as noted above") | Reminders |
| | refer to future text ("as we'll later see") | Announcements |

| | | |
|--|---|--|
| Evidentials | refer to sources of information to back up a claim (e.g. “according to...”) | Attributors (in Dafouz-Milne an Interpersonal marker category) |
| Code glosses | elaborate or explain (e.g. “in other words”, “for example”) | Code glosses |
| | | |
| Interactional markers | address and engage the reader in discourse | Interpersonal markers |
| Evidentials (in Hyland and Tse an Interactive marker category) | refer to sources of information to back up a claim (e.g. “according to...”) | Attributors |
| Hedges | indicate writer’s hesitation or withholding of commitment (“perhaps”, “might”) | Hedges |
| Boosters | indicate writer’s certainty to truth-value (“clearly”, “this demonstrates”) | Certainty markers |
| Attitude markers | writer’s affective attitude to propositions (e.g. “I prefer”; “unfortunately”; “it is logical that...”) | Attitude markers Commentaries |
| Engagement markers | deontic modals or semi-modals marking necessity | |
| | interact with the reader by directives or questions (“note how”, “right?”) | |
| Self-mention | explicit presence of author (I, my, exclusive we) | |
| N/A | asides, commentary separate from the text flow with dashes or parentheses | |

Table 1: Summarising and comparing metadiscourse frameworks by Hyland (2005) and Dafouz-Milne (2008)

Most studies of metadiscourse exclude any mention of intertextuality from their definition of metadiscourse (Boggel, 2009:34). Ädel (2006) sees metadiscourse only as commentary on the “current text” in question and its self-references, not in how the text refers to other texts. She argues that intertextual references, text about other texts, do not refer to “the ongoing discourse as construed by the current writer” (2006:26), which is essential to her narrow approach definition of metadiscourse. Meanwhile, Boggel proposes a broad approach model where “the interpersonal dimension is assumed to underlie all types of metadiscourse” (2009:62). She argues that since intertextuality is a vital part of textuality, a model of metadiscourse should be applicable to intertextual material.

Both Ifantidou (2005) and Boggel (2009) bypass the division into textual/interactive and interpersonal/interactional metadiscourse entirely by dividing metadiscourse into *intertextual* and *intratextual* metadiscourse. While the current study stems from the interactive/interactional division, it adopts Ifantidou’s and

Boggel's viewpoint on intertextuality as part of metadiscourse. As Ifantidou (2005:1337) argues, intertextual metadiscourse markers convey whether a statement is made by the writer themselves or an external source and, in the case of external sources, how the writer interprets the utterance attributed to the other source. This is a form of conveying attitude and persuasion. Citing authoritative sources can increase the credibility of the author. According to Boggel (2009:43–47), metadiscourse often occurs with propositions from other texts, either to explain, expand or evaluate. She further defines two sub-types of intertextual metadiscourse, intertextual *attributing* metadiscourse links a proposition to a source, whereas intertextual *explicating* metadiscourse provides an explanation or interpretation of an external source. Both types can be expected to be found also in journalistic writing, although Boggel's distinction will not be used in the present study. Hyland and Tse (2004) and Dafouz-Milne (2008), for example do not explicitly comment on the status of intertextuality, but in terms of function, Boggel's attributing and explicating metadiscourse correspond largely to their categories of Evidentials/Attributors and Code glosses.

With metadiscoursal markers referring to elements of a text, Mauranen's study on reflexivity (1993:158) further makes a distinction of markers with high explicitness and markers with low explicitness. While Mauranen discusses explicitness as either low or high level, she considers it a continuum that denotes how markers referring to the text at hand acknowledge the different elements of the text. Markers with high explicitness make mention of "text as text (as opposed to, say, a study, a theory, or an argument)" and uses expressions such as "*this section*", or "*we shall explore*" (1993:172). Elements of the communication process are thus labelled and referred to explicitly. Meanwhile, markers with low explicitness do not explicitly refer to said elements, they connect ideas using expressions such as *additionally* and refer to sections in text as *above* or *here*. The reference point is merely understood to be an element of the communication event. As Mauranen (1993:186) notes, sometimes the reference point is not actually an element of the textual act per se, but rather of an argument act. However, she also notes that there is a blurred distinction between markers referring to or connecting textual elements and markers referring to or connecting real-world elements (1993:180). The distinction between text and the real world is crucial for a definition of metadiscourse to be

workable. Section 2.2 will provide further description on what cannot be counted as metadiscoursal.

2.2 What metadiscourse is not

As previously described, metadiscourse is a set of linguistically and overtly expressed functions. Thus, the scope of a metadiscourse does not extend to typography, hyperlinks, emoticons or images – although some researchers have included typographical markers in their definitions (Ädel, 2006:28). Neither is metadiscourse traditionally concerned with multimodal features, such as layout or the content of images, although this too can share the functions of metadiscourse and would certainly be interesting to analyse. Yet even after having restricted metadiscourse to the linguistic, both Hyland (2005:17) and Ädel (2006:22) note that metadiscourse is a functional category lacking definite straightforward boundaries. A coherent analysis thus requires a set distinction between metadiscoursal and non-metadiscoursal items. Below I discuss how metadiscourse is limited to non-propositional content and explicit language and compare the extent of metadiscourse to those of related theories.

Many of the functions of metadiscourse have also been studied under theories such as *stance*, *evaluation*, and *appraisal* (e.g. Tardy, 2012). Stance, developed in the *Longman Grammar of Spoken and Written English* (Biber et al., 1999) is concerned with how language expresses feelings, attitudes, judgements and assessments. Stance is divided into *epistemic* stance, which marks certainty of truth; *attitudinal* stance, marking personal judgement or emotional response; as well as *stylistic* stance, which comments on the way a particular text is said or written. Evaluation, too, is a cover term for things concerned with expressing opinion, maintaining reader-writer relationships and organizing the discourse (Thompson & Hunston, 2000:6). The appraisal framework, developed by Martin and White (2005) is a systemic functional framework that describes evaluation in terms of whether it relays emotions, moral judgement or aesthetic assessment of events or participants; this is considered under the system of *attitude*. The framework also analyses *engagement*, i.e. how statements are presented: whether the author doubts or agrees with the information and whether the proposition allows heteroglossia and alternative viewpoints. Finally, it analyses *graduation*, a complex category that relates to how texts refer to the number or

amount of participants or qualities, or to their prototypicality as member of a semantic group (e.g. *real* vs *sort of*) (Martin & White, 2005).

As with stance, the scope of metadiscourse is limited to the writer's/speaker's personal attitudes and evaluations. Epistemic and attitudinal stance fall roughly under metadiscoursal categories of Hedges, Boosters and Attitude markers. The functions of stylistic stance, however, have no clear-cut place in metadiscourse even though commentary on style in an adverbial phrase such as *seriously speaking* could be seen as "text about text". The function of such phrases can then be interpreted as marking topic shift or personal attitude depending on the co-text. As for functions under appraisal, resources for engagement fall mostly under the metadiscoursal category of Engagement, although metadiscourse does not offer as complex a taxonomy for these functions as the appraisal theory does. Some aspects of graduation (*focus* and *intensification*) fall under Hedges and Boosters. However, attitude in appraisal and attitude markers in metadiscoursal are widely different: this is because appraisal is concerned with all evaluation, whereas metadiscourse is concerned only with non-propositional content.

Non-propositional content is the kind of content that does not add information about participants or events but functions more as commentary; this is the chief topic of interest in metadiscourse analysis (Hyland & Tse, 2004; Hyland 2005). As Crismore (1984) puts it, metadiscourse directs rather than informs the reader. Using the terms of SFG, propositional content is concerned with the *ideational* metafunction of language, the level that construes the experiences or information that is being conveyed. Non-propositional content, then, serves the textual and interactive metafunction that shapes the information into an organised discourse (see Halliday & Matthiessen, 2014:30-31). Textual organisers, to name an example, depend on their co-text: for instance, the organizing words *first* or *then* are metadiscourse if they express non-propositional discourse-internal relations, that is, they are used for organizing the textual flow with a textual metafunction. When describing the temporal order of an actual event, they express propositional discourse-external relations and are not counted as metadiscourse (Hyland & Tse, 2004; Hyland, 2005:24–25). Likewise, deixis is not considered metadiscoursal if it is used to refer to entities described in the text; it functions as interactive metadiscourse only if used to refer back to the text or sections thereof, e.g. *the passage above*. This is an

example of multifunctionality of common metadiscourse markers: metadiscourse is a pragmatic phenomenon not tied to a set of words or constructions.

Another common criterion for identifying metadiscourse is explicitness, with only explicit markers being counted as metadiscourse. Note that the word *explicitness* here has a different meaning than is used in Mauranen's works, briefly discussed in Section 2.1 of this paper. Here explicitness refers to how well the metadiscoursal marker can be observed and identified (see Hyland, 2005:28). Metadiscourse focuses on lexico-grammatical devices of organization or interaction, but it ignores stance-indicating devices that work on an implicit level, such as word-choice, allusions, metaphor or intended inference. Contrastingly, in the theory of stance, evaluation (Thompson & Hunston, 2000:14) and in the appraisal framework (Martin & White, 2005:144) graduation of word-choice, e.g. *startled* vs *terrified*, is held significant. While metadiscourse explains a pragmatic phenomenon, its scope is limited to explicit devices. This is due to practical constraints, but also because the main interest of the analysis is in "the writer's or speaker's overt attempt to create a particular pragmatic or discoursal effect" (Hyland 2005:28). Hyland (2005:30) admits that many features of language convey metadiscoursal meanings and that they certainly do express authorial attitude, as does his example of allusion, *the chocolates he sent were actually a Trojan horse*. Expressions like this, however, are opaque and implicit, and the analysis can in some cases be difficult for an outsider of the particular discourse community (2005:30). The organisation of text can also be implicit: for example, lexical cohesion such as repetition of a word is implicit metadiscourse and thus Crismore et al., for example, opt to leave it outside their framework of metadiscourse. On the other hand, logical connectives such as conjunctions and conjunctive adverbs are explicit as they are not required to form a syntactically well-formed clause (Crismore et al. 1993:49).

The limitation of metadiscourse to explicit features is a practical choice; as Boggel (2009:24) notes, implicit metadiscourse generally overlaps with propositional content, which is not to be studied. From this follows that even if a metadiscoursal function is not expressed using metadiscourse markers, it does not mean the author has not expressed the function at all – it may merely be coded grammatically or in word-choice.

Returning to the issue of propositional and non-propositional content and the fuzzy boundary in between, some scholars have suggested syntactic criteria to

determine how multifunctional words can be analysed. Khabbazi-Oskouei (2013) criticises Hyland's (2005) notion that all forms of writer judgement are expressions of writer's attitude, whether they modify participants of a proposition or a full sentence. In search of an alternative, she suggests that an expression can be considered non-propositional and therefore metadiscoursal if it is separated from the main clause. For example, adverbs of frequency (e.g. *occasionally*, *usually*) can be considered metadiscourse only if "separated from the main sentence by appearing at the beginning of a sentence followed by a comma or within the sentence but between commas" (2013:96). For the sake of consistency, she uses this criterion also for classifying attitudinal markers. That is, a word separated by an impersonal clause or a comma is used to evaluate an entire proposition and thus metadiscourse (e.g. "it is *wonderful* that [proposition]", "*Fortunately*, [proposition]"). Yet the same word used within the main clause to modify an element will count as propositional and is therefore not counted as metadiscourse (e.g. "it was a *wonderful* day", "it all ended *fortunately*"). A similar division is also used by Dafouz-Milne (2003:35). However, especially Ifantidou (2005) criticises the notion that parentheticals do not contribute to propositional content. Genuine parentheticals "may or may not be perceived as making an essential contribution to the proposition expressed by the host clause" (Ifantidou, 2005:1338).

On the other hand, Khabbazi-Oskouei also suggests that what she terms "negation expressing counter-expectancy" be counted attitudinal metadiscourse. This kind of counter-expectancy occurs when "the editorialist implicitly announces that there are alternative positive positions which need to be rejected", as in her example reproduced below as example (3) (Khabbazi-Oskouei, 2013:103, her emphasis):

- (3) "More than 2,000 died in a pogrom in the state of Gujarat in 2002, for which *the perpetrators have never been brought to justice*. (The Economist, 27 Nov. 2008)"

The author is indeed trying to convey emotions or attitudes here by mentioning a fact they see as important; Thompson and Hunston (2000:13), too, consider the comparison of "what is not with what might be" as a form of evaluation. However, in the present study I hold negation of a proposition not as metadiscourse but as part of propositional content. Paraphrased, the example above states that "X has happened and Y (which relates to X) has not happened": any authorial attitude is implicit and it is up to the readers to deduce the evaluation by using their background knowledge. Furthermore, the statement is within the matrix clause and the logical connection is

not made explicit by a conjunction: the propositions are merely juxtaposed in a certain way to create a certain effect. Thus, I will not count these as instances of metadiscourse in this study.

Meanwhile, evaluative language, e.g. lexical items like *worry* or *love*, can mark metadiscourse when they are evaluations made by the writer or speaker (*I worry, I love*), but not when they have been attributed to some other entity in the text (*he worries, she loves*) (Gray & Biber, 2012). The attributed evaluation may well be shared by the author, but the author has nevertheless chosen to attribute it to an entity other than themselves. On the other hand, as pointed out by Ifantidou (2005), if an author comments on an evaluation by someone else, the author evaluates the truth-value of the utterance but removes the speaker's original evaluation. Consider this obvious case, where the verb (4) portrays author's own opinion (metadiscoursal) and in (5) the evaluation of someone else's opinion (non-metadiscoursal):

(4) I think *this is significant*.

(5) She claims *it is significant*.

In example (5) the author casts the original statement "it is significant" into doubt. From this follows that "inter-textual metadiscourse expressions make a contribution to the truth conditions on an utterance" (Ifantidou, 2005:1337). An evaluation made by someone else can thus not be considered truly metadiscourse of attitude although the evaluation by the author can mark evaluation or signpost intertextuality.

As stated above, punctuation does not function as a metadiscourse marker. Items such as exclamation marks, ellipses and symbols such as emoji are used for an effect, but they are as a rule excluded from metadiscourse framework. In certain ways question marks, dashes and parentheses are an exception. However, it is not the punctuation itself that is metadiscoursal, rather the content to which they are attached. Question marks, of course, mark questions used to engage the reader. Dashes and parentheses are used to add information, either to explicate (Code glosses) or to "give writer's opinion towards a particular issue" (Asides) (Dafouz-Milne (2003:39). The section surrounded by dashes or parentheses defines whether the parenthesis is explicating or expressing the author's opinion. The examples below illustrate the difference between Code gloss and Asides.

(6) "Syria continues to bleed, and the possibility of a confrontation between Russia – *Assad's key ally* – and the West becomes more real." (article342)

- (7) “Generations of Irish-American Catholics (*including my own paternal great grandparents*) [...] watched their hard work and spiritual touchstone sold off in pieces to different entities.” (article019)

Example (6) is from a text concerning the conflicts between Western leaders and the leader of the Syrian government, Bashar al-Assad. Here the text separated from the rest of the sentence is a Code gloss that provides information necessary to know in order to understand why relations between Russia and the West are strained. By contrast, example (7) is from a personal story: the phrase in the parentheses is not needed for understanding the content itself. Rather, it is an Aside with the purpose of adding human interest and explaining the author’s relationship with the matter.

In summary, there are several reader-guiding or reader-persuading rhetorical devices for writers to use, but with metadiscourse being limited to non-propositional and explicit items, many of them fall outside the scope of metadiscourse. Table 2 summarizes textual devices that can express authorial voice but are not treated as metadiscoursal in this study.

| Non-metadiscoursal rhetorical device | Example |
|--------------------------------------|--|
| Punctuation and typography | bold , <i>italics</i> , ALL CAPS, ☺, !, ...; |
| Organisers referring to real events | <i>First</i> , we..., <i>after that</i> , we...; <i>Here</i> , in this place |
| Attitudinal modifiers and graduation | <i>a wonderful</i> story; <i>a very</i> similar situation |
| Attitude expressed by third parties | “She thinks <i>it’s significant</i> ” |
| Word-choice | “I was <i>startled</i> ” vs. “I was <i>terrified</i> ” |
| Metaphors | “the chocolates were <i>a Trojan horse</i> ” |
| Counter-expectancy | “pogrom, for which <i>perpetrators weren’t brought to justice</i> ” |

Table 2: Non-metadiscoursal items

Text flow can certainly also be organized through paragraphing or typography; the reader can be engaged and persuaded by word choice or merely by presenting the right facts in the right order. However, the model of metadiscourse employed in this study is not designed to analyse these resources. Instead its focus is on the linguistic markers outside of the proposition, how the writer embellishes a propositional clause.

2.3 Summary of the concept of metadiscourse

In the context of this study, *metadiscourse* is understood as non-propositional, explicit linguistic rhetorical devices. To be considered non-propositional the devices should refer to, connect, or evaluate ideas – paragraphs, sentences and clauses – rather than individual events or participants. Explicitness is a criterion for limiting the definition to rhetoric that shows writer’s overt awareness of the text: it is assessed

based on how unambiguous the pragmatic meaning is without contextual information. However, no previous metadiscourse framework seems to limit the notion of metadiscourse to items that are explicit in the sense that they make overt reference to a textual element (see Mauranen 1993:158). This study, too, will include both metadiscourse markers that refer to the text as text and markers that make no overt reference.

The study adopts a broad approach to metadiscourse. Following Hyland and Tse (2004) *interactive metadiscourse* refers to markers building the logical connections between ideas or to items of text (e.g. meanings of words, sections of text). The markers share the function of facilitating the reading experience. *Interactional metadiscourse* refers to explicit markers that acknowledge the presence of the reader or the writer, whether through direct mention or by alluding to the reader's or the writer's opinions or modifying said opinions. Intertextual metadiscourse will be considered a part of interactive metadiscourse, inasmuch as reporting verbs and other linguistic markers can be used to refer to content from other sources that has been quoted or paraphrased in the text. Although the terminology is used, the distinction of metadiscourse marker categories into interactive and interactional is not necessarily of great importance for this study, as the goal here is to group the categories according to their co-occurrence rather than function.

The metadiscourse marker categories analysed in this study will largely follow the frameworks by Hyland (2005) and Dafouz-Milne (2003) summarized in Table 1. More examples of the categories will be provided in Section 4.2 and descriptions of their functions in Section 5.1.

3 Genre and Register Analysis

As metadiscourse is a rhetorical device, one needs to consider the intention of the writer when analysing metadiscourse features: why would the author wish to interact with the reader in a certain way? Author presence in a text usually follows the style typical for the particular discourse community. Metadiscourse in a text must therefore be analysed with reference to the conventions of the genre it belongs to (Tardy, 2012). As Hyland notes, the rhetorical environment of metadiscourse is “the context which conditions its use and gives it meaning” (1999:6). On the other hand,

the amount of metadiscourse also depends on the writer's background and the conventions they have been socialised into (see Mauranen, 1993:39). Studying academic writing, Mauranen notes that different language groups have different preferences when it comes to rhetorical strategies such as reader guidance. Although what she calls "Anglo-American culture" is highly diverse, there are clear tendencies that differ from those of other language groups (1993:253–258). The material of the present study, opinion writing from newspapers and blogs, is produced by writers based in English speaking nations: much else cannot be said of the author's backgrounds. It is thus the purpose, in this case the persuasiveness, that is the uniting factor of the genre.

Section 3.1 will define the key concepts of genre and register analysis in the context of this study. This is followed by a discussion of the genre at hand, Online Opinion Texts, in Section 3.2.

3.1 Concepts of genre and register analysis

The concepts of genre, register, and text type are quite nebulous, not least because of their various definitions by different scholars. Clear definitions of the related terms are therefore in order. This study follows the functional definition of Swales (1990:46), where *genre* refers to the texts created within particular communicative events that share a communicative purpose or a set of communicative purpose³. The communicative purpose, i.e. the goals the text is intended to achieve, could be for example reporting, describing, explaining, entertaining or persuading. This in turn reflects whether the text is factual or speculative and whether it expresses overt stance or remains objective (Biber & Conrad, 2009). The content and form of a genre are constrained by conventions that are recognised and upheld – but also developed – by a discourse community (Swales, 1990:53). Biber (1994) suggests seven parameters of situational characteristics for genres, all relating to external or contextual aspects of the situation where the text was produced: Participants, Relations among participants, Channel, Production circumstances, Setting, Communicative purpose, and Topic. While some genres do not have set norms in terms of content or style, the working assumption in genre analyses is that texts with a specific communicative purpose will become conventionalized in form and content.

