Reference and Identity in Public Discourses

EDITED BY

Ursula Lutzky
Minna Nevala

John Benjamins Publishing Company

Reference and Identity in Public Discourses

Pragmatics & Beyond New Series (P&BNS)

ISSN 0922-842X

Pragmatics & Beyond New Series is a continuation of Pragmatics & Beyond and its Companion Series. The New Series offers a selection of high quality work covering the full richness of Pragmatics as an interdisciplinary field, within language sciences.

For an overview of all books published in this series, please see
http://benjamins.com/catalog/pbns

Volume 306

Reference and Identity in Public Discourses
Edited by Ursula Lutzky and Minna Nevala

Reference and Identity in Public Discourses

Edited by

Ursula Lutzky
Vienna University of Economics and Business

Minna Nevala
Tampere University

John Benjamins Publishing Company
Amsterdam / Philadelphia

The paper used in this publication meets the minimum requirements of the American National Standard for Information Sciences – Permanence of Paper for Printed Library Materials, ANSI Z39.48-1984.

DOI 10.1075/pbns.306

Cataloging-in-Publication Data available from Library of Congress:
LCCN 2019029452 (PRINT) / 2019029453 (E-BOOK)

ISBN 978 90 272 0420 2 (HB)
ISBN 978 90 272 6205 9 (E-BOOK)

John Benjamins Publishing Company · https://benjamins.com

Table of contents

Pragmatic explorations of reference and identity in public discourses

Minna Nevala and Ursula Lutzky
Tampere University / Vienna University of Economics and Business

This introductory chapter sets the scene for the contributions to this volume which are all focused on the analysis of reference and identity in public discourses. In this chapter, we provide an initial overview of previous research findings on reference and identity, and thus offer a foundation on which the following chapters will build. The focus is on the study of identity that is constructed or represented in public discourses through reference markers or the discussion of linguistic and extralinguistic defining features. It is the purpose of this introduction to shed light on these discursive mechanisms behind historical and present-day identity construction, as well as to introduce the studies themselves.

Keywords: identity construction, reference markers, public discourses

1. Introduction

The focus of this book is on constructing and expressing identities in various public fora that are of interest in both historical and present-day discourse studies and (socio)pragmatics. From a historical perspective, there is surprisingly little research that explicitly thematises or theorises the notion of identity and its relationship with language use. In research on present-day language, this gap is less wide. Identity and reference, and their influence on language use and communication, have been of interest in present-day language studies for the past decade or so, but investigations focusing on public discourse are still rather scarce.

The volume has been produced in international collaboration and is mainly based on the papers given as part of a panel at the International Pragmatics Association Conference held in Belfast, UK, in July 2017. The focus then was on three partly overlapping areas: (a) communication patterns in public discourse, (b) reference and identity markers, and (c) micro- and macrostructures in text and discourse. These three themes are also addressed in the chapters of the current

https://doi.org/10.1075/pbns.306.int

volume. The group of contributors includes several internationally well-known scholars. In addition to nine authors who presented their work at the conference panel, we invited four further contributors in order to offer a wider perspective on the topic, reaching from historical to multilingual concerns.

The chapters in this book study various aspects of identity which can be defined broadly as the positioning of the self and other (Bucholtz and Hall 2005: 586). In research, identity construction is generally viewed as an interactive process where others affect the way we see ourselves. Linguistic identity construction has become an active area of research where identity is understood as a complex and many-sided phenomenon built in interaction and dependent on time and space (see e.g. Bucholtz and Hall 2005; De Fina et al. 2006; Auer 2007). It has mostly been studied in present-day language and discourse, supported by the use of various modern ethnographic tools and methods, and the focus has often been on the individual language user, emphasising their active linguistic choices and agency.