³ Although the use of the word genre described here reflects the concept of genre in literary studies, the communicative purpose is generally "unsuited as a primary criterion" for literary genres (Swales, 1990:47).

Form and content, as well as structure and audience expectations, become characteristics that can be used to determine how *prototypical* of a particular genre a given text is (Swales, 1990:52).

To describe the linguistic form of a text, one can employ the term *register*. The present study follows Ferguson (1994), who gives the following working assumption for register:

A communication situation that recurs regularly in a society (in terms of participants, setting, communicative functions, and so forth) will tend over time to develop identifying markers of language structure and language use, different from the language of other communication situations. (Ferguson, 1994:20)

In other words, register refers to the language structure that is used in a particular situation or purpose, this structure being a repertoire defined through its distinct use of linguistic features such as syntax, intonation, formulaic sequences, or metadiscourse markers as is the case in this study.⁴ One can thus speak of different registers existing within different uses of language, one for example used in academic journals, another in conversation between friends. Because of the repeated setting and communicative function, opinion texts, too, would be expected to use a certain register. It should be noted that the term *register* has e.g. in Biber (1988) and Biber (1995a) been used to describe what here has been defined as *genre*, that is, Biber uses the word to describe texts with a certain purpose irrespective of their grammatical form.

Finally, there is *text type*, which refers to groups of texts that share a register. Different text types are distinguished from each other only by their linguistic structure, not their genre or features thereof. Thus, it is possible for texts from different genres, say a personal letter and public blogpost, to belong to the same text type. The term text type has been used in MD analyses, as a group of texts that are “maximally similar with respect to their linguistic characteristics”; different text types are “maximally distinct with respect to their linguistic characteristics” (Biber, 1994:52).

3.2 Online opinion texts as a genre

The genre which I have given the broad name of “online opinion texts” is rather a collection of genres or sub-genres, which share many features but differ in others.

⁴ The description of register above has in some research been termed *style*, but Biber and Conrad (2009) separate *register* from *style*, with *style* seen as an aesthetic property, not a functional one.

This section will delve into the characteristics of texts produced online that could be described as opinionative and journalistic. The types of texts relevant for this study are posts from opinion blogs, and columns and op-eds from the websites of magazines or newspapers. Further details on the corpus, its texts and its sources are given in Section 4.1.

Op-eds, columns and opinion blogs share a communicative purpose, which is why they could be argued to belong to the same broad genre. Their primary aim is to give a personal opinion on a matter and they are not bound by objectivity. They are opinionative in the sense that they are first and foremost polemic texts on issues that are already in-the-know. They can also be journalistic and concerned with presenting the factual content as an interpretive report, that is, as a report of a verifiable fact that is also being analysed, explained and evaluated (Rivers et al, 1988:8). In this way, they differ from news reports and other journalistic content, where the purpose is to cover the news in a detached and seemingly objective tone (although certainly news stories, too, mix evaluation and values with its factual content) (Greenberg, 2000). According to Greenberg, besides evaluating the issues on normative or prescribing bases, op-eds and columns can also encourage the reader to form their own opinions. Meanwhile, blogs can be categorised along two scales: as either personal or topical and as either individual or community oriented (Krishnamurthy, 20002; cited in Puschmann, 2013). Keeping in line with the persuasive purpose, opinion blogs here refers only to the kinds of weblogs that are concerned with more topical and community or society relevant topics, as opposed to more personal, diary-like blog publishing.

A principal characteristic of the texts in my corpus is that they all have single writers. While editorials would represent the opinion of an institution such as the newspaper, op-eds and columns in newspapers reflect the point-of-view of the author only. Blogs, too, are usually personal websites, where the author is responsible for all content. On the other hand, some blogs, such as a corporate blog of a think tank or research institute, may represent an institution in which case certain institutional guidelines may apply. Columns can be written by journalists working for the newspaper or by syndicated columnists. Op-eds is originally the shorthand for “opposite the editorial page”, which functions as space for columns and interpretive articles by journalists not in the editorial board. The abbreviation has later come to be interpreted as “the opposing editorials” and the content can function as a counter-

argument to the views presented in the editorials (Rivers et al, 1988:67). As Alonso Belmonte (2009:52) puts it, editorials represent “Our view” whereas op-eds represent “Other views”. Yet, an editor may oversee op-eds, unlike for personal blogs where the author acts as their own editor. The authors of editorials and op-eds are professional writers and are thus part of a professional discourse community that is aware of the conventional forms of the genre (see Swales, 1990:53). In terms of the communicative purpose, the authors try to state their personal opinion to persuade a reader who is not yet convinced of the matter. However, authors of op-eds are somewhat less concerned with creating a sense of urgency or “maximising the problem” than the authors of editorials (Alonso Belmonte, 2009:65).

Although sometimes online news may be located behind a paywall, the texts are public and the audience can in principle consist of anyone who has access to internet. The audience online is thus theoretically worldwide: Reese et al. (2007:237) consider specifically blogs as they write that the audiences “organize around issues and political affinities rather than geographical proximity,” but in a time where people can access the online news of any nation, this holds true of newspaper sites, too. However, the worldwide readership means that the audience is anonymous to the author, who in most cases addresses merely an imaginary ideal-reader. The addressee can sometimes participate in discourse, e.g. by sending a letter to the editor or by using the interactive commenting space provided on most online publishing forums; the discussion is however not equal, as the commenting space may be moderated and is in any case subordinate to the article itself. For blog writers the audience is closer than to an author of traditional journalism, partly due to the size of the audience. Wall (2005:161) terms the blogs’ audience as “co-creator” as readers may be “invited to contribute information, comments, and sometimes direct financial support” to the blog.

As news of the 21st century have shifted online, almost all newspapers have their own websites. Some newspapers also exist as purely online news, as news sites exclusively online where they originated. Both old media newspapers’ websites and online news sites often host feeds they term “blogs”. These news blogs share some of the characteristics of personal blogs. While news blogs can and do preserve a journalistic authority on the content, they make “less of a claim to know what readers want or to know what an event means” (Matheson, 2004:460). In a study on the blog upheld by BBC News, Hermida (2009) notes that the blog allows readers to

comment, which makes the communication less one-way. Blogging also allows journalists to convey insights and opinions about the story in a more personal way. However, while the tone and style is not restrained, at least BBC's bloggers must remain within editorial guidelines and not appear too impartial. Thus, blogs upheld by news institutions belong to the domain of journalism and are not, at least in the present study, considered blogs.

Returning to news sites, the format of online journalism differs considerably from that of traditional journalism: online news can be interactive and customized, whichever is the best way to reach the target audience (Deuze, 2003). Since the past two decades, many independent newspapers are exclusively online, the notable ones including for example *The Huffington Post*, *Business Insider* and *BuzzFeed*. Deuze (2017:11) associates the growing number with a global "emergence of a startup culture in the field of journalism." These news sites are not part of the so called old media, but are nevertheless sources upheld by a team of reporters – who are usually highly committed because of emotional engagement rather than economic reasons, seeing as funding is unpredictable and working conditions are prone to long days and underpayment (Deuze, 2017:14–15). While big news sites such as *The Huffington Post* raise money in the millions, small local digital news publishers rely on advertising revenues. Yet in 2015 only 47% reported turning a profit (Pew Research Center, 2016:59).

Journalism online has changed since its early days, its new features placing expectations on both journalism that is exclusively online and traditional media that has expanded online. Beckett and Deuze (2016:3) note that changes in media consumption habits put requirements on media. First, news sites have ever increasing competition online, especially in social media. Second, as news are made available at every moment thanks to technology, news sites must compete with the reader's attention by engaging the reader, for example by using so-called click-bait to create curiosity-gaps that reveal only some information and leave the reader wanting more. Third, people respond to emotion rather than facts, wherefore news need to stir feelings in order to pique the interests of their readers. Still, generally 1SG expressions of emotion and evaluation in newspaper reports are not authorial but attributed to other participants, e.g. people interviewed (Bednarek, 2008).

Turning to blogs, Matheson (2004:451) notes that blogs are seen as a democratic space that challenges mainstream journalism. The idea is that anyone can

write a blog on what they personally find interesting or important. However, blogs do not conduct any independent reporting; in general, they rely on mainstream media for topics and information (Haas, 2005). They can also bring back topics that mainstream media has not followed up or challenge ideas presented by the media. Undoubtedly, some blogs may do their own reporting, but most comment on issues that have already been reported elsewhere; the exception is if a blogger can report on news stories the sources of which mainstream media cannot access. (Campbell et al., 2009). However, with the arrival of social media, even this function is no longer exclusive to blogs.

As Wall (2005) notes, blogs that do conduct journalism have a more personal and opinionated styled narrative than traditional journalism. Myers (2010:120) observes that blogs have less need of hedging to soften the claim, as statements are based either on personal belief or “considered by default to be speculative and revisable.” Yet, Reese et al. (2007:277) note that many bloggers acknowledge different perspectives and are not necessarily easy to pin down in a single political standpoint. Pinjamaa and Cheshire (2016:6) observe that Finnish bloggers feel that blogs are increasingly becoming more professional and less “personal and introspective.” Stylistically, the influence of mainstream news media on blog coverage “is further strengthened by their appropriation of the format in terms of conventional journalistic norms and practices, and by their strict surveillance of the private weblogs of employees” (Haas, 2005:394). Nevertheless, while bloggers may be experienced or professionally trained writers aware of conventions in mainstream news, blogs remain personal platforms that can exhibit more freedom than columns and op-eds in terms of style and form. The story form of blog journalism can be more fragmented and provide less background details. Instead they direct the reader to other sources using hyperlinks, which also provide credibility to the blogger’s claims (Wall, 2005).

Metadiscourse in opinion articles has been studied to some extent, and the current study is joining to a still growing field. Le (2004) notes that, compared to academic writing, editorials are concerned with relaying less complex topics to a less specialized audience. On the other hand, academic writing is less limited in space. This holds somewhat true online, too, at least because of norms of the editorial genre if not of actual layout restrictions. The shorter format means that the reader has less need to be guided through the text with interactive metadiscourse (Le, 2004). The

same tendency is observed by Dafouz-Milne (2003): editorials in British newspapers contain less interactive than interactional metadiscourse. Interestingly, editorials in Spanish newspapers use more interactive than interactional metadiscourse, another sign of the impact of culture.

It is important for opinion articles in newspapers to persuade readers. Although Biber (1988) finds editorials to be impersonal and uninvolved, Alonso Belmonte (2009), analyzing the textual patterns of Question-Answer and Claim-Response, notes that a common tactic of both editorials and op-eds is the writer asking questions and countering them with an adequate answer of their own. Metadiscourse theory sees this anticipation of the reader's arguments as engagement of the reader and a marker of interactional metadiscourse. Fu and Hyland (2014) compare the use of interactional metadiscourse in newspaper opinion texts and popular science articles. They find all forms of interactional metadiscourse to be much more common in opinion articles. Engagement markers are the most frequent metadiscourse markers in opinion text, but especially Self-mentions are much more frequent in opinion texts than in popular science articles. Fu and Hyland observe that Attitude markers are common in both genres, but in opinion texts they are used to convey a wide range of writer's affective feelings (e.g. *unfortunately*, *dramatically*) rather than stance towards information (e.g. *importantly*, *surprisingly*), as is the case in popular science articles. Metadiscourse in blogging is to my knowledge yet an understudied field, a gap the current study aims to fill, but for example Myers (2010:78–86) discusses audience address in blogs. While operating beyond the metadiscourse framework, he discusses interactional metadiscoursal devices such as audience reference (e.g. *you*, *the readers*), questions and directives. Puschmann (2013), too, notes that the use of pronouns is crucial in audience design among bloggers.

4 Material and Methods

The method chosen for this study, Multi-dimensional (MD) analysis, requires an annotated corpus from which normalized frequencies can be counted for each text in the corpus. The data is described in Section 4.1 and the annotation of the specialized corpus in Section 4.2. After the annotation of the data, a MD analysis as described by Biber (1988) is conducted on the normalized frequencies of metadiscourse markers

in order to find correlation patterns between the markers. A walkthrough of the steps of MD analysis is given in Section 4.3.

4.1 Compiling the corpus

The material for this non-commercial research project is from sources that either explicitly state permission to do so on their webpage, or whose copyright holders or representatives have given the author written permission to use their material for the project. The corpus was collected in October and November 2016. To avoid authorial or search engine bias in selecting sources, it was decided that sources would be selected from an objectively collected list. Thus, the register of the texts should not impact the selection of the data. Alexa.com, a website traffic analytical intelligence tool, lists top 500 sites on the web by categories such as Society, News, Recreation etc.⁵ For this study the sites were chosen from *News>Analysis and Opinion*, *News>Weblogs*, *Society>Politics>News and Media*. In addition, some sources were chosen from Feedspot's regularly updated list of UK political blogs⁶ ranked by social metrics. Details on the sources can be found in Appendix 1. Although the resulting corpus does not make as fine grained a division, the sites to provide material can be divided into three source types, or sub-genres, which I identify as follows:

- *Blogs*: websites upheld by private individuals or non-journalistic institutions
- *Purely digital news*: websites of journalistic institutions that are fully digital, without hard copy print editions.
- *Partly-digital news*: websites of journalistic institutions that also issue hard copy print editions

A division into blog, column and op-eds was also considered, but the distinction of these categories would have had to rely on the self-identification made by the site. The various terms news sites use for their opinion writing section – including “column”, “opinion”, “op-ed”, “blog” – would not have provided as clear-cut researcher independent categories.

Having website traffic as a criterion means that the results of the study are indicative of the styles of experienced and possibly professionally trained writers but hardly representative of small private blogs. As many of the sites are from the United States, this study focused on the genre will make no attempts at analysing differences between regional varieties of internet language or genre specifics, although it bears to

⁵ <<http://www.alexa.com/topsites>>

⁶ <http://blog.feedspot.com/uk_political_blogs/>

be said that U.S. news are very much in focus in the topic content of the corpus. For a strictly synchronic study, the material consists of the most recent opinion pieces on the site at the time of collection and thus represents the issues that dominated the headlines in the autumn of 2016. Only articles that self-identified as opinion pieces or were clearly representative of such were collected: any articles clearly belonging to other genres (e.g. book reviews, obituaries, announcements on author's personal life or site maintenance) were excluded, even if they had been published under the same section as the opinion pieces. It is not, however, uncommon for articles to go off-topic where political or societal issues are discussed in connection to popular culture, faith or personal life. Such borderline cases were included, so as to not skew the data with restrictions based on pre-determined definitions or assumptions of the style or contents of the genre.

The corpus of 285,056 word tokens consists of 343 texts from a total of 27 sources. While images and audio-visual material as well as headlines, leads and bylines were not collected, the bodies of the texts were otherwise collected in their entirety. The texts vary greatly in length but are 831 word tokens in average. Some of the sources publish texts consistently shorter or longer than average: to make up for this, rather than collecting the same number of texts from every source, the goal was a similar total number of tokens (ca 10,000) from each source. Alternatively, the problem of different lengths could have been solved by cutting texts to match a certain length and leaving out the rest of the text: however, as this could skew the results, it is desirable to consider the texts in their entirety. For instance, some types of metadiscourse could be more prominent in a certain part of the text, such as in the end that would be cut off if a limit was set to the length of the text (see Zhang, 2016:222). Furthermore, some metadiscoursal items may occur only in a particular section of a text and leaving out the section would again misguide the analysis. While a multi-modal analysis of the texts would provide interesting results, the corpus is intended for purely linguistic research, wherefore no illustrations, photos, audio or video were collected. For a similar reason, embedded hyperlinks were not retained in the corpus: while they may serve the same purpose as Attributors (metadiscourse markers attributing information to a source) they are not explicit or linguistic markers and fall thus outside of the scope of the present study.

As the interest lies in metadiscourse markers as used by the blog/column writers, metadiscourse markers in quoted passages were excluded, that is, the

analysis does not count metadiscourse markers that appear in paragraph-long quotes from other text sources the author has copied from elsewhere (e.g. other news articles, official documents or social media). For instance, one article goes on to give lengthy citations from a speech by presidential candidate Donald Trump to describe his supporters (ellipses in brackets original):

- (8) Just a reminder of the kinds of things they are cheering:
 "[...] we are in fact controlled by a handful of global special interest rigging the system and our system is rigged. [...] They will seek to destroy everything about you including your reputation. They will lie, lie, lie and then again they will do worse than that. They will do whatever is necessary. The Clintons are criminals. Remember that. [...]" (article166)

The quote goes on for another paragraph. Citations in the article allow for sharing information without referring the reader to other pages. However, the quoted texts are from contexts entirely different from the article itself. Without deleted quotes like this, the word token count of the corpus is 285,056 tokens; the deleted quotes would in total account for 16,639 word tokens. Quotes integrated within the text as shorter quotes were retained: these integrated quotes appear in the same paragraph as author's own text, usually interrupted by reporting verbs. They were used to source statements or support the author's reporting. Meanwhile, paragraph-long quotes often served to republish longer sections of other texts for the reader's convenience.

4.2 Annotation of metadiscourse

Whether a word actually has a metadiscoursal function very much depends on its context, which problematizes annotation in a corpus of this size without precise annotation tools. Ädel and Mauranen (2010:2) distinguish between what they term "thick" and "thin" approaches to metadiscourse, arguing that the choice of approach has "implications not only for the method of identifying metadiscourse, but arguably also for how the category is understood." The thick approach is a qualitative data-oriented approach where potential metadiscourse markers are retrieved from the data itself. All relevant markers are then analysed in the context of their lexicogrammatical frames or discourse functions. For example, Ädel (2006) examines some aspects of metadiscourse using a thick approach. In contrast, the thin approach is quantitative and uses a pre-defined list of inherently metadiscursive markers. This approach is useful for corpus studies where large quantities of data needs to be analysed. A thin approach is used in Hyland's quantitative analysis (2005). The approach in the present study is quite thin, as the selected metadiscourse

markers are largely based on previous research rather than on the dataset of the present study. On the other hand, as metadiscourse is a rhetorical function, not a grammatical category, there is no closed class of potential metadiscoursal markers, although previous studies on metadiscourse usually provide examples of markers. Furthermore, an automated annotation based on the assumption of inherent metadiscourse can be superficial if no closer analysis is conducted, and a pre-defined list may not be suitable to analyse a genre it is not intended for.

When deciding what lemmas or constructions could mark metadiscourse, the list of potential markers was created by studying earlier research as well as reviewing and becoming familiar with the texts in the data. Because of the limitations of this thin approach method, the lists were complemented by adapting thick approach methods, that is, by extracting wordlists for single-word markers from the corpus itself. Verbs, adjectives and adverbs that appeared more than once in the corpus were categorised according to whether they could mark metadiscourse. Closer analysis was then conducted only on the lemmas that could potentially express metadiscourse (see Appendix 3 for a list of potential metadiscourse markers). To make a statistical study feasible, potential markers that occur only once in the corpus have not been included in the wordlists⁷.

The extraction of the wordlists and the annotation itself was carried out using *UAM CorpusTool*-software, a corpus annotation tool suitable especially for tagging segments of texts and tagging a section on several layers, e.g. on both a semantic-pragmatic and a syntactic level (O'Donnell, 2008). The software also adds part-of-speech tags to the data, which allows for automated queries. Decisions on the annotation criteria were made beforehand for the analysis to be systematic. This is a necessary step to tackle the multifunctionality of words that endangers the precision of automatic annotation. For example, conjunctives have a cohesive function if they are used to express relations between clauses, but not if they express only a relation within a sentence (Halliday & Hasan, 1976:6–7, 233). Likewise, in metadiscourse analysis, in order to be considered having metadiscoursal function they must connect

⁷ In many cases, even the potential markers were found to be non-metadiscoursal. For example, verbs such as *brag*, *decline*, *offer*, *raise* and *recognize* could in theory introduce a speech act: they were thus all checked for in the corpus to determine whether they could function as Attributors, a metadiscoursal marker that attributes a proposition to a written or spoken source. However, these verbs only appeared as propositional content in contexts where they described someone's action or reaction, but never where it could be considered as sourcing a statement (e.g. "brag about <np>" rather than "brag that <clause>"). It is likely that a clear majority of the rare items that went unnoticed are lexemes that have only non-metadiscoursal functions, at least within this data.

and explain relations between propositions rather than single items (see Crismore et.al, 1993:49). Compare the examples below: in example (9) the word *but* is used to contrast the propositions of two matrix clauses – its purpose in the text is cohesive in logical sense and as such it functions as a Transitional marker. In example (10) *but* merely connects two adverbials within a single proposition and does thus not function within the scope of metadiscourse.

(9) “I tried to tabulate the gains of the open house *but* soon gave up” (article174)

(10) “McTernan did come in for some heckling as the meeting wore on, which he dealt with politely *but* robustly” (article143)

As the example shows, fully automatic tagging would not be reliable because of the multifunctionality of the potential items. Often either the search query was precised or initial criteria for features were revised if found imprecise or insufficient. In most cases, manual checking of the query results was also employed by necessity as syntactic restrictions were not enough to extract metadiscoursal uses only. A similar combination of automatic and manual analysis has also been used by Dafouz-Milne (2003) and Ädel (2006) among others. It must be acknowledged that it is unlikely for a thin approach to have perfect recall of metadiscoursal markers simply because authors have quite different writing styles. Even with wordlists extracted from the corpus there could very well remain metadiscoursal *hapax legomena* that have not ended up in the wordlist. On the other hand, a conscientious manual check of a wordlist that contains the entire used vocabulary would defeat the purpose of computerised annotation of a larger amount of data. Thus, even if the chosen method does not have a perfect recall, the results reach at least an even and systematic recall: items that have been annotated in one text have been annotated the same way in all the other texts where they occur.

In the present study, the metadiscourse markers are grouped into 13 categories, which are given in Table 3. Initially, 30 separate marker types (also listed in Table 3) were analysed as separate categories, but as some were low in frequency, many categories were merged to create larger macro-categories with higher overall frequencies, as categories with higher frequency will yield better results when using the quantitative MD analysis. For example, organisers such as Sequencers or Topic shifts are far less common in shorter opinion texts than in Academic writing, on which e.g. Hyland and Tse (2004) base their framework of metadiscourse. Thus, in this framework the markers were merged with Announcers and Discourse goal

markers into a macro-category termed Textual organisers. As the focus of this study is on metadiscourse as a rhetorical device, metadiscoursal markers serving a similar function were also merged even though they differ in grammatical form (e.g. Hedging adverbs and Hedging adjectives). Conversational markers are not listed as a category in the metadiscourse frameworks by Hyland and Tse (2004) and Dafouz-Milne (2008). Their inclusion in this framework is inspired by Myers (2010:84–86), who discusses the use of interjections and paralinguistic features (e.g. *sigh*, *wink*) in computer-mediated communication as an expression of humour and conversational response to projected readers.

| | |
|--------------------------------|---|
| 1. Transitional Markers | |
| a. Additive transitions | <i>and, also, not only, in addition, moreover</i> |
| b. Similarity transitions | <i>likewise, similarly</i> |
| c. Contrasting transitions | <i>but, yet, however, although</i> |
| d. Concluding transitions | <i>so, therefore, because, thus, accordingly</i> |
| e. Countering transitions | <i>nevertheless, even so, still</i> |
| 2. Textual Organisers | |
| a. Sequencers | <i>secondly, finally; (1), (2); on the other hand</i> |
| b. Announcers | <i>this graph, click here, above, the following</i> |
| c. Topic shift | <i>as far as, meanwhile, back to; Let's switch...</i> |
| d. Discourse goal markers | <i>I argue; in short; Let me make it clear...</i> |
| 3. Code Gloss | explaining items added in parentheses; <i>such as, for example, that is</i> |
| 4. Attributors | General reporting verbs: <i>say, suggest that...</i> |
| 5. Hedges | |
| a. Epistemic adjectives | <i>likely, possible, uncertain, unclear...</i> |
| b. Probability adverbials | <i>probably, maybe, allegedly, I think...</i> |
| c. Hedging verbs | <i>seem to; appear to, tend to...</i> |
| d. Approximators | approximate quantity: <i>nearly, around...</i> |
| e. Epistemic modals | <i>may, might, could</i> |
| 6. Boosters | |
| a. Boosting adjectives | <i>it's clear/certain/evident/indisputable</i> |
| b. Boosting adverbials | <i>obviously, certainly, of course, no doubt</i> |
| c. Boosting reporting verbs | <i>we know; it confirms/proves/guarantees</i> |
| d. Necessity modals | <i>must</i> (when epistemic) |
| e. Boosting noun expressions | <i>the fact is, proof that</i> |
| 7. Attitude Markers | |
| a. Attitudinal adjectives | as predicative: <i>it's fun/interesting...</i> |
| b. Attitudinal adverbs | Author's attitude: <i>unfortunately, hopefully...</i> |
| c. Attitudinal verbs | Author's attitude: <i>I hate/want, I'm afraid...</i> |
| d. Cognitive verbs | Author's cognition: <i>I wonder/believe...</i> |

| | |
|----------------------------------|---|
| 8. Reader Address | |
| a. 2SG-pronouns | <i>you, your</i> , etc.; also indirect e.g. <i>our readers</i> |
| b. directives | <i>Consider...; let's hope...; don't...</i> |
| 9. Self-Mention | |
| a. 1SG-pronouns | <i>I, me, my, myself</i> , etc. except where attitude marker |
| b. exclusive 1PL-pronouns | <i>we, us, our</i> , etc. which do not refer to reader |
| 10. Inclusive We pronouns | <i>we, us, our</i> , etc. which also refer to reader |
| 11. Deontic Modals | <i>must, ought to, need to, should not</i> , etc. |
| 12. Questions | Rhetorical questions: <i>Remember how...?</i> |
| 13. Commentary | |
| a. Asides | Commentary separated by dashes or parentheses, e.g. <i>(including my own paternal great grandparents)</i> |
| b. Conversational devices | <i>oh, hey, wow, damn, yes, sigh...</i> |

Table 3: Metadiscourse marker categories

Inclusive We pronouns are not counted separate from Self-Mentions in Hyland's framework. The separation of the categories in this study is based on Ädel (2006:31), who discusses the uses of the 1PL-pronoun. Based on its reference point, it is quite natural that inclusive *we* is a separate entity situated in a continuum between Self-mention and Reader-address. Like with other multifunctional metadiscourse markers, the inclusive/exclusive function of the pronoun was something that required manual annotation on automatically extracted concordances.

The difficulties between separating propositional and non-propositional content were already discussed in Section 2.2 of this study. As mentioned, Khabbazi-Oskouei (2013) suggests that Attitude markers must be separated from the proposition by commas whether they appear at the beginning of a proposition or in the middle of it. This approach is adopted here as it is convenient in a computerised corpus study. When annotating attitudinal adjectives, it was noted that a modifier of an entire clause is more likely metadiscoursal than a modifier of a noun phrase, as the attitude is made more explicit and non-propositional. Compare for instance the word *surprising*, which in example (11) modifies a noun, and in (12) is the subject complement of a dummy subject referring to a whole that-clause:

(11) "Wind energy has made some *surprising* advances." (article360)

(12) "it is scarcely *surprising* that this enthusiasm is not wholly shared by some of his regional colleagues" (article014)

Example (11) is considered propositional, whereas example (12) is non-propositional and therefore metadiscoursal. This set of criteria is needed also when classifying less explicit metadiscourse: for example, a modifier of a single phrase may indeed be the attitudinal evaluation, but if it occurs as part of propositional content, it cannot be

considered metadiscoursal (Khabbazi-Oskoueï, 2013:99). The fewer compromises the researcher must make, the more reliable the results are.

As touched upon above, to keep the workload manageable, the search queries on the *UAM CorpusTool* were narrowed down to syntactic frames where the usage as metadiscoursal markers was likely. For example, for reporting verbs to function as Attributors, a device to link source to statement, they should be proximal to nouns or 3rd person pronouns as in constructions like *she stated* or *the report suggests that*. In constructions such as *somebody has said* or *they have yet to report* the verb does not serve its function as Attributor and cannot thus be considered metadiscoursal. To retain reliability, manual checking was conducted on the results. The context-based distinction between certain Code glosses and Asides has already been discussed in Section 2.2 of this paper. Manual annotation was necessary also for multifunctional words that could mark different types of metadiscourse, such as the 1PL pronoun *we*, which counts as Inclusive We when including the reader in the referent, but marks Self-mention when excluding the reader.

While most metadiscourse markers were analysed within the whole text, markers that by definition refer to the author, i.e. Self-mentions and Attitude markers were analysed only within text written by the author of the article, not in sections quoted from elsewhere⁸. This is to distinguish between author's own opinions and those of others. It can be argued that all metadiscourse marker categories should be limited to text written by the author themselves, but on the other hand the author has made the conscious choice of including the metadiscourse marker of someone else rather than paraphrasing the borrowed idea. Thus, metadiscourse marker apart from Self-mention and Attitude markers were automatically annotated also in shorter quotes or paraphrased content from other sources. In the big picture, the frequency of metadiscourse markers is not very high, and it was noted that the markers found in quotes have little effect on the results of the factor analysis. As noted above, lengthy quotations were excluded from the analysis.

4.3 Multi-Dimensional Analysis

Multi-dimensional (MD) analysis is a corpus-driven quantitative method for exploring the co-occurrence patterns of a large group of linguistic features across

⁸ It should be noted that to avoid double annotation, 1SG/1PL constructions marked as Attitude markers are not counted as Self-mentions in spite of the presence of the first-person pronoun.

registers. As the second research question of the present study concerns co-occurrence of metadiscourse marker categories, MD analysis is ideal for the purposes of this study.

In a corpus-driven approach the aim is to formulate new theories based on evidence from the corpus itself. In the words of Tognini-Bonelli (2001:87), corpus-driven studies aim “to derive linguistic categories systematically from the recurrent patterns and the frequency distributions that emerge from language in context.” As MD analysis is a method used to extract correlation patterns directly from the data, it is classified as corpus-driven, as opposed to corpus-based, where corpora are used to empirically examine pre-existing conceptions or theories that are not based on corpus-data (Tognini-Bonelli, 2001:68).

This section will first consider the theory and the presuppositions behind the method developed by Douglas Biber (1988). Section 4.3.2 describes the more technical aspects of conducting a MD analysis. It presents and explains the more statistical findings of the dataset used in this study, which will then be interpreted and analysed more in-depth in Section 5.

4.3.1 Multi-Dimensional Analysis as an empirical method

Biber et al. (1998) note that while linguists have long understood the importance of patterns in the occurrence of linguistic features, these patterns were difficult to study without adequate methods. Detecting patterns becomes challenging especially when studying large quantities of data or when the number of possibly significant linguistic features is high. Yet a large sample is needed to make generalizable observations and only a higher number of variables can provide a fuller picture. MD analysis has been developed to solve this problem: by merging correlating variables, it decreases the number of variables and can be applied to big corpora. Developed and described in full in Biber’s highly influential work *Variation across Speech and Writing* (1988), MD analysis is a form of Exploratory Common Factor Analysis, a statistical method where co-occurring variables are grouped together to form a single unit called a *factor*. Essentially, a large number of variables is reduced to a smaller number of variables in a process facilitated by statistical tools. The advantage of the method is that it is easier to find and interpret patterns between sets of variables than to study all the variables separately and compare them to each other all at once. Indeed, the goal of a MD analysis is not to operate on the level of an individual variables or to

describe a single feature. Neither is the aim to separate categories that have no overlap. Rather the purpose of MD analyses is to find “multiple parameters of variation” that function as continua along which different kinds of texts vary (Biber, 1995b:343). A comparison of parameters is easier but also more reliable in distinguishing genres than the analysis of individual features, where idiosyncratic texts of the variables can distort the numbers (Biber, 2009:824). Biber (1988) used MD analysis on the LOB (Lancaster–Oslo–Bergen) corpus to analyse variation in spoken and written genres of English. The corpus was tagged with as many as 67 grammatical and functional linguistic features, including constructions such as nominalisations and *wh*-relative clauses, which were merged into seven parameters. Since then, MD analysis has become a popular method for analysing co-occurrences in different registers and languages. Because of its popularity (see e.g. Biber, 1995b; Sardinha & Veirano Pinto, 2014), MD analysis has been tested and established a position as a trusted and valid method for studying variation of linguistic structure. The present study is focused not on grammatical features but on a smaller number of pragmatic variables not bound by form and the aim is to find variation within rather than across a corpus. The study will nevertheless follow the method as described in Biber (1988) quite closely.

In a MD analysis, correlating variables are merged into factors. Variables have *factor loadings* which determine how strongly the variable belong to a specific factor. When more than one variable clusters in the same factor, the working assumption is that the correlation can be explained by a shared underlying latent variable (Pett et al., 2003:3–4). The function of this underlying variable can be understood by interpreting the functions of the variables with the highest factor loadings in that cluster. Thus, if a set of linguistic features tend to co-occur in texts, it indicates that the features “work together to mark some underlying [linguistic] function”, sometimes more than one (Biber, 1988:55). Factors that are assigned a meaningful function are termed *dimensions*. In genre and register analyses the dimensions are thought to represent a function of the text, such as its setting or communicative purpose. However, when several dimensions are extracted, a text can be compared to others across the dimensions. This yields a broad characterisation of the functions at play. Some functions are typical in particular genres or text-types, and texts may be similar in some dimensions but different in others.

While MD analyses are useful for finding variation between genres, the goal of the present study is to seek variation within a genre, that is, among texts that have similar communicative purposes. When studying the sub-genres of a single, it is beneficial not to use external criteria such as situational context as starting point, as the researcher's preconceptions on genre, author or audience could affect the results. Instead the starting point of the present study is in the linguistic variables and the underlying functions emerge as dimensions from a statistical analysis. The use of tested statistical methods also means that pre-determined definitions and intuitions play a lesser part in the description of the correlation. On the other hand, to understand the function of a given feature and its use in a specific text, the quantitative analysis needs to be complemented by a qualitative analysis of the shared functions of the relevant features (Biber, 1988:62–63). The qualitative analysis can verify that the interpretations of the underlying functions of the dimensions are correct. This is also important in a study on metadiscourse, where a “thin” approach leaves an analyst only with a general overview of the use of metadiscourse in the data, with no insight in the collocations or co-text of the markers.

In Biber's (1988) study, the dimensions represent a linguistic register. Although most studies using MD analysis do focus on grammatical features, some previous studies have successfully used it to study stance and metadiscourse phenomena. Precht (2000:2) argues the method is well suited for the study of stance “because of the complexity of this construct”, and I believe the same can be said on the study of metadiscourse. Zhang (2016) has used MD analysis to study metadiscourse across several registers. Although Zhang, following Ädel's (2006) narrow approach, makes use of a modified reflexive model of metadiscourse and extracts dimensions from a corpus of 1 million words, the results consider also variation within sub-genres. There is thus reason to believe that a MD analysis can provide insights into different uses of metadiscourse also in a specific corpus. As metadiscourse is a more pragmatic phenomenon not tied to grammatical structure, the functions of the dimensions are better thought of as rhetorical strategies. Following Mauranen (1993:34), the term strategy here does not “imply a real psycholinguistic process, but an abstraction of observations of text.” As discussed in Section 2.2, a study in metadiscourse is a study of writer's overt attempts at discursial effects. Texts that share a rhetorical strategy, that is, score high in a certain

dimension or use the same metadiscourse markers would according to the outlined theory of genre belong to the same text type.

4.3.2 Conducting a Multi-Dimensional Analysis

When conducting a MD analysis, the size of the corpus is of secondary importance – what matters is the range of variation (Biber, 1995b:364). Admittedly, as the corpus used in the present study consists of texts with more or less the same mode, purpose, and topic, it is to be expected that there is less variation to be found than in a study considering several very different genres. If statistically significant variation is found in this corpus, the data contains variation and texts of different text types. This would support the hypothesis that the use of metadiscourse cannot be predicted from a too broadly defined communicative purpose or genre, as well as the assumption that more qualitative analysis is worthwhile from a discourse analytical point-of-view. The steps of the MD analysis in this study are as follows (cf. Biber, 1988:64):

1. The metadiscourse markers to be analysed are annotated in the corpus
2. Frequencies of the markers in the texts are counted using a computer program
3. A Factor Analysis is used to cluster the markers into groups, *factors*, according to their co-occurrence
4. The functions of the markers in each factor are assessed, from which the underlying function, the function of the *dimension*, is interpreted
5. A *factor score* is computed on each dimension for each text
6. Dimensions and individual texts are interpreted further using qualitative methods

Once metadiscourse markers had been annotated into the corpus using *UAM CorpusTool3*, the same software was used to count the raw number of instances for each type of marker category. Some marker types were found to be very rare throughout the corpus; these were combined with other closely related categories (see Section 4.2). To adjust for differences between text lengths, the raw numbers were then normalized for frequency per 1000 words, by using the formula:

$$f = \frac{\text{raw occurrence}}{\text{word token count}} * 1000$$

As the corpus consists of almost 350 texts, the sample size is small but likely to be sufficient, especially given that the number of variables (i.e. metadiscourse marker

categories) is only 13 (see Field, 2005:640). Furthermore, the Kaiser-Meier-Olkin index (KMO)⁹ of the data is 0.608, which indicates it is fit for factor analysis.

The factor analysis and the statistical analyses of this study were conducted using RStudio, a software environment for statistical computing using the R programming language¹⁰. A Principal Component Analysis (PCA), which analyses the correlation of variables, was conducted using the *factoextra* package¹¹ to find the number of factors that can be extracted. PCA determines how variables correlate and what their variance is, that is, how the scores deviate from the mean (Pett et al., 2003:56). There are several methods to determine a suitable number of variables to extract, but here a Scree test was used.

In a Scree test the eigenvalues¹² are presented as a line plot: factors are extracted based on where the line levels (Pett et al., 2003:119). The plot in Figure 1 shows the first break between factors 2 and 3, the second between factors 4 and 5, but the actual levelling happens only after the sixth factor. However, a six-factor solution has a p-value of 0.68. On the other hand, as Biber (1988:88) notes, extracting too few factors would lead to too many variables being loaded under the same dimension, which would lead into a problematic interpretation of the underlying variable. In the present Scree test a four-factor solution divides the variables more evenly, explains their variance more, and is statistically significant. It was thus decided that a four-factor solution would be optimal for the present dataset.

⁹ The KMO index determines whether the sample size is fit for a factor analysis: if there is too much or too little correlation in the data, factors cannot be extracted. The index can range from 0 to 1 but should be above 0.50 for the data to be usable. However, an index above 0.60 is desirable for the results to be worthwhile (Pett et al., 2003:78; Field, 2005:640).

¹⁰ For R, see <<https://www.r-project.org/>>. For RStudio, see <<https://www.rstudio.com/>>

¹¹ Developed by Alboukadel Kassambara and Fabian Mundt, the *factoextra* package is available at <<http://www.sthda.com/english/rpkgs/factoextra/>>

¹² A factor's *eigenvalue* essentially indicates how much of the variance of the variables is explained by that factor, with higher eigenvalues explaining more variance (Pett et.al, 2003:99). The more variance a factor explains, the more important it is to retain (Field, 2005:632).