In addition to identity, the chapters of this volume discuss the concept of reference. While people's perception of themselves is influenced by other participants and how they see and treat them in interaction (Ochs 1993: 298), the use of reference markers is influenced by speakers' view of other participants in a communicative situation. That is to say that, reference markers are chosen on the basis of who the addressee is and what the addresser knows about and understands in respect to the referent (Schegloff 1996: 458–459). Consequently, they can be used to express more than mere reference (cf. Schegloff's reference *simpliciter*, 2007) and may, for instance, be used when engaging in other actions, such as describing. "When we *describe* a person, we commit to selecting some features and not others as constituting 'the description'" (Stivers et al. 2007: 4), and the same applies to the description of different social and societal groups and communities.

Furthermore, the studies in this book focus on the analysis of public discourses. Historically, it was not until the late modern period that the distinction between private and public seems to have emerged with the rise of the public sphere in major European countries. As for communication, the growth of the public domain has frequently been related to the major technological developments, such as the invention of print, and to the gradual rise in the accessibility of printed material throughout the sixteenth century. However, the concept of public discourses that the contributions to this volume rely on has become increasingly elusive, especially in the digital era. In characterising the new media and the rise of computer-mediated communication (CMC), researchers have underlined the blurred distinction between newsmakers and consumers, as well as between what is considered public and what is considered private, between what is mass and what is interpersonal communication (Landert 2014; Ratia et al. 2017).

This volume spans a variety of contexts: it approaches the importance of reference and identity markers in both synchronic and diachronic data, including studies from as early as the early modern English period, and investigates their contribution to the construction of a potentially evaluative stance. It studies texts that are or were (intended to be) public, as reflected in their accessibility or distribution in the public domain, and draws on a range of different text types, including letters, newspapers, parliamentary debates, and online communication in the form of reader comments, discussion pages or fora. Finally, this volume does not exclusively focus on a single language area but combines chapters that study in addition to English also Spanish and Polish data, including, for example, political tweets about Poland and recordings of presidential interviews in Ecuador.

It is the aim of this book to uncover new insights into discourse specific reference and labelling patterns, both from a historical and present-day perspective, and to discuss their impact on self- and other-representation in the construction and representation of identity in public texts. While the diachronic dimension of this volume allows for trends to be observed in the development of certain reference markers, the synchronic approach facilitates the immediate application of findings to enhance our understanding of reference practices in more or less institutionalised contexts.

2. Identity

Identity has been described in more recent research as something that is created and performed rather than innate and possessed (see e.g. Joseph 2009: 14). Particularly, Bucholtz and Hall's (2005) seminal work on dynamic identity construction integrates different approaches within sociocultural linguistics to achieve an understanding of a complex, multifaceted concept of identity. According to their theory, an identity has multiple layers, which range from situational positions in conversations to more fixed demographic categories. Identity is thus seen as something that is contextually constructed through acts and stances and is constantly changing.

A similar view was already introduced in the discourse analytic study of multifaceted identity by Schiffrin (1996). The equally multidirectional framework created by Bucholtz and Hall comprises the actual discourse, the ideological processes behind it, the linguistic structures implementing it, as well as the local cultural practices and social groupings behind it (Bucholtz and Hall 2005: 607). Gordon (2013) describes this general perspective as "currently the dominant paradigm in the analysis of identity in interaction". In a more recent study on online identity, Georgalou (2017) also adopts a constructionist approach, in which identity is seen as a constantly shifting phenomenon being produced and reproduced through language use every day. In other words, identity is not something one *is* but rather something one *does* via discourse.

Personal identity implements our individual characters or attitudes, whereas social identity is related to groups that we do or do not belong to or identify with. Georgalou notes that there is an in-built need in us all to represent ourselves to others: our identities are "mediated, constituted and reconstituted, via the varied discursive practices in which we participate" (2017: 12; see also De Cock, Kluge this volume). The theory of identity construction is thus built on various premises, according to which identity is the fusion of social identity and personality, as well as a form of socially meaningful practice. It is constructed in discourse and defined as a contextual and interactive task. It involves a semiotic process of representation and is thus a multifaceted and complex entity.