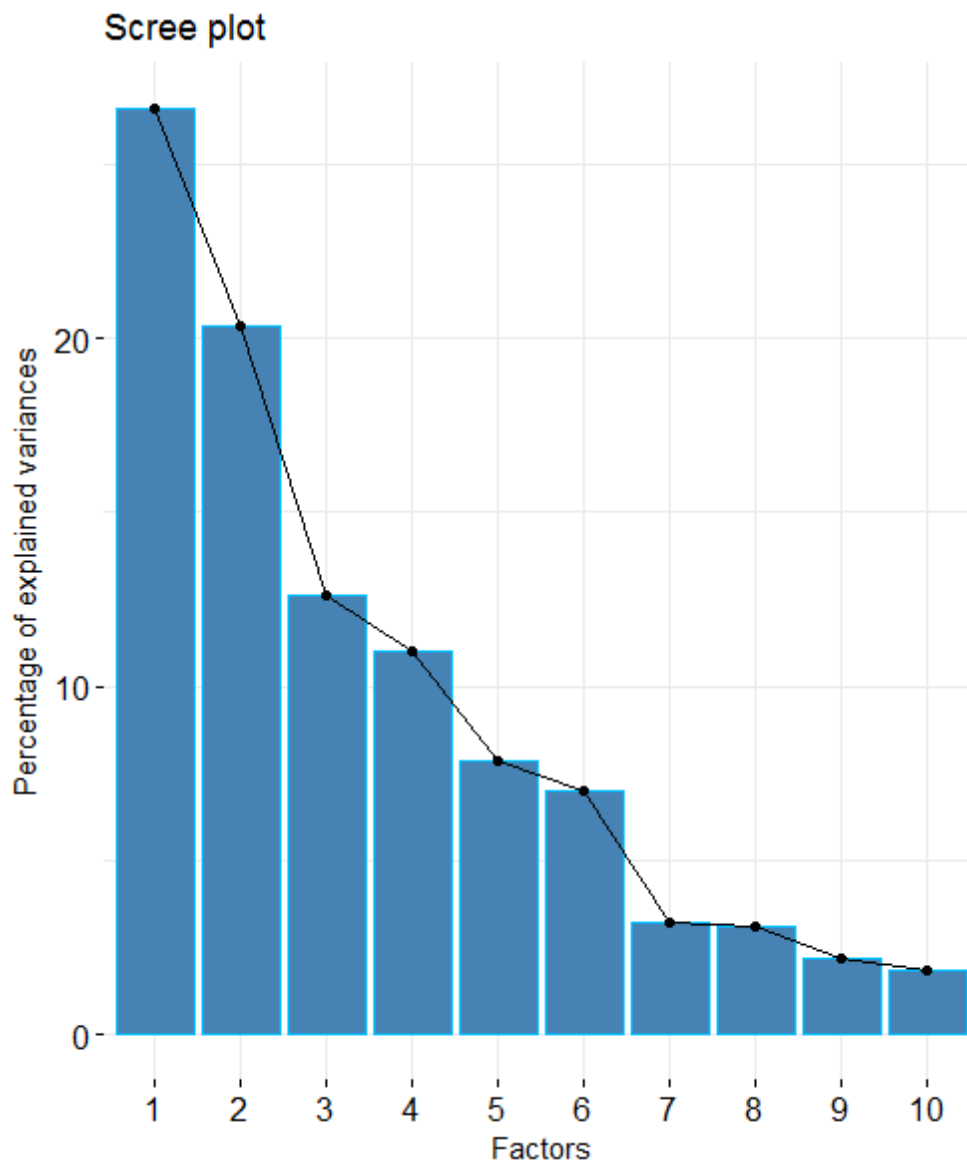


Figure 1: Scree plot for Principal Component Analysis

As shown in Table 4, cumulatively the first four factors explain 70.5 % of the variation. The first two factors 26.6 % and 20.4 % of the variation respectively. The factors were extracted using `factanal()`-function with Promax-rotation. Promax is an oblique rotation suitable for when factors can be assumed to be dependent of each other, as is the case with metadiscourse markers (see Zhang, 2016:211; also Biber, 1988:85).

| Factor | Eigenvalue | Variance (%) | Cumulative variance (%) |
|--------|------------|--------------|-------------------------|
| 1 | 71.314328 | 26.621969 | 26.62197 |
| 2 | 54.613120 | 20.387331 | 47.00930 |
| 3 | 33.814696 | 12.623183 | 59.63248 |
| 4 | 29.552474 | 11.032076 | 70.66456 |
| 5 | 21.108285 | 7.879821 | 78.54438 |
| 6 | 18.790673 | 7.014645 | 85.55902 |
| 7 | 8.629369 | 3.221383 | 88.78041 |
| 8 | 8.257219 | 3.082458 | 91.86287 |
| 9 | 5.856703 | 2.186334 | 94.04920 |
| 10 | 5.000809 | 1.866825 | 95.91603 |
| 11 | 4.446530 | 1.659910 | 97.57594 |
| 12 | 3.538357 | 1.320885 | 98.89682 |
| 13 | 2.955169 | 1.103178 | 100.00000 |

Table 4: Eigenvalues and variance explained in unrotated factor analysis

Factor loadings denote how big a part a variable plays in the factor, that is, how strongly it belongs to that factor. A negative factor loading indicates an inverted correlation, that is, the two variables do not tend to co-occur. A marker is considered to belong only to the factor where it loads the highest as this is the factor it defines the strongest. Generally, only variables that load higher than $|0.30|$ are retained in the factor. Biber (1988:93) uses a conservative cut-off point of $|0.35|$. In a MD analysis on a fairly homogenous genre it is not surprising that many of the variables have lower variance and load on more than one factor. Sometimes variables with low loadings on all factors are eliminated from the data altogether. However, Pett et al. (2003:173) suggest that variables with low loadings can still be retained in the factor analysis if they are of importance to a subset of a study. Thus, all categories of the present study were retained.

After the factors have been extracted, *factor scores* can be calculated for each text (Biber, 1988:94–95). First the standardised scores, Z-score, of each metadiscourse marker category is calculated using the mean frequency and standard deviation (see Appendix 2 for the mean frequencies and standard deviations of the metadiscourse marker categories in this dataset). Working under the assumption that the frequencies representing the population follow a normal distribution (i.e. a Gaussian bell curve), the standardised score Z has the following formula:

$$Z = \frac{f - \mu}{\sigma}$$

f = normalised frequency, μ = mean, σ = standard deviation

The score essentially shows the difference between the normalized frequency and the mean frequency in units of standard deviation, with negative scores meaning the

frequency is below the mean. Therefore, a low Z-score does not indicate a low frequency of a marker: rather it merely shows that the frequency is close to the mean frequency ($f - \mu = 0$). The maximum Z-score is theoretically ± 4.0 , but the normal distribution means that 68.2% of observations will have a standardised score Z that is $-1.0 \leq Z \leq 1.0$.

The factor score F of a text is the sum of standardised scores for the metadiscourse categories that belong to a Factor X. For example, the metadiscourse marker categories loading on Factor 1 are Attitude markers, Self-mentions, Transitional markers and Hedges. If one text scores -0.32 in Attitude markers, -0.60 in Self-mention, 1.52 in Transitional markers and -0.22 in Hedges, its factor score for Factor 1 would be -0.39 as according to the following:

$$\begin{aligned} F_1 &= Z_{Attitude} + Z_{Self-mention} + Z_{Transitional\ marker} + Z_{Hedges} \\ &= -0.31 - 0.60 + 1.52 - 0.22 \\ &= \underline{0.39} \end{aligned}$$

If normalized frequencies were used, a high-frequency category would have “inordinate influence on the factor score” (Biber, 1988:94). The standardising of the score is done so that each metadiscourse category is only compared to itself. As Factors 2 and 4 only consist of one metadiscourse category, their factor scores are equal to the standardised scores of Inclusive We and Attributors respectively. It should be reminded that scores from two separate dimensions are not comparable in unit, due to the different structures of the dimensions.

Once the underlying function of an extracted factor has been interpreted, the factor is termed *dimension*. After each text has been given a factor score, now *dimension score*, the score can be used to rank texts in how well they represent a particular dimension, that is, the underlying variable that explains the correlation. Once again, a dimension score close to zero, such as 0.39 above, does not indicate a low frequency of the metadiscoursal markers in that dimension. It merely signifies that the use of categories adds up to a score that is close to the mean of the data.

A text with the dimension score that equals the mean can be considered to have a use of the dimension that is very typical to the genre. A text where all dimension scores are close to the mean is typical across all dimensions. To get a clearer picture of what counts as a typical opinion text according to the current corpus, I use a separate criterion which I term *prototypicality*. This cross-dimensional property is here defined as the distance to the mean score. Note that this concept is

not present in Biber's (1988) model of MD analysis. The concept is introduced here in order to facilitate comparison of individual texts with respect to the mean score. In a four-dimension solution a perfectly prototypical opinion text would have a dimension score of 0 in all four dimensions. Thought of as Cartesian coordinates, the mean is per definition represented by the origin O. The position P of a text is thus composited of its dimension scores. Adapting a standard formula for the distance between origin O and position P, the distance between the mean and text, the *prototypicality score*, can thus be measured as the squared sum of dimension scores to the power of two, which in a four-dimension solution is solved as below¹³:

$$\overline{OP} = \sqrt{(a^2 + b^2 + c^2 + d^2)}$$

The lower the prototypicality score, the higher the prototypicality and the text's resemblance to the mean. Prototypicality according to this definition only accounts for the dimensions and ignores both the use of variables outside the dimensions (Boosters, Code glosses, Commentary and Deontic modals) and the inner structure of a dimension, that is, how the variables of a dimension account for the dimension score. It is thus intended to explain prototypicality only across the four extracted dimensions, not across the metadiscourse framework overall.

The functions of the four extracted factors will be interpreted and discussed in the following section.

5 Findings

Section 4.3.2 presented the results of the first part of a MD analysis: the quantitative Factor Analysis yielded four statistically significant factors. This chapter will interpret and consider the four dimensions extracted in a more qualitative manner. In line with the steps for MD analysis as outlined in Section 4.3.2, the following will consider the interpretation and usage of the dimensional functions, the factor scores of each text and the comparison of texts based on their dimension scores.

¹³ The formula is standard for calculating the distance between two points in space. It follows essentially the same logic as the Pythagorean theorem used to calculate the length of the hypotenuse in a two-dimensional right angled triangle, $c^2 = a^2 + b^2$, only here it is used for a triangle in a four-dimensional space.

5.1 Interpretation of dimensions

This section will interpret the rhetoric purposes of the dimensions extracted and thus focus on the second research question of this study, that is, how different metadiscourse marker categories co-occur with each other. The interpretation of the underlying latent function is based on the functions of the items within that factor.

The Factor Analysis conducted for this study resulted in four factors. The structure of these factors is given in Table 5.

| | |
|---|--|
| <p><u>Factor 1</u></p> <p>Attitude markers 0.748</p> <p>Self-mentions 0.468</p> <p>Transitional markers 0.414</p> <p>Hedges 0.348</p> <p>Boosters 0.230</p> <p>(Deontic modals) (0.200)</p> <p>(Attributors) (-0.124)</p> | <p><u>Factor 2</u></p> <p>Inclusive We 0.983</p> <p>Deontic modals 0.253</p> <p>(Self-mention) (-0.109)</p> <p>(Textual organisers) (-0.113)</p> |
| <p><u>Factor 3</u></p> <p>Reader address 0.701</p> <p>Questions 0.407</p> <p>Textual Organisers 0.321</p> <p>Commentary 0.210</p> <p>(Self-mention) (0.146)</p> <p>(Boosters) (-0.110)</p> <p>Code gloss -0.219</p> | <p><u>Factor 4</u></p> <p>Attributors 0.620</p> <p>(Transitional markers) (0.256)</p> <p>(Boosters) (0.118)</p> <p>(Self-mention) (-0.158)</p> <p>(Textual organisers) (-0.137)</p> <p>(Code gloss) (-0.197)</p> |
| <p style="text-align: right;">$\chi^2 = 51.29, df = 32; p\text{-value} = 0.0167$</p> | |

Table 5: Extracted factors and loadings above 0.1

The cut-off point in the present study is |0.30|: only variables scoring higher (bolded in Table 5) are taken into consideration in the interpretation of the dimensions and only in the dimension on which they score the highest. In Table 5 lower loadings are marked in parentheses. Those with low loadings will not be considered further as part of the dimension. As noted earlier, negative loadings indicate inverted correlation: for example, the above-average use of Reader address, scoring 0.70 in Factor 3, predicts the below-average use of Code gloss, scoring -0.22. However, none of the negative loadings are lower than -0.30, meaning that no inverted correlation is present in the dimensions.

Dimensions are factors for which the underlying function has been interpreted. The dimensions found in the present study are the following:

Dimension 1: *Writer-oriented text*

Dimension 2: *Solidarity*

Dimension 3: *Reader-oriented text*

Dimension 4: *Intertextuality*

The reasoning behind these names will be clarified in Sections 5.1.1–5.1.4 along with a closer interpretation on the metadiscourse markers. Only variables with loadings above $|0.30|$ are considered in the dimensions. Variables that scored below $|0.30|$ will be considered separately in Section 5.2.

5.1.1 Dimension 1: Writer-oriented text

The first dimension is to be related to expressing attitude. Out of the four dimensions, Dimension 1 accounts for most variance, 26.8 %. Four metadiscourse markers load on this dimension: **Attitude markers**, **Self-mentions**, **Transitional markers** and **Hedges**, the metadiscourse marker category with the highest loading being Attitude markers, scoring 0.75. Attitude markers express the author's own viewpoint on a matter, both through verbs expressing emotion or cognitive process or through adjectives or adverbials where the source of the viewpoint is not generally stated, merely understood to be the author's (e.g. "*hopefully*, it is..."). The purpose of Hedges is to withhold commitment to a claim (Hyland, 2005:52). It appears that these occur frequently when the author makes predictions on their own opinions, so as to allow alternative opinions or present the author as reasonable, something also noted by Fu and Hyland (2014). For instance:

- (13) "And *I suppose* that's also one of the basic functions it serves in the society *in general*." (article178)

On the other hand, one sub-category for Hedges is Approximators (e.g. *around*, *nearly*) used to estimate numbers or values, which is less related to opinion. Thus, the function of Hedges as softening an estimate or likelihood is perhaps more related to making generalizable claims by rounding numbers. Nevertheless, softening statements in opinion writing conceals the persuasiveness and increases reasonableness, making the persuasion perhaps more effective, as readers are allowed to draw their own conclusions (Fu & Hyland, 2014).

Self-mentions relate to the author's self, but their function is twofold: they are used for telling of one's own actions and experiences, of self as a private person, but also for commentating on the communication situation and its ideas. Ädel (2006:30-31) distinguishes between 1SG pronouns that refer to the self as an autobiographic

so-called “text-external” individual and ones that refer to the self as the writer of the text (e.g. *I will analyse*). Such a distinction has not been made in the present study. Both functions are nevertheless used to the purpose of creating and maintaining a writer persona, as in example (14):

- (14) “I’ve seen the climate movement build from handfuls of people here and there, to hundreds and thousands -- and even hundreds of thousands, in rare moments. *I believe* that we’re poised for a growth far more explosive” (article151)

The first metadiscourse marker is a Self-mention, relating to personal experiences to contextualise and back-up an idea. In a personal topic, such as climate activism, the line between the two selves may be fuzzy, but the second marker, an Attitudinal cognitive verb, primes a thought by self as a commentator of the world and creator of discourse: the reader is given space to disagree. Using Attitudinal markers an author with their own perspective enters a dialogue where Self-mentions and Hedges make room for the reader’s critique. In this corpus, this commentator function of the pronoun was found to be more common than the reference to self as a private person. Following Crismore and Farnsworth’s (1989) connection between metadiscourse and ethos, here the *ethos* of the author is acknowledged as a narrator who is involved, has attitudes but is also willing to hedge¹⁴.

Less related to the author’s self, Transitional markers also load on the first dimension. Transitional markers connect propositions (*also, likewise, moreover*) and explain causalities (*thus, because, since*), but also indicate how the propositions contrast to expectations (*but, however, nevertheless*). These conjunctions are thus used to connect the writer’s ideas and combine them into a narrative. Because of the rhetoric uses of the two highest loading markers I will term Dimension 1: *Writer-oriented text*.

5.1.2 Dimension 2: Solidarity

Although Dimension 2 explains 20.5 % of the variation, only one metadiscourse marker category loads high enough to be part of it, namely **Inclusive We**, i.e. 1PL-pronouns including in its referent both the speaker and the audience. This category is not present in the framework suggested by Hyland (2005), who considers all 1PL-pronouns Self-mentions. However, Ädel (2006:31) argues 1PL-pronouns can

¹⁴ One may note that there is another metadiscourse marker category, concerned with author’s commitment to a claim, namely Boosters emphasising the certainty of a statement. Recalling Table 5, Boosters, too, load on Dimension 1, although not high enough to be considered as part of it.

function as either audience-inclusive or audience-exclusive, but only the latter is a form of Self-mention. The distinction is also acknowledged by Zhang (2016). Admittedly, the metadiscourse categories in Zhang's MD analysis differ from those of the present study, but it is interesting to note that in his corpus of various genres, Reader-address and Inclusive We fall in the same dimension (2016:212). In the present study, Inclusive We forms a dimension of its own, separate from both Reader-addresses and Self-mentions.

Inclusive We is often used in contexts that refer to shared knowledge, what we all can be expected to know. However, in the data analysed here, Inclusive We is most frequently used to create a sense of unity or solidarity where *we* refers to 'citizens' in collocations such as *our country* or *our laws*. Another common group reference is to the part of society that could be assumed to relate to the text content, e.g. *us liberals*. Deontic modals (e.g. *must* and *need to*) do not load high enough to be considered as part of Dimension 2, but it can be noted that often they appear together in statements where the writer reaches out and urges "us", the writer and the reader, to do something, as in example (15). Observe that this example is from the same text as example (14), but here the author shifts from personal experience to proposing what common people could do together to help people in power:

- (15) "And *we* can help them [people in power]. *We* can't leave the hard work to other people. *We* have to stand up for the world *we* love. Look outside your window, and understand that if *we* don't succeed everything you see will be profoundly changed in the coming decades." (article151)

The plea comes off as strongly persuasive. Calling a reader to action is easier, if the writer themselves is ready to act in the same team as the reader.

On the other hand, Khabbazi-Oskouei (2003:101) notes that the Inclusive We can be used "indirectly to refer a third party, usually the government." In many of the texts in the data the pronoun is indeed used in contexts referring to a society or a nation (e.g. "We are at war"). It reminds the reader of the fact that they may be personally affected by society. Whether or not the writer and the imagined reader have actual agency in the decision-making of the society, the rhetoric constructs common ground.

Inclusive We appearing in a dimension separate from the writer-oriented *I* or the reader-oriented *you* suggests that texts using this metadiscourse marker are unwilling to make a distinction between the reader and the writer when it comes to

real-world matters. Thus, the function of Dimension 2, used to unite writer, reader and/or nation, will here be termed *Solidarity*.

5.1.3 Dimension 3: Reader-oriented text

The metadiscourse marker categories on Dimension 3 are **Reader-address**, **Questions** and **Textual organisers**. While the first two are prototypical interactional markers, Textual organisers are interactive markers structuring the text by referring to the position or order of sections or the discourse topic of a marker. The dimension explains 12.6 % of the variation.

Whereas Dimension 1 is concerned with the writer-oriented side of interactional metadiscourse, reader-address and rhetorical questions in Dimension 3 are rather on the reader-oriented side. These metadiscourse markers overtly acknowledge the presence of the reader. Questions and directives are presented to the imaginary reader in an attempt to appeal to them or to make assumptions on their thoughts in order to answer their questions. Question-Answer patterns in a text are also a form of this, as the question is presented as if a query the reader might have. Directives are used to position the reader with expressions such as *consider*, *imagine* or *look* but on less equal grounds than with Inclusive We in expressions like *let's go*.

The pronoun *you* is often used to ascribe a role to the reader, such as the role of a person with certain characteristics or someone for whom the issue discussed is relevant (see Thompson & Thetela, 1995:120). Ädel (2006:34–35) notes that the pronoun *you* does not necessarily refer to all imaginary readers but to a category defined in the context. For instance, it can refer to females even though it is known not all readers will be women. In certain contexts, Ädel continues, it can be taken to function as a generic *you*, much like impersonal *one*, and sometimes it can even be taken to mean the writer themselves. Such a fine-grained distinction has not been made in the annotation here as the use of *you* can engage the imaginary reader whether or not it actually represents the real-world reader. In the data studied here, *you* usually has a generic function as a more casual version of *one*, but in some instances it appears to refer to the hypothetical reader. The distinction is not always clear-cut, such as in the example below:

- (16) “If *you* are not for abolishing solitary confinement, then *you* must not believe that solitary is torture. [...] Once *you* accept that solitary is torture, then there is no moral alternative to supporting the fight to end solitary.” (article149)

Here the hypothetical reader still needs convincing of the main argument of the text, although many regular readers of the site may already agree with the writer. Here *you* can be argued to be more impactful in terms of persuasion than a generic *one*.

Textual organisers load low on the factor and are quite different from the other two marker categories. In many ways, they are similar to Transitional markers, but rather than connecting single ideas, the organisers refer to sections or order of sections in the text or to the writer's intentions. However, many of the Textual organisers are much like directives, such as the Announcer *see table below* or the Discourse goal *let me offer an alternative*. Textual organisers vary in how they refer to the text: Announcers and Discourse goal markers often refer to sections of text explicitly, e.g. *the following section*. Conversely, Sequencers, typically refer to the order of sections, e.g. *secondly, finally*, but do not refer to these as textual items. Nevertheless, they all structure arguments and create a linearity, signposting a path for the reader to follow. Compared to Transitional markers, Textual organisers work across sections and thus serve to inform the audience of what is happening in the text on a higher scale: where the section is in relation to other sections (e.g. arrangement or order of sections); what the purpose of a paragraph is (e.g. to summarise or to provide an example). It can thus be argued that Textual markers guide the audience more than Transitional markers do. In terms of the opinion text genre, it should be mentioned that almost all Textual organisers in the corpus appear to refer to sections of text as opposed to illustrations, graphs or photos. This may be because the illustrations that accompany the texts are thought of as self-explanatory or as merely decorative.

As the two highest loading categories are both reader oriented interactional metadiscourse marker categories, the main rhetoric function of Dimension 3 appears to be in some ways the opposite of Dimension 1 (Writer-oriented text). The Textual organisers, too, serve chiefly to guide the reader's comprehension. Thus, Dimension 3 will here be termed *Reader-oriented text*, as its chief purpose is to engage the reader through Addresses, Directives and Questions and to direct the reader using Textual organisers.

5.1.4 Dimension 4: Intertextuality

As many of the items in Dimension 4 loaded higher in other factors, only one metadiscourse category loads on this dimension. However, its importance should not

be neglected: **Attributors** are expressions used to source a statement or to refer to someone else's statement; note that hedging verbs (*seem, appear*) and expressions of hearsay are not Attributors despite the similarities. This fourth dimension explains 11.0 % of the variation.

The purpose of Attributors is often to refer to an authority and so add persuasive force to the argument, but Crismore et al. (1993:52–53) note that often the intended rhetorical purpose remains the same even when the quoted source lacks the status of an authority. On the other hand, as noted by Khabbazi-Oskoueï (2013:98), Attributors can be divided into “certainty markers” when the source is a reliable authority and “uncertainty markers” when the source is discredited. In the corpus of the present study, the quoted sources are usually attributed to reports in newspapers or television and to spoken or written statements by individual people of some political or societal power. Yet, without a close reading of the context it is difficult if not impossible to say whether the source is considered credible, as even the quotes of authority figures, e.g. politicians, are not always considered trustworthy claims. Instead they are being reported on as a part of the journalistic aim to transmit information. In example (17) authorities are sourced using the neutral phrase *according to*. However, in the sentence that follows the use of the word *insists* implies that the writer may not consider the source reliable.

- (17) “*According to* Turkish authorities, the raids killed some 200 Kurdish fighters. Ankara *insists* that the YPG is affiliated with the outlawed Kurdistan Worker’s Party (PKK).” (article240)

At times, not even the word choice or the immediate context reveals the author’s stance and one needs the whole text to know how the author feels about a particular source. Arguably, even if the source is considered unreliable, referencing to other sources is beneficial to the author, as this indicates the author’s familiarity with the topic, the latest developments and awareness of a larger discourse community, one that the author themselves may belong to as a commentator.

Attributors often connect a source to quoted passage but generally they seem to source paraphrased content. Especially with paraphrased content, sourcing is important for separating between the ideas of the writer and those of others. While perhaps not representative of a rhetorical strategy, Dimension 4 is concerned with sourcing of both reliable and unreliable statements. Acknowledging Ifantidou’s

(2005) and Boggel's (2009) distinction between intratextual and intertextual metadiscourse, this fourth dimension will here be termed *Intertextuality*.

5.2 Other metadiscoursal markers

Four metadiscourse categories did not load above ± 0.30 on any dimension (see Table 5, Section 5.1). This means that while the categories do occur, there is not enough variation in their use. The functions of these four categories, **Boosters**, **Deontic modals**, **Code glosses** and **Commentaries** will be briefly discussed here.

Boosters, briefly touched upon in the discussion of Dimension 1, are interactional metadiscourse markers that emphasise the certainty of the author. In opinion articles Boosters are needed to make the author's views explicit (Dafouz-Milne, 2003:45). With Boosters the author can close opposing arguments, but in doing so they can create a sense of solidarity with the audience who draw the same conclusion as the author (Hyland, 2005:52–53). In the corpus Boosters were most often expressed as adverbials *of course*, *in fact*, *clearly*.

Deontic modals were retained as their own category in this study due to differences in previous frameworks. In the framework by Hyland and Tse (2004), based on academic writing, Deontic modals and semi-modals function as Engagement markers alongside of Directives, but in the framework by Dafouz-Milne (2003), based on opinion articles, they function as Attitude markers. The modal *must* often collocates as *we must* as a form of engagement (Inclusive We), but most modals (*should*, *have to*, *need to* inter alia) refer to third parties or occur in passive constructions. They are used to describe e.g. what political personages or entities should do in the opinion of the writer. At times, the recommendation is in passive voice. Example (18) gives instance of both:

- (18) “the full biosafety dossier for GM mustard *must* be uploaded on the Ministry's website, the GEAC *should* disclose its full agenda notes and minutes for each meeting” (article183)

The topic being an official dossier, the author's request cannot not apply to a generalised reader. As most uses of Deontic modals in opinion texts are of this type, the Deontic modals of the corpus at hand chiefly function not as Engagement markers but as Attitude markers, as proposed by Dafouz-Milne (2003).

Code glosses are an interactive metadiscourse category used to explicate concepts or words by giving a definition or an illustrating example. This category was difficult to annotate automatically: often it occurs with parentheses or with

certain set expressions such as *for example* or *in other words*, but a definition could just as well be inserted into the text flow without any priming, and thus the analysis focuses on the explicit expressions of glossing. Example (19) illustrates the use of a parenthetical Code gloss. The article concerns a Supreme Court ruling on Board of Control for Cricket in India (BCCI) according to which members' terms in the board should be limited.

- (19) "The BCCI's argument is that in no other sphere of activity is there a "cooling off" period (*three years followed by a cooling off period of three and a maximum of nine years in office*)."

Here the Code gloss informs the uninitiated of the length so called cooling off period, a detail necessary for readers to form their own view on the issue.

The category of Commentaries in this study includes what Dafouz-Milne (2003) terms "Asides", as well as conversational devices. Conversational devices gain their pragmatic meaning from their context, but as they imitate the spoken language they are used for interacting with the reader and can sometimes set a joking tone (Myers, 2010:111). In the corpus, Conversational devices were not frequent and their use followed standard spelling and use, the most common being *yes* used for emphasis and *oh* to express being reminded of something (*oh, and by the way...*). With Asides, a form of Attitudinal marker, the author adds their explicit appraisal on the topic as a parenthetical, creating solidarity and familiarity. According to Dafouz-Milne (2008:104, 108), Asides can be interpreted more freely than explicit opinions formed as full sentences but they nevertheless signal the author's stance. In the corpus, passages separated from the text flow was marked as an Aside, if the information it contained seemed subjective or superfluous, that is, if it cannot be considered as Code gloss. The following is one example:

- (20) "And it's reasonable to believe that people with high incomes are richer than the rest of us. So if we have progressive tax rates – *without too many loopholes* – then the income tax could be a tool to narrow the gap between rich and regular."

The first sentence is humorous in its remark on the connection between income and wealth. The Aside in the following sentence continues this style, seemingly specifying which kind of progressive tax rate, but in fact providing a somewhat snide remark, perhaps suggesting rich people are likely to commit morally questionable activities in order to save money. Humour is a frequent function of Asides in the corpus.

5.3 Characterising Opinion Texts

Having named the functions of the dimensions in a macrolevel analysis in section 5.1, a closer study is required for the results to be validated and the descriptions be expanded upon. In this section texts representative of each dimension will be looked at using discourse analytical methods. The goal is to find out how aspects of a text as communication explain its dimension scores, i.e. the sum of the standardised scores Z for each metadiscourse marker belonging to the dimension.

As dimensions are structured according to the correlation of metadiscoursal categories, texts can be expected to generally not have high dimension scores in more than one dimension. The dimensions provide the answer for the first research question posed in this study: for what purposes is metadiscourse used in online opinion texts? From the interpretation of the dimensions in Section 5.1, one finds that metadiscourse is used to negotiate the role of the writer and the reader: Writer-oriented, Reader-oriented and Solidarity rhetoric, that is (1) a rhetoric with the writer in focus, (2) a rhetoric where the reader is approached, and (3) a rhetoric where the writer and the reader are united. These can be seen as three distinct rhetorical strategies in approaching the reader. Unlike the previous three, the fourth dimension, Intertextuality, does not perhaps translate into a rhetorical strategy, but it is used to source intertextual content, whether it is to reliable or to somewhat less trustworthy authorities.

As dimension scores denote only the distance from the mean, they do not tell an analyst about the absolute or relative frequencies themselves. The frequencies of individual marker categories will be briefly described in Section 5.3.1. Section 5.3.2 considers the dimensions qualitatively: in describing metadiscourse in texts that have dimension scores close to the mean, it explores the concept of a *prototypical* opinion text. Following this, Section 5.3.3 considers with the amount of variation that exists within the genre, primarily with texts with particularly high or low dimension scores.

5.3.1 Overview of frequencies and validity issues

All texts in the corpus have at least some metadiscourse markers, but it is common for a metadiscourse marker category not to be represented in a text whatsoever. On average, a text contains 51.48 metadiscourse markers per 1000 words, but the standard deviation is 20.40, which indicates that texts vary considerably. Focus on frequencies is not integral for MD analyses, which generally concentrate on findings

within the dimension continua. This section is included here as a comparison point to more traditional metadiscourse studies as well as to address validity issues. Appendix 2 features the frequencies and standard deviations of all 13 metadiscourse marker categories used in the framework of this study.

The categories with the two lowest frequencies are Deontic modals and Code glosses, which is most likely the reason for neither category loading high on a dimension: small variation is not enough for a variable to load significantly on a dimension. The low frequencies may be caused by the annotation process. Unlike broad categories such as Attributors or Attitude markers, Deontic modals is a category made up of a small, closed class of words, thus easily dwarfed by other categories. Meanwhile, the annotation of Code glosses is based on explicit markers including parentheses and punctuation. It should be noted that Code glosses are in reality probably more common than indicated by an analysis focusing on a limited set of explicit markers. The same, of course, is true for the fourth least frequent category, Attitude markers: as was observed in Section 2.2, an absence of metadiscoursal attitude says nothing about the implicit ways a writer can use to express their opinion on a matter.

Textual organisers form the third least frequent category: here the low frequency is likely caused by the properties of the genre. The short length of most texts translates to less need of guiding the reader, especially when it comes to announcing future or reminding about previous points. In general, categories listed as interactive metadiscourse are less frequent than interactional markers. The connection between short length and lack of interactive metadiscourse was also noted by Le (2004), who studied editorials. This is not to say that all interactive metadiscourse markers would be rare: the metadiscourse marker category with the highest mean frequency is Transitional markers, at 10.94 occurrences per 1000 word tokens. This is perhaps unsurprising, as the category contains common conjunctions used to create cohesion between sentences. Unlike Textual organisers, they do not presuppose the text to contain quotes, hyperlinks or multimedia that is being referred to. Moreover, their function is stylistically neutral, as opposed to, say, Self-mentions, which create a more subjective text. Transitional markers include very common conjunctions (*and*, *but*, etc.), but only those with a metadiscoursal function were counted in the corpus, wherefore the multifunctionality of the lemma should not skew the results.

Hedges have the second highest frequency and thus they are the most frequent interactional metadiscourse marker category, but at 6.12 occurrences per 1000 word tokens it is much less frequent than Transitional markers. The goal of Hedges, too, may be a style that sounds objective, especially considering how Boosters occur only 3.35 times per 1000 words. Admittedly, Boosters as a category is more narrow in terms of possible markers than Hedges, which also contains approximators (see Appendix 3). However, Hedges have been found to be more frequent than Boosters also in the opinion texts studied by Fu and Hyland (2014). Fu and Hyland's frequencies for Attitude markers in opinion texts and popular science texts, 1.8 and 1.1 respectively, also resemble the frequency observed in this study, 1.94. Their frequency for Self-mentions is 6.7 in opinion texts and 4.5 in popular science. The corpus of the current study, with a frequency of 4.69, bears closer resemblance to the popular science genre, but with a standard deviation of 7.86, this closeness is of less consequence.

In conclusion, while some of the frequencies may be misleading due to the annotation process, the results in general appear to be in line with previous studies on metadiscourse in related genres. The metadiscourse marker categories are not equal in size, wherefore categories with several sub-categories may naturally have a higher frequency than categories consisting of only one specific function. However, if the corpus is systematically annotated, the size of the category should not play part in a MD analysis.

5.3.2 The prototypical Opinion Text

The observations made so far, have mainly concerned the frequencies of individual metadiscourse marker categories. This section will turn to describing dimensions qualitatively through close readings of two prototypical texts from the corpus, that is, texts which represent the mean dimension score. As defined in Section 4.3.2, *prototypicality* refers here to a text's distance to the mean, with low prototypicality scores indicating high prototypicality. In the previous section, the mean frequencies of metadiscourse categories were discussed. A prototypical text may not have a mean of a close-to-mean frequency of each metadiscourse marker category, or even a close-to-mean dimension score on a specific dimension. Instead the prototypicality is evaluated across all dimension: a prototypical text is evenly using metadiscourse markers from each dimension – no dimension is over- or underrepresented. Thus,

studying a prototypical piece of online opinion writing is useful in order to get an idea of what texts in the corpus look like in terms of the metadiscourse dimensions.

The text with the highest prototypicality has a prototypicality score of 0.55. From a purely digital news site, the article concerns media's portrayal of the Syrian conflict. The article has a Dimension 3 score of 0.09 and the dialogic Reader-oriented dimension comes into play for instance when Questions are presented as if something the reader is asking.

- (21) “*So* what is the truth on Syria? In the last five and a half years, since a regional uprising turned into an armed rebellion – *turned into civil, regional and international war* – ‘the truth on Syria’, has been segmented into many self-tailored ‘truths’” (article218)

Given the question and the informal *so*, the tone is conversational, but given the Code gloss, perhaps also didactic. *So* also indicates a topic shift, further aiding the reader. The conversational tone continues later on, when the Conversational device *yes* functions as if a reply to an expected question.

- (22) “*Yes*, these journalists exist, *but* they fight against many odds. [...] *And* good journalists, are either forced to, albeit begrudgingly, toe the line or to stay out of the discussion altogether.” (article218)

Example (22) also illustrates the use of Transitional markers as logical connectors between sentences. The Transitional marker loads on the Writer-oriented dimension, but explicitly writer-oriented markers, such as 1SG pronouns, are not present in the text. Moreover, while the text was observed to be dialogic and make use of markers loading on the Reader-oriented dimension, it lacks explicit address in the form of *you* pronouns.

One text from a purely digital source discusses the so called “war on drugs” politics in the US. In prototypicality it measures 0.67. The premise the article tries to convey to the reader is that overly harsh criminalization of the selling and using of drugs will merely hurt society as a whole. The paragraphs are short in length although this may be due to the site layout. A new paragraph often opens a new argument or a new topic, and this is often signalled by a metadiscoursal marker, frequently a Transitional marker, but at times also by Textual organisers or Attitudinal markers. Example (23) is an extract that follows just after a passage that has given examples of laws that give hard punishments (e.g. murder charges or life imprisonment) to people found guilty of selling drugs that have resulted in a death.

- (23) “*Of course*, these laws aren't applied equally: The systematic targeting of people of color by police and prosecutors means that many drug arrests are a result of racialized criminalization.” (article150)

Here the author uses *of course* to introduce a new topic (racialized criminalization) but also to anticipate a question or counter-argument the hypothetical reader might have had in mind. Besides the Reader-oriented Textual organiser, the rhetorical style uses the pronoun *you* at times, but mainly as a general *you*.

- (24) “Those who sell drugs have been classified as the “bad guys,” and when *you* end up in that category, the current public distaste for the drug war will hardly save *you*.” (article150)

It cannot be claimed that the generality of the pronoun fails to create a rhetoric effect. Personal pronouns, while used sparingly, bring the issue closer to the reader. Sellers of drugs are not otherised with a *they*, but rather as people who could potentially be among the audience. The text has a Dimension 2 score of 0.06 and Inclusive We is used in connection to Questions: after the writer has laid out the facts, the audience is reached out to as members of the same society as the writer. This is illustrated by example (25):

- (25) “[...] the drug war will not end until *we* stop blaming drug-related problems on “criminals.” *We* need to ask *ourselves*: If no one were “criminal,” what would *we* do to build a society that fostered health and life?” (article150)

The issue discussed should be of concern also for the readers. Attributors are used to point to authoritative sources, researchers or activists. It is, however, also used to attribute quotations to sources presented less favourably. For instance:

- (26) “In August, Maine's notorious Gov. Paul LePage, who has deployed overtly racist myths to advance his state's drug war, *announced* that the vast majority of dealers arrested in his state are Black and Latino.” (article150)

The paraphrased content is here attributed to a *notorious* governor, whose past misconduct is also mentioned. Clearly this is not a source the author feels can be trusted.

The close reading in this section sheds some light on the uses of metadiscourse, both confirming some of the functions of the dimensions and giving examples of actual usage. For instance, paragraphs with new or alternate arguments can be opened by Transitional markers, Textual organisers or Attitudinal makers. In texts where the use of *you* or the inclusive We is prototypical, the purpose is to bring the argument to the reader and the effect is thus used more sparingly or, as may be the case for *you*, only as a general reference. More importantly, texts use metadiscourse

from different dimensions, both to ask questions in a Reader-oriented rhetoric and to organize arguments using Transitional markers from the Writer-oriented dimension. It is also to be noted that although a text appears to load on a Writer-oriented dimension, it does not necessarily contain many explicit mentions of the writer: its metadiscourse may merely be working on the writer's ideas through Hedges and Transitional markers.

With the notion of prototypicality based on a calculable score, the texts for close reading can be chosen without author bias. However, by reading prototypical texts, little information is gained on outliers. To understand the use of metadiscourse in texts with exceptionally high or low dimension scores, a look on the whole corpus is in order.

5.3.3 Opinion Texts across dimensions

The discussion so far has considered the mean of the frequencies of metadiscourse categories and dimension scores. As variation in the corpus does exist, this section will look at causes for a particularly low or particularly high dimension score. The aim is to see how the dimension scores reflect the topic or purpose of the text. This is not done in order to comment on individual texts but rather to evaluate the interpreted dimension functions and to observe them in action. As discussed above, the ranking of two dimension scores can also be used for visualisation in plots, where the scores serve as coordinates. This visualises the dispersion and helps in noticing outliers. Note, however, that the scores of different dimensions are based on different factor structures and thus not comparable in unit.

As Dimensions 1 (Writer-oriented text) and 3 (Reader-oriented text) both relate to the parties involved in the communicative event, a description of the dimensions will provide some fundamental insight in how authorial voice and the audience are construed in the text. Figure 2 compares the ranking of texts across the two dimensions.