Group identity,[1] or social identity, is "defined by individual identification with a group: a process constituted firstly by a reflexive knowledge of group membership, and secondly by emotional attachment or specific disposition to this belonging" (Benwell and Stokoe 2006: 25). The actual group behaviour (what people actually think and do as members of a group) is shaped by more macro-level dimensions of social identity processes (Hogg 2005: 208–209). This also involves such processes as building and maintaining different kinds of stereotypes and beliefs. Tajfel (1982) sees stereotypes not only as descriptive but also as functional, in that when we hold a certain stereotype of a group, thinking, for example, that it is morally depraved, the stereotype serves to justify and legitimise our actions. In the social identity process, we tend to exaggerate differences among groups as well as similarities within groups.

Therefore, it is not unusual for a group to define its identity by its common opposition to some enemy or 'out-group'. While this process can be very effective in strengthening in-group bonds, it does so by significantly intensifying intergroup conflicts. Intragroup consensus can be reached by conforming to group norms. This process is called "referent informational influence", which occurs in three stages: self-categorisation (in which a person defines a social category or identity for themself), norm formation (in which a person creates or learns the stereotypical norms of the social category), and norm representation (in which a person assigns the norms to him/herself and starts behaving accordingly) (Hogg and Abrams 1988: 172). In other words, we are influenced by others to the extent that they are in a position to be knowledgeable about group beliefs, norms and

1. Kleinke et al. (2018: 12) make a distinction between 'group identity', which can be seen as "highly dynamic and negotiable", and 'collective identity', which comprises more stable and long-term characteristics.

values. That is particularly true of individuals who are most representative (prototypical) of the in-group. They guide and lead discussions about 'who we are' and hence 'what we should do'.

As already mentioned, social identity may be constructed through dissociation from other groups. Social identities "imply both the perception of features shared with fellows in an ingroup and the accentuation of perceived differences with other groups' members" (Grad and Martín Rojo 2008: 12). The most typical way of manifesting social identities and intergroup relations is in the so-called in-group and out-group discourse. Often the majority culture comes to be seen, and talked about, as the norm, the 'us', and that of the minority group as the 'other'. Wodak (2011: 61), in her study of discursive exclusion and inclusion strategies, notes that the meaning of who *we* are varies according to prevailing ideologies and power relations: sometimes *we* means 'all of us reasonable people' and, at other times, it denotes clearly defined and restricted groups. Therefore, the use of *us* and *them* characterises linguistic contexts including, for instance, "many debates in all European member states (such as polarized discussions on EU enlargement, gender mainstreaming, on definitions of citizenship, immigration and participation in decision-making, and many more)" (Wodak 2011: 75).

Van Dijk (2009: 52) discusses this state of polarised public discourse in terms of "our own place". Such places are where we want to be autonomous socially, politically, and culturally: there we do not want interference from above, or from outside our own place, i.e. group, including invasion into our way of using language (i.e. our sociolect, idiolect, dialect, language). As Van Dijk (2009: 141) states, giving attributes to the self and others concerns interactional and societal contexts. This means that defining the self is not only governed by macro-level norms or shared knowledge, but also by micro-level interactions and situations (see e.g. Kluge, Marcus and Evans, Nevala, Thielemann, Tyrkkö this volume).

A specific field of identity studies where the 'us vs them' discussion prevails is the construction of national identity. According to Verdugo and Milne (2016: 3), there are three ways of understanding national identity: (a) seeing it as a fixed, cultural characteristic which is based on birth and ancestry, (b) describing it as an invented or imagined entity, or (c) characterising it as a civic identity, i.e. a sense of belonging to a geopolitical entity. Following Wodak et al. (1999), studies in this volume see national identity as "constructed and conveyed in discourse, predominantly in narratives of national culture. National identity is thus the product of discourse." (Anderson 1983: 22; see e.g. Dossena, Kopf, Thielemann this volume). Nations and national identities are created in language use and are therefore also socially constructed; they are established and maintained in daily social life (Billig 1995: 8). In McCrone and Bechhofer's (2015: 15) view, people "do" national identity within certain contexts. In places like Scotland and England, national identity can have such a strong cultural, rather than political, foothold that it changes only a little in time (2015: 20).