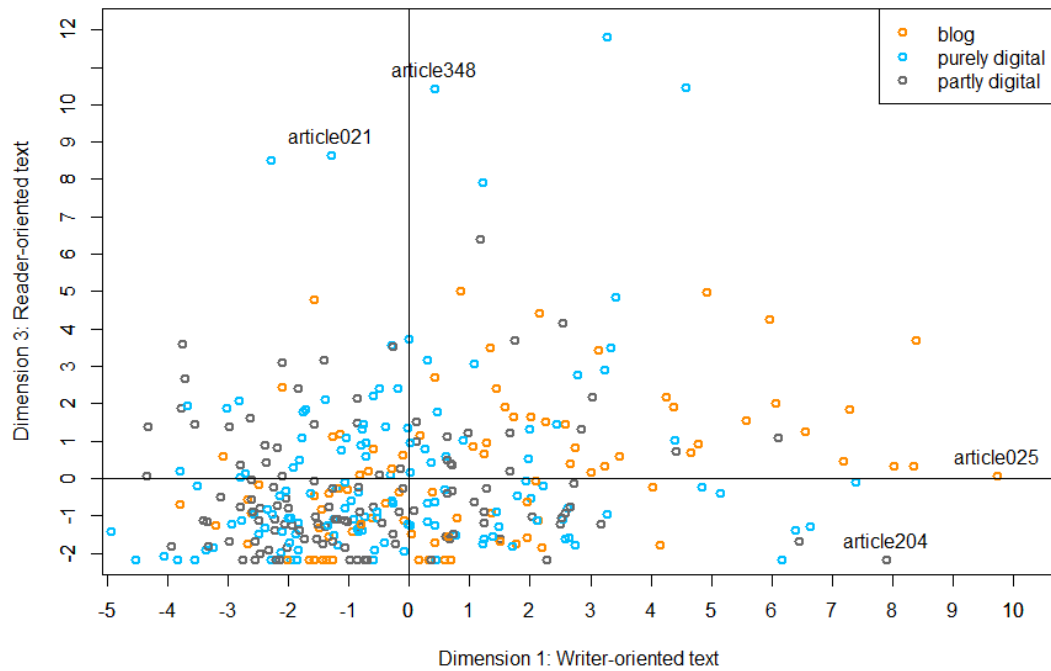


Figure 2: Texts in corpus plotted on Dimensions 1 and 3

Comparison of Dimensions 1 and 3 seems natural as their function is, in a sense, the opposite. Moreover, these two Dimensions consist of more than one metadiscourse marker category. As in Figure 2, most texts do not score particularly high or low on either dimension and cluster in an area of $-3 < x < 3$ (Dimension 1) and $-2 < y < 2$ (Dimension 3). This implies that there are certain metadiscoursal tendencies that are common among the majority of texts of this genre. In Dimension 3, many texts have a score of just below -2. These are texts where the frequencies of Reader Address, Questions and Textual organisers are all zero. However, most texts plot higher than this and especially on Dimension 1 the texts are more evenly dispersed. Several texts load low on Dimension 3: these texts use no Writer-oriented metadiscourse whatsoever.

The results indicate that the source of a text affects its metadiscoursal style. For instance, most of the texts higher in the Writer-oriented Dimension 1 are blogs. Though not all blogs are concerned with presenting their own viewpoint, unsurprisingly there appears to be a clear continuum and reference to the author's self or their own opinions is more common in blog writing, with blogs serving as outputs for personal thoughts, as a soapbox of sorts. Some opinion texts from news sites also take on a more personal, blog-like style and score high on Dimension 1 but

low on Dimension 3. However, the mean on Dimension 1 is lower for these categories than for blogs. For example, texts from partly digital sources show some variation but in general rank quite low especially on Dimension 1. Meanwhile, while the mean dimension scores on Dimension 3 for blogs and purely digital are quite alike, few blogposts take on as high a score in Dimension 3 as select texts from news sites. Partly digital sources generally score low on each dimension.

Texts scoring high on Dimension 1 use the Writer-oriented rhetoric to explain private viewpoints or to relate personal experiences. Take a text titled *Hillary's Night* as an example: the text's dimension score on Dimension 1 is 9.74 but on Dimension 3 it only scores 0.08. The blogpost is the author's reaction to a televised debate between presidential candidates in the U.S. 2016 election. The metadiscoursal markers are chiefly Attitude markers and Hedging (which by default coincide with Self-mention). Metadiscoursal markers of Dimension 1 has been italicised in example (27):

- (27) "*I mostly disagree* with Jeb Lund that Clinton should have been more assertive and gone in for the kill. Yeah, she missed a couple of opportunities to stick a knife between his ribs, so to speak, *but* she was playing rope-a-dope very well, *I thought.*" (article025)

Through the use of Hedge *I thought*, the author explicitly shows that these are their personal thoughts brought into a dialogue. As a rule, hedging is common in blogging, which accounts for their high score in Dimension 1. By acknowledging a differing opinion but hedging their disagreement, authors are less open to criticism on their viewpoint. A column from a partly-digital news source that scores high on Dimension 1 (7.90) but low on Dimension 3 (-2.18) is article204. Compared to article025, this article contains less Attitude markers and more Self-mentions as in example:

- (28) "As a woman, people assume *I'm* voting for Hillary Clinton. *And* as a conservative, people assume *I'm* voting for Donald Trump. *But*, the truth is - *I'm* still undecided and there are so many others out there like *me.*" (article204)

The author also uses Transitional markers to connect the contrasting expectations laid on her. The narration is very personal and more monologic in style: the author uses her personal experience to build a perspective from which she weighs the presidential election candidates against each other. By doing this she can speak with the voice of all women feeling the same. This particular section is more monologic than dialogic or persuasive, and thus there is no need for Hedges to defend the argument. The article scores low on Dimension 3 as it does not contain any of the

metadiscourse markers of this dimension, that is, it makes no overt reference to the reader.

Texts scoring high in Dimension 3 do not necessarily make overt reference to the writer's presence; instead they are concerned with the reader or textual elements. Questions are used especially in purely digital sources. Article348 scores fairly high in Dimension 3 (10.43) but low in Dimension 1 (0.42). Published on an purely digital newssite, the article "Ten questions about the Richmond Park by-election" is arranged in a question-answer format, which adds to its above-average number of questions and, consequently, to its high score in Dimension 3. Some of the questions are presented as something a reader might be asking and the article provides an answer; others are questions the author, too, is asking as though voicing the shared concerns of the reader and paper. For instance the second out of the ten questions:

- (29) "2. *How posh is Richmond Park?* Well someone did have the task of delivering our tabloid to a certain Royal elector who actually lives inside the park – *you can't* get much posher than that." (article348)

Notice that the question is numbered. This systematic numbering of the topics covered is a Textual Organiser and accounts for a large part of the high dimension score and there is only a small amount of explicit reader address. Compare this to the dialogic metadiscourse marker example (30):

- (30) "*Here's* another thing to consider: The thief may keep the new phone, which still has *your* number, to gain access to *your* online accounts via the two-factor authentication process" (article021)

Compared to example (29) where *you* is quite general, example (30) utilises Reader Address in the form of a more specific *you*-pronoun. The point of reference can be seen as referring to the reader given how the situation described (identity theft) is hypothetical. The tone throughout the text is advisory as it goes on to describe what "you" could do in said hypothetical situation. Note that the excerpts above could be rephrased using passive voice or impersonal expressions: "one can't get much posher", "the new phone which still has the owner's number." Thus, besides reader-oriented rhetoric, the use of general *you* could simply indicate a more informal tone. However, the informal tone itself can be argued to be conversational and approachable to the reader. It is advisory or instructional, but also the voice of a peer or a friend.

In summary, texts with a high score for Dimension 3 are conversational by informing advising the reader. This style is more favoured by blogs and partly digital

sources than by purely digital sources. If blogs, indeed, see regular readers as “co-creators”, as suggested by Wall (2005), a Reader-oriented rhetoric makes sense. The same could hold true also for smaller specialised online news sites with limited audiences. If the score for Dimension 1 is low the writer remains more detached; this style is common in blogs and rarer in partly digital sources, potentially because the writers working for the online news wish to keep themselves, their private persona, detached from the issue at hand, thus imitating traditional newspapers, the opinion writings of which score low on both Dimensions 1 and 3.

As text of the same source type (blog, purely digital, partly digital) exhibit variation, it is fair to ask what kind of variation exists between different sources that technically belong to the same source type. Biber (1988) interprets the dimension score of a genre as the mean of the dimension scores of all texts belonging to that genre. Thus, in the current study the dimension scores of the sources can be analysed through the mean of scores for texts that have been published in the same source, that will here be termed as *source dimension scores*. It should be noted that some source dimension scores have a high standard deviation: this can be related e.g. to different authors or to different topics or approaches. As noted by Puschmann (2013:100), there is variation between the different writers for a single newspaper, but also internal variation in the style of an individual blogger. Another factor is the different number of texts chosen per source; as mentioned in Section 4.1, the corpus was collected to contain ca 10,000 word tokens per source, irrespective of number of texts. This means that the mean and standard deviation of the source dimension scores is based on slightly varying numbers of texts.

The source dimension scores confirm some earlier findings: while variation exists, blogs have the highest use of writer oriented metadiscourse and partly digital sources are the most homogenous. The source dimension scores of Dimensions 1 and 3 are plotted in Figure 3. Note how the range of the source dimension scores is smaller than that of text dimension scores. In fact, the ranges of source dimension scores mostly match the clustered area noted in Figure 2, which was noted to fall where $-3 < x < 3$ and $-2 < y < 2$. Again the partly digital sources fall together in a fairly constrained area. It can be noted that they share it with many sources that exist only in digital forms. Three blogs also fall into this area: two of them, Cato Institute and Worldwatch Institute, are in fact corporate blogs upheld by research institutions.

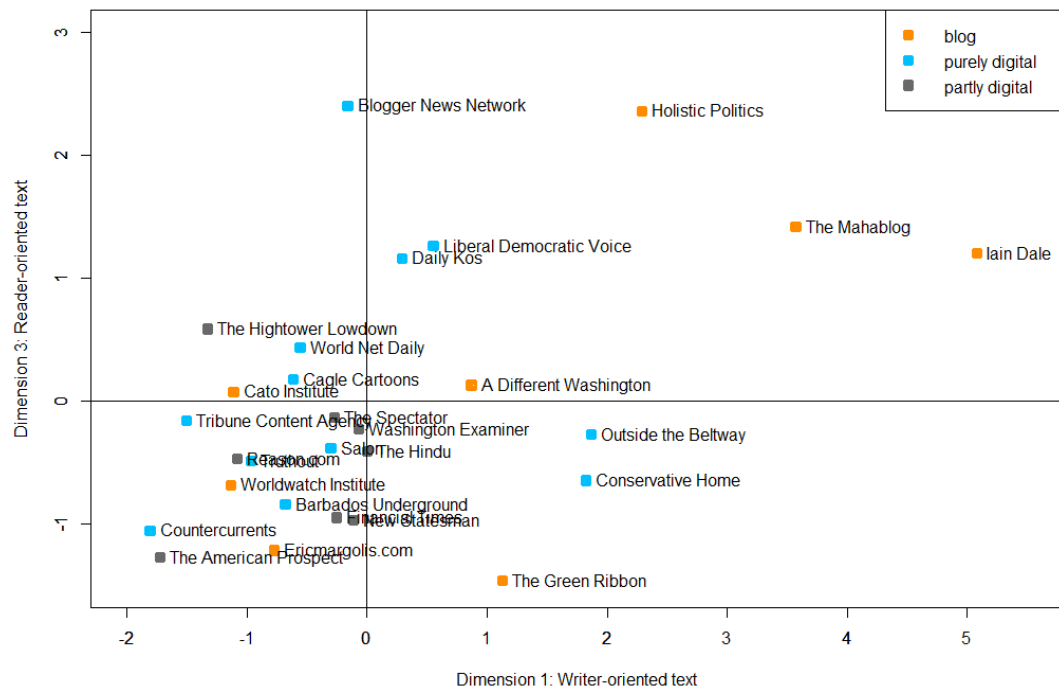


Figure 3: Source dimension scores plotted on Dimensions 1 and 3

However, the sources categorised as blogs make up the widest continuum, as is expected, considering the stylistic freedom and often personal approach of blogs. Sources that exist in purely digital format also appear to spread out, but more so along Dimension 3, indicating a less personal and a more Reader-oriented rhetoric. As hypothesised above, the authors of purely digital sources are generally not as inclined to make explicit mention of themselves but some do attempt to persuade by engaging their readers through the use of pronouns.

Like Dimension 3, Dimension 2 (Solidarity) is also concerned with reaching out to the reader, but Inclusive We brings forth the writer's presence. One text ranks low in all dimensions but Dimension 2: it is also a prime example of the Inclusive We used to equate *us* with a nation, namely the UK after its vote to leave the EU:

- (31) “The Government’s insistence that *we* should not give away *our* hand in negotiations with the EU has backfired. [...] *We* should define the worst that can happen and prepare for it. *We* can’t control how *our* opponents behave, but *we* should define all the things *we* can control and make sure *we* control them.”
(article383)

The Inclusive We does not refer to the government, however, as *the Government* is mentioned and construed as a separate entity from *us*. Instead it refers to the British people with whom the author stands. On an individual level, the people have little

agency in the negotiations between Britain and the EU, but their lives are affected by the result and may therefore feel as if they, too, play a part or are invited to be involved. Since the dimension only consists of one metadiscourse marker category, dimension scores are based entirely on the frequency of the Inclusive We.

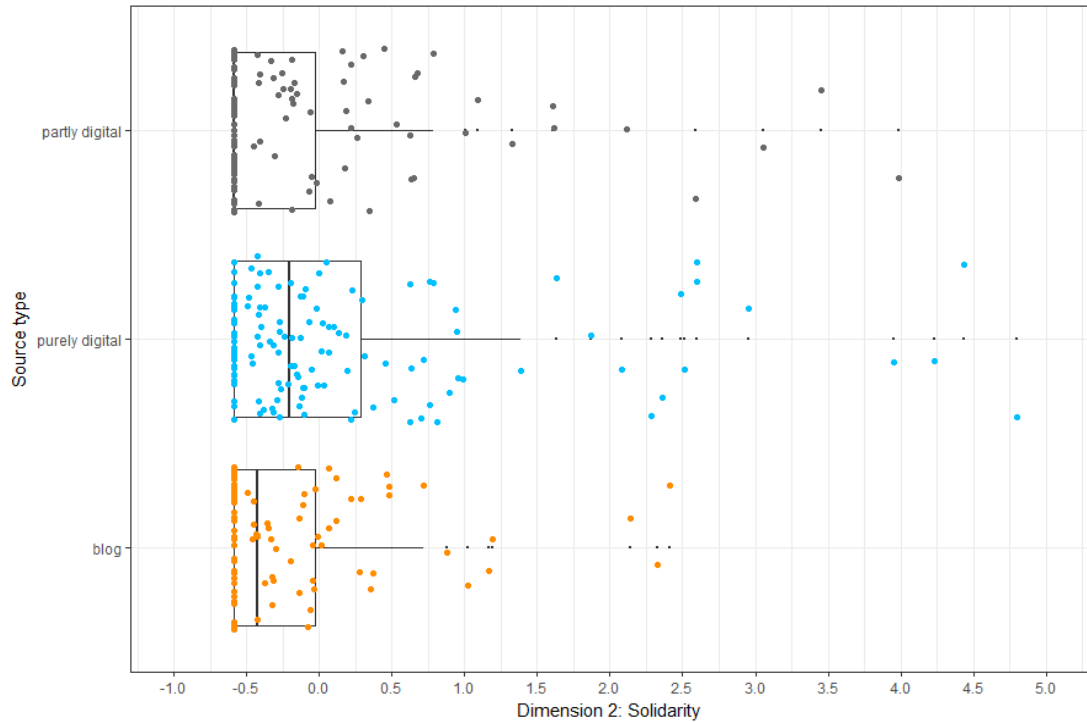


Figure 4: Texts in corpus plotted on Dimension 2

As can be seen in Figure 4¹⁵, quite a few texts in the corpus of the present study avoid the Inclusive We altogether, especially texts from partly digital sources. Here the lack of a lower 25% whisker indicates that at least 25% of the texts have zero instances of Inclusive We. Moreover, as shown in Figure 4, the median for partly digital texts is in fact also the minimum value, which means that at least 50 % of the texts have no instances of the marker category. While blogs were above shown to score high especially with Writer-oriented but also with Reader-oriented styles, a rhetoric of Solidarity is not as common. Meanwhile, purely digital sources display the highest median for dimension score in Dimension 2 and quite a few texts score exceptionally high. The use of Inclusive We is more prevalent in texts from purely digital sources than in blogs, the other online-only medium. Thus, while partly digital sources are less likely to use a Writer-oriented rhetoric than blogs, Inclusive We is

¹⁵ In a boxplot such as Figure 3 and Figure 4, the box marks where the middle 50% of observations fall, the horizontal whiskers the lowest 25% and the highest 25%, excluding outliers. The median is marked by a black vertical line.

favoured. This choice of rhetoric could reflect the sense of closeness between writers and readers.

An analysis of source dimension scores reiterates these findings. Blogs tend to score below or just above the mean, even for blog that favour a writer-oriented style. A fair number of partly digital sources do use Inclusive We, but the low median is caused by a few select sites that seem to lack or have exceptionally low frequencies of Inclusive We in all of their texts. This low use of the feature could imply stylistic preferences or guidelines, which would make the usage site-specific, a choice made by the writers or the editors to adopt a cohesive rhetorical stance towards the reader. The variation between sources is particularly noticeable among purely digital sources, where some use Inclusive We frequently and others stay closer to the distanced style favoured by partly digital sources.

Intertextuality is used to refer to authorities, but also to show awareness of a wider context. According to Hyland and Tse (2004:171) in academic writing Attributors are used for persuading and justifying arguments, but also for establishing credibility and displaying familiarity with the discipline. These aims are similar to that of the Intertextual Dimension 4, as outlined in Section 5.1.4: Attributors in opinion texts are used to source statements that support the opinion, but also to show awareness of the existing community and familiarity with the latest developments of the news story. Comparing texts that rank high in Dimension 4 (Intertextuality), one can notice that texts ranking high in this dimension resemble each others in terms of topic: they all report on statements made by an outside source. For instance:

- (32) “The paper *says* the mystery surrounding the deal that convinced Nissan to stay put in Sunderland [...] was evidence of a ‘drift’. It *calls* on Theresa May and the government to ‘settle what its broad aims are for Brexit’ and *says* that the current approach of offering assurances to individual companies will form a pattern” (article388)

Example (32) is a summary of the viewpoint of another newspaper, a sort of metareporting, that introduces perspectives whose original sources reader might not visit themselves. Its high score in Intertextuality is thus expected. Here the word choice is fairly neutral, but by changing the verb used (e.g. *claim*, *insist*, *show*), authors can also evaluate the source text, show their personal stance and indicate where the reader should stand.

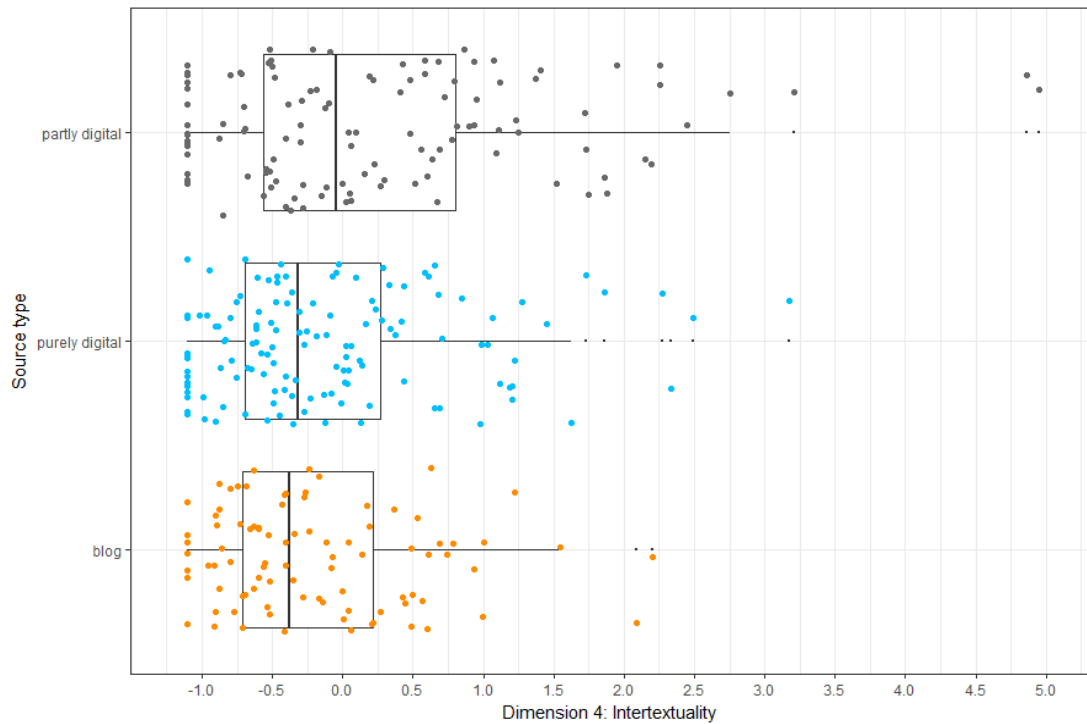


Figure 5: Texts in corpus plotted on Dimension 4

Many texts in the corpus have no Attributors and therefore fall low in Intertextuality, although the medians here are considerably higher than for Dimension 2. When comparing Intertextuality across source types, partly digital sources have the highest median and also the highest individual dimension scores, shown as outliers in Figure 5. Meanwhile, the medians of blogs and purely digital sources are quite even. These findings are confirmed by the source dimension scores. Looking at the texts above the medians, blogs do tend to have less Intertextuality than texts from news sites. This may reflect the purpose of news sites: investigating sources is a journalistic priority even when the text in question is a column or an op-ed. Blogs on the other hand are more concerned with personal opinion or prefer to refer to sources in other, non-linguistic ways (e.g. hyperlinks). However, in blogs the dispersion is more even and fewer blogposts lack Intertextuality altogether. This may reflect the authors of news sites being a more heterogenous group than individual bloggers. Meanwhile, partly digital sources score even higher than purely digital sources. This may reflect the modes of partly digital sources as compared to purely digital sources: physical prints of news must make intertextuality explicit and cannot use hyperlinks to source materials. Another possible explanation is that start-up like purely digital news sites are less concerned with reporting news than webpages of “old media”. However, this seems less likely as purely digital news sites, too, are

committed to journalistic professionalism (Deuze, 2017). In blogs, linking is used to refer to more information, to provide evidence for a claim, or even to add humour (Myers, 2010:38-45)

If a text is low in Dimensions 1, 2 and 3 it is not Writer-oriented, not Reader Oriented, and not even Solidarity seeking. To borrow the terminology of Puschmann (2013), such a text would be Topic-centric. The function of a text using Topic-centric style is generally to “inform others, indicate stance or opinion to others, gain recognition, acquire expert status” as opposed to relate and reflect on one’s own thought process (Puschmann, 2013:101). From the metadiscourse point of view, Topic-centric style is characterised by a lower frequency of personal pronouns as well as higher frequency of quotes and intertextuality and reference to more distant or generic participants. The depersonalised register bears similarities to academic writing, another genre where Attributors are noticeably frequent compared to other metadiscourse markers. For example, Hyland and Tse (2004:170) show that doctoral dissertations contain 7.62 Attributors (named Evidentials in their framework) per 1,000 words, making it the third most frequent metadiscourse type just after Transitional markers and Hedges. In comparison, Self-mentions score 4.02 and Attitude markers 1.85. This makes the the genre generally Topic-centric.

In the current corpus, Topic-centric style is best represented by the partly digital group, where Intertextuality is high, but references to reader or writer are less common. As an exemplar of a Topic-centric text, article335 scores -4.54 in Dimension 1 for a few Transitional markers and -1.79 in Dimension 3 for an Organisation marker, that is, it only uses select interactive metadiscourse markers. It has no uses of Inclusive We or even of Intertextuality. The text concerns the costly maintenance of Route 99 in California, yet neither the writer nor the reader is projected as Californian through possible phrases such as *us Californians* or *my Californian readers*. In fact, the reader is not addressed in any way. Consider for instance the following, all metadiscourse marker categories italicised:

- (33) “Ensuring that California’s freeways were all six lanes, well-lit and safe would have been a gargantuan but practical task that *could* have been completed long ago and *would* have saved thousands of lives (*though it would have required the admission that the mundane modern automobile was here to stay*). *Instead*, cool bureaucrats and hip politicians preferred to blow money on visions of grandiose space-age rail.” (article335)

The authorial voice and stance is detached. The hedged opinions can be understood to belong to the author especially in the cases where Asides are used as commentary (as in the example above, marked by parentheses), but this connection is not made explicit metadiscoursally. Certainly the topic is thought provoking, but the linguistically explicit purpose is to inform the reader of the issue, but readers are not called to action or to engage in discussion. Thus, much like academic writing, Topic-centric opinion writing utilise chiefly interactive metadiscourse, although with opinion writing this, too, may be limited as the texts are rarely so long or complex as to require further signposting. However, as shown by the variation explored in this section, all source genres mix rhetorical styles and a fully Topic-centric style is unlikely to be found in this corpus.

5.4 Summary of findings

The dimensions listed in the results of this empirical study are based on statistical grounds rather than perceived connections. While qualitative analyses on individual texts rely to some extent on intuition, the interpretation of the dimensions is based on previous literature on metadiscourse.

Four functionally meaningful dimensions were extracted from the corpus consisting of blogposts and columns and op-eds from news sites. While the broad, overarching communicative purpose of the texts is expression of opinion, the underlying variable of the dimensions can be interpreted as metadiscoursal and rhetorical strategies for approaching the reader. The first dimension relates to conveying and organising the writer's personal opinions and ideas using Attitude markers, Self-mentions, Hedges and Transitional markers. The second, consisting only of one metadiscourse marker category, aims to create solidarity between the writer and the reader by use of Inclusive We. In contrast to the first dimension, the third dimension is directed towards engaging and guiding the reader through Reader Address, Questions and Textual organisers. The fourth dimension consists of Attributors, which mark intertextuality. While not a rhetorical strategy per se, the category is important in persuasion as reference to authorities and as proof of awareness of societal context.

The purpose and the positioning of the audience was noted to play part in the choice of rhetoric, as the three first dimensions relate to the rhetoric of the writer/reader-relationship. Unsurprisingly, Writer-oriented text is used when the

author relates private opinion or experience, and the reader is merely an observer. However, Hedging is in a central position in Writer-oriented texts. On the other hand, Reader-oriented texts using the pronoun *you* can be conversational, dialogic and perhaps advisory in tone, but also persuasive towards a supposedly sceptic reader, bringing the issue closer to them, the reader becoming an active listener and a participant of the discourse event. Where the reader could be anticipated as less disagreeing, a rhetoric of Solidarity is often employed: the writer and reader are equals, face the same problems and should act as united. Both the Reader-oriented and the Solidarity based strategy are more explicitly persuasive or prescriptive in that they make mention of the reader and possibly of the reader's assumed pre-existing opinion.

While the opinion texts are similar in communicative purpose, some of the variance across texts can be linked to whether the text is from a blog, a purely digital news site, or the website of a print-newspaper. As expected, blogs were found to generally opt for a writer-oriented strategy more than opinion articles from news sites. This would indicate that at least on the surface they are more focused on relating the authors' own stories than changing the mind of the reader. It is possible that the author believes the readers read the blog specifically because they share its values – such readers would of course need less explicit persuasion. Both blogs and online news use reader-oriented strategies and Inclusive We to some extent, but blogs use less Inclusive We. Purely digital news sites were found to have the strongest preference for these strategies. As purely digital sources use less reader-oriented rhetoric, this could suggest that the authors, as representatives of online newspapers, are less likely to mention themselves unless it is under a uniting *we* that can refer to a nation or humanity as a whole. Some news sites score low in both the writer-oriented and the reader-oriented dimension, indicating that both the writer's and the reader's presence remain unacknowledged. Texts of this kind can be characterised as Topic-centric texts, using a term used by Puschmann (2013). As the texts have low use of metadiscourse categories, it can be hypothesised that either these texts are more reporting than persuasive or, as more likely, given how the texts are column and op-eds, their persuasion is implicit, relying on implicit devices such as word choice. However, based on these results it remains unclear why texts from purely digital sources would be more explicit in their persuasion.

The fourth dimension is used to source of intertextual content, but also to evaluate the reliability of the source and thus guide the interpretation of the reader. However, the evaluation of a source as reliable or unreliable depends on implicit linguistic factors such as word choice, which strictly speaking is not a focus of metadiscourse and thus not considered in this study. Intertextual references are most common in partly digital sources, that is, websites of newspapers with printed editions, which thus are the most likely to take on a Topic-centric style. Blogs exhibit variation but generally use less intertextual references. The differences between the comparable sub-genres are noteworthy, but the dispersion within the sub-genres must be borne in mind. Furthermore, it should be noted that even individual sites show some variation in the dimension scores, even when the site content is written by a single author, as is the case with blogs. One possible factor causing the variation is the topic of the text, another is slightly differing persuasive communicative purposes (e.g. reach a compromise with the readers vs challenge the readers).

Although there is variation in the use of metadiscoursal marker categories, most texts cluster around or slightly below 0, the mean dimension score. In general, only a small number of texts have high scores with the majority remaining lower on the continuum. In this study, prototypicality has been thought of as the distance to the mean scores. This accounts for variation within the whole corpus and assumes that the outlying texts are equally part of the genre. Admittedly, the method presented here is merely an experimental suggestion that to my knowledge has not been used in previous studies. It could be argued that the “prototypical text” would be better described without the outliers, that is, by focusing on the clusters that were observed in each dimension. A member of the denser cluster could provide a better picture of the prototypical opinion text than a measure influenced by the scores of the outliers. For example, the median of Dimension 2 was noticeably lower than the mean. However, without the outliers, the structure of the dimensions would look very different and describe a different dataset altogether. Furthermore, in a larger sample, it is likely that the outliers pattern out differently. Including the outliers in the definition and formula of prototypicality is thus essential for the prototypicality score to reflect the sample at hand.

The corpus here is intended as representative of the whole population, the sources having been chosen without author bias and the texts representing whole articles by various authors. What affects the results significantly is the fact that many

metadiscourse marker types had to be merged into broader categories. All dimension scores being based on normal distribution ensures that broad categories do not skew the scores unduly. The merged categories themselves are relatively coherent, and the annotation was made to be consistent even when a perfect recall could not be achieved. However, the merging does not allow a more fine-grained analysis of differences between sub-categories. For instance, Sequencers and Discourse goal markers are both analysed as Textual organisers, but have slightly different functions. However, the merging was done in accordance with the established theoretical framework, and if it were not for the merging a MD analysis could not have yielded dimensions of statistical significance in this limited corpus. For practical reasons, the size of the corpus was not expanded. This is hopefully something that can be explored in future studies.

6 Conclusion and Future Considerations

In this paper, the focus has been in the dimensions extracted, and individual metadiscourse marker categories, especially ones that do not load on any dimension, have been largely ignored, apart from when they are the only category loading in a dimension. Thus, it does not delve into a very detailed analysis of the texts, and the interpretations are meant as examples of the range of possibilities from which an author can choose. However, this is essentially the goal of a MD analysis, to see the bigger patterns, the forest for the trees. On one hand, the results of the present study have been fruitful in gauging the extent of the variation in the genre; on the other hand, this study has raised some questions that could be picked up in future studies.

Overall, the findings are promising given that a MD analysis on a pragmatic linguistic phenomenon in a small specialised corpus could just as well have yielded no variation whatsoever. Based on the relative success of this experimental study, the MD analysis could be adopted as a more common method in the field. They are useful in gauging the range of variation of a pragmatic function even across a seemingly homogenous genre: this study could thus be replicated using a larger specialised corpus. The statistically significant findings indicate that variation in metadiscourse exists and is primarily exemplified by the choice of rhetoric in creating a writer/reader-relationship. By testing the results of this study, future studies could further specify the functions of different metadiscourse markers and

further develop the metadiscourse framework to suit analyses in opinion discourse. For example, in this study Self-mentions load on a different dimension than Inclusive We, indicating that they represent separate underlying functions of a text. As such, future analyses on metadiscourse could benefit from making greater distinction between exclusive and inclusive *we*. The distinction is promoted by Ädel (2006), but not acknowledged (at least explicitly) by Dafouz-Milne (2003, 2008) or Hyland (2005). On a larger scale, the division into interactive and interactional metadiscourse often still holds in frameworks. The interpretation of correlating metadiscourse markers as set strategies could reveal new framework division models, opening fruitful avenues for future metadiscourse research.

As most texts in the corpus used here resemble each other, the next step could go on to analyse metadiscourse and rhetoric in even more types of opinion texts, including texts by less experienced writers or texts aimed for more specific audiences, say, the readers of a personal blog and or comments in web forum discussion. Because of their shorter length, such texts may not contain many interactive metadiscourse markers, in which case it would be interesting to conduct a similar research using especially interactional metadiscourse and frameworks such as stance, appraisal, or evaluation, which do not limit the interactional aspect to non-propositional and functionally explicit content. Admittedly, in corpus studies on pragmatics the challenge lies in annotating large amounts of data manually, or at least checking the reliability of automatic annotation, an effortful task when a pragmatic function is not limited to one form.

Some generalisations of metadiscourse use in a text can be made based on the source type of the text. An interesting focus of future studies would be opinion texts beyond the text, as especially texts online rely increasingly on multimodality. In general, much has changed since the early days of online journalism. In this study the data was limited to texts from blogs, purely digital news sites and websites of printed newspapers. In the 2010's the consumption of news has been ever transforming, news outlets becoming ever more personalised. As noted by Deuze (2017:11), online audiences tend to find their news on social media rather than on news websites. To meet the expectations and beat competitors, journalism must prioritise human interest and conduct affective storytelling: it is “emotional authenticity” that interests the reader (Beckett & Deuze, 2016:4). At the same time, interactivity and social media presence are ever increasing in importance: stories spread by “trending” on Twitter

and news on Facebook feeds are condensed into videos that must grab the attention of the social media users. Blogging in the 2010's and beyond is also facing the challenges of networking, professionalization and commercialisation. Visualisation is increasingly important: according to Pinjamaa and Cheshire (2016:10), blog readers assess that blogs will increasingly have to rely on visual media rather than text. This calls for future studies in writer-reader interaction to include the multimodal aspect of online journalism. De Groot et al. (2016) have already proposed a way of how to extend the metadiscourse framework by Hyland and Tse (2004) to include the multimodality framework by Kress and Van Leeuwen (2006) in order to study visual metadiscourse in corporate documents. They propose their extended model could be used to analyse other professional genres. However, visual metadiscourse could also be found in less professional and less conventionalised genres, and future research could expand to analyse all metadiscourse in persuasive or journalistic genres, possibly including the realm of social media presence including micro-blogging platforms. Yet another possible future endeavour could be to expand the corpus to cover also spoken data, e.g. online podcasts and vlogs (truncated from "iPod broadcast" and "video logs" respectively). The comment section of an article could be studied in comparison to the article itself.

Future studies should also expand to consider other variables than what have been studied so far. Dafouz Milne's (2008) studies on metadiscourse have compared British editorials with Spanish ones. Blogs using English as a lingua franca (ELF) have been considered from a grammatical point of view for example by Vettorel (2014). A study on metadiscourse in the cross-cultural and multilingual world of ELF could provide a closer look at differences between cultural variation in the rhetorical style of bloggers worldwide. In the present study, the opportunity to find connections between metadiscourse practises and authors' cultural and linguistic backgrounds was not possible due to most of the data sources being based in the United States. Factors such as variety of English as well as gender, age, education, or political ideology were ignored due to the limited dataset and the focus and scope of the study. The ever-evolving modes of computer mediated communication will always be a fruitful and topical focus for the field of metadiscourse in online opinion writing.

References

- Ädel, A. (2006). *Metadiscourse in L1 and L2 English*. Amsterdam: John Benjamins Publishing Co.
- Ädel, A. & Mauranen, A. (2010). Metadiscourse: Diverse and divided perspectives. *Nordic Journal of English Studies*, 9(2), 1–11.
- Alonso Belmonte, I. (2009). Positioning the reader: A study on the use of interactive textual patterns in English written newspaper editorials and articles of opinion. *English Text Construction*, 2(1), 48–69.
- Beckett, C. & Deuze, M. (2016). "On the Role of Emotion in the Future of Journalism", *Social Media + Society*, 2(3), 1–6.
- Bednarek, M. (2008). *Emotion talk across corpora*. Houndmills, Basingstoke, Hampshire: Palgrave Macmillan.
- Biber, D. (1988). *Variation across speech and writing*. Cambridge: Cambridge University Press.
- Biber, D. (1994). An analytical framework for register studies. In D. Biber, & E. Finegan (Eds.), *Sociolinguistic perspectives on register* (pp. 31–56). New York: Oxford University Press.
- Biber, D. (1995a). *Dimensions of register variation: A cross-linguistic comparison*. Cambridge: Cambridge University Press.
- Biber, D. (1995b). On the role of computational, statistical. and interpretive techniques in a Multidimensional analysis of register variation: A reply to Watson. *Text*, 15(3), 341–370.
- Biber, D. 2009, Multi-dimensional approaches. In A. Lüdeling & M. Kytö (Eds.), *Corpus Linguistics: an international handbook*. Walter de Gruyter, Berlin, pp. 822–854.
- Biber, D, Johansson, S., Leech, G., Conrad, S., & Finegan E. (1999). *Longman Grammar of Spoken and Written English*. London: Pearson Education Limited
- Biber, D., & Conrad, S. (2009). *Register, genre, and style*. Cambridge: Cambridge University Press.
- Biber, D., Conrad, S., & Reppen, R. (1998). *Corpus linguistics: Investigating language structure and use*. Cambridge: Cambridge University Press.
- Boggel, S. (2009). *Metadiscourse in Middle English and Early Modern English religious texts: A corpus based study*. Frankfurt am Main: Peter Lang.

- Campbell, V., Gibson, R., Gunter, B., & Touri, M. (2009). News blogs, mainstream news and news agendas. In S. Tunney, & G. Monaghan (Eds.), *Web journalism: A new form of citizenship* (pp. 29–45). Eastbourne: Sussex Academic Press.
- Crismore, A., & Farnsworth, R. (1989). Mr. Darwin and his readers: Exploring interpersonal metadiscourse as a dimension of ethos. *Rhetoric Review*, 8(1), 91–112.
- Crismore, A., Markkanen, R., & Steffensen, M. S. (1993). Metadiscourse in persuasive writing: A study of texts written by American and Finnish university students. *Written communication*, 10(1), 39–71.
- Dafouz-Milne, E. (2003). Metadiscourse revisited: A contrastive study of persuasive writing through professional discourse. *Estudios Ingleses De La Universidad Complutense*, 11, 29–52.
- Dafouz-Milne, E. (2008). The pragmatic role of textual and interpersonal metadiscourse markers in the construction and attainment of persuasion: A cross-linguistic study of newspaper discourse. *Journal of Pragmatics*, 40(1), 95–113.
- De Groot, E., Nickerson, C., Korzilius, H. & Gerritsen, M. (2016). Picture This: Developing a model for the analysis of visual metadiscourse. *Journal of Business and Technical Communication*, 30(2), 165–201.
- Deuze, M. (2003). The web and its journalism: Considering the consequences of different types of newsmedia online. *New Media & Society*, 5(2), 203–230.
- Deuze, M. (2017). Considering a possible future for Digital Journalism. *Revista Mediterránea de Comunicación: Mediterranean Journal of Communication*, 8(1), 9–18.
- Ferguson, C. A. (1994). Dialect, register, and genre: Working assumptions about conventionalization. In D. Biber, & E. Finegan (Eds.), *Sociolinguistic perspectives on register* (pp. 15–30). New York: Oxford University Press.
- Field, A. (2005). *Discovering statistics using SPSS: and sex, drugs and rock'n roll*. (2nd ed.). London: Sage.
- Fu, X., & Hyland, K. (2014). Interaction in two journalistic genres: A study of interactional metadiscourse. *English Text Construction*, 7(1), 122–144.
- Gray, B., & Biber, D. (2012). Current conceptions of stance. In K. Hyland, & C. Sancho Guinda (Eds.), *Stance and voice in written academic genres* (pp. 15–33). Houndmills, Basingstoke, Hampshire: Palgrave Macmillan.
- Greenberg, J. (2000). Opinion discourse and Canadian newspapers: The case of the Chinese "boat people". *Canadian Journal of Communication*, 25(4)
- Haas, T. (2005). From "Public journalism" to the "Public's journalism"? Rhetoric and reality in the discourse on weblogs. *Journalism Studies*, 6(3), 387–396.
- Halliday, M. A. K., & Hasan, R. (1976). *Cohesion in English*. London: Longman.

- Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *Halliday's introduction to functional grammar* (4th edition ed.). London: Routledge.
- Hermida, A. (2009). The blogging BBC: Journalism blogs at "the world's most trusted news organisation". *Journalism Practice*, 3(3), 268-284.
- Hyland, K. (1998). Persuasion and context: The pragmatics of academic metadiscourse. *Journal of Pragmatics*, 30(4), 437-455.
- Hyland, K. (1999). Talking to students: Metadiscourse in introductory course books. *English for Specific Purposes*, 18(1), 3-26.
- Hyland, K. (2005). *Metadiscourse: Exploring interaction in writing*. London: Continuum.
- Hyland, K., & Tse, P. (2004). Metadiscourse in academic writing: A reappraisal. *Applied Linguistics*, 25(2), 156-177.
- Ifantidou, E. (2005). The semantics and pragmatics of metadiscourse. *Journal of Pragmatics*, 37(9), 1325-1353.
- Jacobson, R. (1985). Metalanguage as a linguistic problem. In: *Selected Writings*, VII. S. Rudy (Ed.). (pp. 113-121). Berlin: Mouton Publishers.
- Khabbazi-Oskouei, L. (2013). Propositional or non-propositional, that is the question: A new approach to analyzing 'interpersonal metadiscourse' in editorials. *Journal of Pragmatics*, 47(1), 93-107.
- Kress, G., & Van Leeuwen, T. (2006). *Reading images: The grammar of visual design*. London, England: Routledge.
- Le, E. (2004). Active participation within written argumentation: Metadiscourse and editorialist's authority. *Journal of Pragmatics*, 36(4), 687-714.
- Pew Research Center (2016). *State of the News Media 2016*. Available at: <http://assets.pewresearch.org/wp-content/uploads/sites/13/2016/06/30143308/state-of-the-news-media-report-2016-final.pdf> [Accessed 2017-09-29]
- Martin, J. R., & White, P. R. R. (2005). *The language of evaluation*. Houndmills, Hampshire: Palgrave Macmillan.
- Matheson, D. (2004). Weblogs and the epistemology of the news: Some trends in online journalism. *New Media & Society*, 6(4), 443-468.
- Mauranen, A. (1993). *Cultural Differences in Academic Rhetoric: A Textlinguistic Study*. Frankfurt am Main: Peter Lang.
- Myers, G. (2010). *Discourse of blogs and wikis*. London: Continuum.
- O'Donnell, M. 2008. "The UAM CorpusTool: Software for corpus annotation and exploration". Proceedings of the XXVI Congreso de AESLA, Almeria, Spain, 3-5 April 2008.

- Pett, M. A., Lackey, N. R. & Sullivan, J. J. (2003). *Making sense of factor analysis*. London: SAGE Publications Ltd
- Pinjamaa, N. & Cheshire C. (2016). "Blogs in a changing social media environment: perspectives on the future of blogging in Scandinavia." In: *Twenty-Fourth European Conference on Information Systems (ECIS), Istanbul, Turkey*. pp. 1–16
- Precht, K. (2000). *Patterns of stance in English*, PhD dissertation, Northern Arizona University.
- Puschmann, C. (2013). Pragmatics of Computer-Mediated Communication. In S. Herring, D. Stein & T. Virtanen (Eds.). *Pragmatics of Computer-Mediated Communication* (pp. 83–108). Berlin: De Gruyter Mouton.
- Reese, S. D., Rutigliano, L., Hyun, K., & Jeong, J. (2007). Mapping the blogosphere professional and citizen-based media in the global news arena. *Journalism*, 8(3), 235–261.
- Rivers, W. L., McIntyre, B. & Work, A. (1988). *Writing Opinion: Editorials*. Ames, IW: Iowa State University Press.
- Sardinha, T. & Veirano Pinto, M. (Eds.) (2014). *Multi-Dimensional Analysis, 25 years on: A tribute to Douglas Biber*. Amsterdam: John Benjamins.
- Schiffrin, D. (1980). Meta-talk: Organizational and evaluative brackets in discourse. *Sociological Inquiry*, 50(3–4), 199–236.
- Tardy, C. M. (2012). Current conceptions of voice. In K. Hyland, & C. Sancho Guinda (Eds.), *Stance and voice in academic written genres* (pp. 34–48). Houndmills, Basingstoke, Hampshire: Palgrave Macmillan.
- Thompson, G. & Hunston, S. (2000). "Evaluation: An Introduction". In S. Hunston, G. Thompson (Eds.), *Evaluation in text: Authorial stance and the construction of discourse* (pp. 1–27). Oxford: Oxford University Press.
- Thompson, G., & Thetela, P. (1995). The sound of one hand clapping: The management of interaction in written discourse. *Text-Interdisciplinary Journal for the Study of Discourse*, 15(1), 103–128.
- Tognini-Bonelli, E. (2001). *Corpus linguistics at work*. Amsterdam: J. Benjamins.
- Vande Kopple, W. J. (1985). Some exploratory discourse on metadiscourse. *College Composition and Communication*, 36(1), 82–93.
- Vettorel, P. (2014). English as a Lingua Franca in Wider Networking: Blogging Practices. Berlin: De Gruyter Mouton.
- Wall, M. (2005). 'Blogs of war': Weblogs as news. *Journalism*, 6(2), 153–172.
- Zhang, M. (2016). A multidimensional analysis of metadiscourse markers across written registers. *Discourse Studies*, 18(2), 204–222.

Appendix 1: Data sources of corpus

Below are listed the websites whose texts were used for the corpus in this study. Some sites state permission to use their texts (sans image, video or other multimedia) for non-commercial research purposes. The remaining the author contacted personally by e-mail and received the respective copyright holders' permission to use the texts as material and quoted examples in this study. My sincerest thanks to the copyright holders are in order as this study would not have been possible without their generosity.

| Source Type | Source name | Url (Accessed 15.8.2017) |
|----------------------|---------------------------|---|
| blog (institutional) | Cato Institute Blog | cato.org/blog |
| blog | Ericmargolis.com | ericmargolis.com |
| blog | A Different Washington | geraldlday.blogspot.com |
| blog | Holistic Politics | holisticpolitics.org/ |
| blog | Iain Dale | iaindale.com |
| blog | The Mahablog | mahablog.com |
| blog | The Green Ribbon | tomgriffin.org |
| blog (institutional) | Worldwatch Institute Blog | blogs.worldwatch.org |
| purely digital | Barbados Underground | barbadosunderground.wordpress.com/category/blogging |
| purely digital | Blogger News Network | bloggernews.net |
| purely digital | Cagle Cartoons | caglecartoons.com |
| purely digital | Conservative Home | conservativehome.com |
| purely digital | Countercurrents | countercurrents.org |
| purely digital | Daily Kos | dailykos.com |
| purely digital | Liberal Democratic Voice | libdemvoice.org/category/op-eds |
| purely digital | Outside the Beltway | outsidethebeltway.com |
| purely digital | Salon | salon.com |
| purely digital | Tribune Content Agency | tribunecontentagency.com/premium-content/opinion |
| purely digital | Truthout | truth-out.org/speakout & truth-out.org/opinion |
| purely digital | World Net Daily | wnd.com/category/opinion/ |
| partly digital | The American Prospect | prospect.org/blog/tapped |
| partly digital | Financial Times | ft.com/comment |
| partly digital | The Hightower Lowdown | hightowerlowdown.org |
| partly digital | The Hindu | thehindu.com/opinion/columns |
| partly digital | New Statesman | newstatesman.com/the-staggers |
| partly digital | Reason.com | Reason.com/blog |
| partly digital | The Spectator | blogs.spectator.co.uk |
| partly digital | Washington Examiner | washingtonexaminer.com |

Below is a list of the texts from which extracts have been cited as examples in this study:

- Article014: *Barbados Underground*, "[The demise of West Indian Cricket?](#)" Author: David King. Published: 2016-09-24
- Article017: *Blogger News Network*, "[2016 – An Annus Horribilis like no other in modern times](#)". Author: Paddy Briggs. Published: 2016-09-28
- Article019: *Blogger News Network*, "[Saint Anthony's rises from the ashes!](#)" Author: Hugh McNichol. Published: 2016-09-14
- Article021: *Blogger News Network*, "[Mobile SIMs Hacks Cause Concern](#)". Author: Robert Siciliano. Published: 2016-09-20
- Article025: *The Mahablog*, "[Hillary's Night?](#)". Author: Barbara O'Brien. Published: 2016-09-27.
- Article143: *The Green Ribbon*, "[Tottenham in the nuclear balance of power](#)". Author: Tom Griffin. Published 2016-01-13
- Article149: *Truthout*, "[Solitary = Torture](#)". Author: Alan Mills. Published: 2016-10-13
- Article150: *Truthout*, "[Death Penalty for Heroin Dealers? More Proof the Drug War Is Not Over](#)". Author: Maya Schenwar. Published: 2016-10-12
- Article151: *Truthout*, "[The Time for Direct Action on Climate Change Is Now](#)". Author: Emily Johnston. Published: 2016-10-11
- Article166: *Daily Kos*, "[Supreme Court vacancy watch Day 243: Let's talk about the rule of law, Republicans](#)". Author: Joan McCarter. Published 2016-10-14
- Article174: *The Hindu*, "[For an emotional connect with readers](#)". Author: A. S. Panneerselvam. Published: 2016-10-10
- Article178: *The Hindu*, "[What does sport mean to us?](#)". Author: Nirmal Shekar. Published 2016-10-05
- Article180: *The Hindu*, "[Some problems are genuine, but BCCI's brinkmanship could end badly](#)". Author: Suresh Menon. Published: 2016-10-05
- Article183: *The Hindu*, "[Seeds of Discontent?](#)". Author: Aniket Aga. Published: 2016-09-30
- Article204: *Washington Examiner*, "[A pro-life woman's Election Day dilemma](#)". Author: Kate Bryan. Published: 2016-10-20
- Article218: *Countercurrents*, "[The Many 'Truths' On Syria: How Our Rivalry Has Destroyed A Country](#)". Author: Ramzy Baroud. Published: 2016-10-20.
- Article240: *Cato Institute Blog*, "[Turkey Attacks Anti-ISIS Forces in Syria](#)". Author: Ted Galen Carpenter. Published: 2016-10-20
- Article335: *Tribune Content Agency*, "[Lessons from the Highway of Death](#)". Author: Victor Davis Hanson. Published: 2016-10-27
- Article342: *Conservative Home*, "[How to increase the disposable income of those who are just about managing](#)". Author: James Frayne. Published: 2016-10-25
- Article348: *Liberal Democratic Voice*, "[Ten questions about the Richmond Park by-election](#)". Author: Mary Reid. Published: 2016-10-31
- Article360: *Holistic Politics*, "[Alternative Energy for Fun and Profit](#)". Author: Carl S. Milsted, Jr. Published: 2016-08-03

Article361: *Holistic Politics*, “[Should We Replace the Income Tax?](#)”. Author: Carl S. Milsted, Jr. Published: 2016-06-08

Article362: *Holistic Politics*, “[A Flat Tax for the 99%](#)”. Author: Carl S. Milsted, Jr. Published: 2016-05-24

Article383: *The Spectator*, “[Britain doesn’t need to bluff about Brexit](#)”. Author: David Green. Published: 2016-10-31

Article388: *The Spectator*, “[What the papers say: Should Carney stay?](#)”. Author: Tim Goodenough. Published: 2016-10-31

Appendix 2: Frequencies and variation of metadiscourse markers

Below are listed the mean, the minimum and maximum and the standard deviation of the normalized frequencies (per 1000 words) of metadiscourse marker categories.

These mean and standard deviations were used to calculate factor/dimension scores for the texts.

| | MEAN | MIN | MAX | SD |
|---------------------------|--------------|-------------|---------------|--------------|
| Attitude Markers | 1.94 | 0.00 | 10.56 | 2.10 |
| Attributors | 5.54 | 0.00 | 30.82 | 5.03 |
| Boosters | 3.35 | 0.00 | 17.38 | 2.86 |
| Code Gloss | 1.87 | 0.00 | 9.99 | 2.03 |
| Commentary | 2.24 | 0.00 | 18.07 | 2.46 |
| Deontic Modals | 1.78 | 0.00 | 10.71 | 2.10 |
| Hedges | 6.12 | 0.00 | 28.83 | 4.65 |
| Inclusive We | 4.24 | 0.00 | 58.90 | 7.23 |
| Questions | 2.00 | 0.00 | 17.86 | 2.77 |
| Reader Address | 4.09 | 0.00 | 44.64 | 5.86 |
| Self-mention | 4.69 | 0.00 | 51.81 | 7.86 |
| Textual Organisers | 1.92 | 0.00 | 19.05 | 2.75 |
| Transitional Markers | 10.94 | 0.00 | 33.25 | 5.36 |
| Sum of all markers | 51.48 | 5.41 | 118.12 | 20.40 |

Below are listed the standard deviation, maximum, minimum and range of dimension scores.

| | SD | MAX | MIN | RANGE |
|-----------------------------------|------|-------|-------|-------|
| Dimension 1: Writer-oriented text | 2,62 | 9,48 | -4,88 | 14,36 |
| Dimension 2: Solidarity | 1,00 | 7,56 | -0,59 | 8,15 |
| Dimension 3: Dialogic text | 2,11 | 11,97 | -2,12 | 14,09 |
| Dimension 4: Sourcing | 1,00 | 5,02 | -1,10 | 6,12 |

Appendix 3: List of potential metadiscourse markers

As stated in Section 4.2, the lists of adjectives, adverbs and verbs were automatically extracted from the corpus using the *UAM CorpusTool*, and the potential metadiscourse markers were chosen from this list. The list of other examples was compiled through familiarising with the corpus and with previous research on metadiscourse, such as Hyland's (2005) Appendix on metadiscourse items. Below are lists of words and expressions counted as potential markers. Their use as metadiscourse has been analysed in context, syntactic or pragmatic. Not all potential markers turned out to truly have a metadiscoursal function, and not all occurrences of a marker can be considered metadiscourse as defined by the criteria of this study. Questions and Commentaries were found in the corpus by searching for the relevant punctuation ("?" and "—" respectively). The annotation of Directives, too, is based on syntactical criteria (the imperative), not on word lists. These metadiscourse marker categories are therefore not included in the list below.

Attitudinal adverbs

| | | |
|-----------------|----------------|---------------|
| admirably | fantastically | justly |
| affectionately | fervently | luckily |
| amazingly | fiercely | magically |
| appropriately | fortunately | mentally |
| aptly | frankly | nicely |
| astonishingly | fulsomely | obnoxiously |
| awfully | grievously | oddly |
| backward | handily | ominously |
| bizarrely | happily | ostensibly |
| blatantly | hilariously | painfully |
| blindly | honestly | passionately |
| boldly | hopefully | peculiarly |
| brilliantly | horribly | plainly |
| cleverly | hugely | poignantly |
| competently | ideally | pointedly |
| comprehensively | importantly | popularly |
| conveniently | incorrectly | powerfully |
| courageously | indirectly | predictably |
| critically | inexplicably | profoundly |
| crucially | insufficiently | rightfully |
| curiously | interestingly | sadly |
| disturbingly | intolerantly | sheepishly |
| dramatically | invariably | shockingly |
| drastically | ironically | skilfully |
| exclusively | irrationally | spectacularly |
| | jealously | splendidly |

staunchly
stunningly
surprisingly
suspiciously
uncomfortably
unfairly
unfortunately
unhappily
uninterestingly
unreasonably
unsurprisingly
wildly

Attitudinal cognitive verbs

acknowledge
assume
believe
be convinced
consider
imagine
recognise
remember
want

Attitudinal verbs

accept
agree
appalled
appreciate
confess
disagree
encourage
enjoy
fear
feel
hate
laughed
love
pray
prefer
thank
wonder
worry

Attributors

according to
accuse
acknowledge
add

admit
advise
announce
answer
argue
ask
boast
brag
call
cite
claim
comment
complain
conclude
criticize
declare
discuss
document
emphasise
explain
find
indicate
insist
issue
maintain
mention
note
observe
plead
pledge
provide
question
reads
remind
reply
report
request
respond
reveal
say
shout
speak
state
suggest
sum
tell
tweet
warn
wonder
vows

write

Boosters, adjectives

certain
clear
definite
evident
indisputable
sure
undeniable

Boosters, adverbials

certainly
clearly
definitely
definitively
entirely
exactly
inevitably
in fact
in reality
merely
obviously
of course
no doubt
surely
undoubtedly
unquestionably
visibly

Boosters, necessity modals

must
shall
should

Boosters, noun expressions

evidence
fact
proof
reality

Boosters, reporting verbs

affirm
be convinced
confirm
demonstrate
evidence

know
promise
prove
show
see

Code glosses

namely
e.g.
for example
for instance
i.e.
in other words
known as
namely
put another way
say,
such as
that is to say
that/this means

Conversational devices

blah
bloody
damn
fricking
golly
hell
hey
huh
me-ow
oh
sheesh
sigh
um
ugh
well,
whee
wow
yes

Deontic modals

have to
must
need to
ought to
should

Hedges, Approximators

about
approximately
around
barely
comparatively
fairly
generally
largely
likely
mainly
maybe
mostly
nearly
often
partly
pretty
quite
rather
relatively
roughly
slightly
somehow
somewhat
usually
broadly
widely

Hedges, Epistemic adjectives

likely
probable
unlikely
possible
unclear

Hedges, Probability adverbials

allegedly
almost
arguably
apparently
likely
perhaps
possibly
potentially
presumably
probably
reportedly

supposedly

Hedging verbs

appear
describe
feel
guess
point out
postulate
seem
suppose
suspect
tend to
think

Reader-addresses

you, your, yours,
yourself,
yourselves...
reader, readers

Self-mentions

I, me, my, mine,
myself...
we, us, our, ours,
ourselves...

Transitional markers

actually
also
although
and
apart
as well
aside
because
besides
but
consequently
despite
else
even so
even though
further
furthermore
given
hence
however
in spite of
instead

likewise
 moreover
 nevertheless
 nonetheless
 not only
 notwithstanding
 or
 otherwise
 similarly
 since
 still, ...
 subsequently
 thereby
 therefore
 though
 thus
 too
 while
 yet

Textual organisers

1), 2), 3) ...
 a), b), c) ...
 above
 after all
 as far as
 as for
 as to
 accordingly
 additionally
 admittedly

again
 anyway
 back to
 below
 conclusion
 either way
 example
 figure
 finally
 first
 firstly
 following
 fourth
 here
 hereby
 graph
 I add/argue/explain/
 indicate/maintain/note
 /remind/speak/suggest
 /want/write
 in other words
 in short
 in summary
 in the end
 it's worth citing/
 mentioning/noting/
 pointing/recalling
 let me
 let's continue/
 consider/look/
 switch/revisit

meantime
 meanwhile
 my goal
 next
 note
 now
 on one hand
 on the other hand
 overall
 previously
 regarding
 returning to
 second
 secondly
 section
 So,
 table
 there
 third
 thirdly
 turning to
 then
 under
 underneath
 with regard to