Triandafyllidou (2002: 26–27) sees national identity as "double-edged". On the one hand, it is based on a set of common features that bind the members of the nation together. On the other hand, national identity implies difference, since it involves awareness of 'others' from whom the nation seeks to differentiate itself. In this respect it does not differ from the notion of social identity, which, as we already saw, involves polarised in-group and out-group discourse. According to Triandafyllidou, this means that national identity has no meaning *per se*, but it only becomes meaningful in contrast to other nations. Collective identities, such as a national or an ethnic one, thus structure speakers' perceptions of themselves and others as members of different groups (see also Kleinke et al. 2018: 4–5 for collective identities and political discourse). "In a world organised into nations and national or multinational states, ethnicity or nationality are crucial aspects of identification and social categorisation of an individual, with wide consequences for the individual's life chances" (Triandafyllidou 2002: 28).

3. Reference

Reference is an important means of identity construction. In its basic sense, reference is "a kind of verbal 'pointing to' or 'picking out' of a certain object or individual that one wishes to say something about" (Carlson 2004: 76). The contributions to this volume mainly focus on the concept of 'person reference' (but see, for instance, Kopf this volume) which pertains to the way in which speakers refer to themselves (self-reference) or others (other-reference), such as the addressee or third parties who may be present or absent from an interaction.

Self-reference is closely tied to speakers' construction of their own identity. In this context, first-person personal pronouns play a prominent role as they allow a speaker to either refer to themselves as an individual (e.g. *I, me, my*) or as part of a group (e.g. *we, us, our*). However, especially with the advance of social media, new forms of self-reference have developed. Coesemans and De Cock (2017) show, for example, that Twitter users have further constructions at their disposal, such as Twitter handles (@username) and hashtags (#name), and may even engage in self-reference in the third person by retweeting a message that another user tweeted about them. They explore the use of these medium-specific forms of self-reference in political tweets and find that Twitter is appropriated as "a tool for professional political communication, but also a tool for personal branding" (Coesemans and De Cock 2017: 37).

When considering the traditional linguistic means which may be used to express reference, three main categories emerge: nominal forms (e.g. *Jim, Ms Miller*), pronouns (e.g. *we, they*), as well as other definite or indexical constructions (e.g. *this woman, the tall guy*). The group of nominal forms of reference comprises next to proper nouns also titles (e.g. *Professor*), kinship (e.g. *mum, dad*), as well as occupational terms (e.g. *doctor, reverend*). While these forms may be used when talking about someone, they may equally function as address terms and thus index the interlocutor that a speaker is addressing. As terms of address are defined as "the linguistic forms speakers use to refer to their collocutor(s)" (Braun 1998; see also Mazzon 2010 for historical research), they may be regarded as forming part of the category of reference terms. This is also illustrated by contributions to the current volume (see e.g. Kehoe and Gee this volume) which show that especially in an online context the boundary between address and reference may be blurred.

In addition to their "basic function of establishing linguistic reference to the interlocutor" (Braun 1998), terms of address serve further purposes, such as getting an interlocutor's attention or singling them out from a group of potential addressees. At the same time, they may have an expressive function, conveying the speaker's mood or their evaluation of the addressee, and they may contribute to defining their relationship. Thus, the use of more formal terms of address may indicate a more distant relationship, whereas informal ones may signal closeness and possibly intimacy. Given the often asynchronous and place-independent nature of online media, the way in which forms of address and reference are used to "establish and maintain relationships through [Computer-Mediated Communication] CMC" (De Oliveira 2013: 307) is of particular relevance. Parts II and III of this volume offer further research into the affordances of platforms such as Twitter (see Jaworska and Sogomonian, Thielemann this volume), YouTube (see Kluge this volume), and Wikipedia (see Kopf this volume).

The second category of reference expressions includes pronouns. As Cramer (2010: 624) notes, "[p]ronouns carry a great deal of power in their use", especially as they can be employed "in strategic ways for the purposes of presenting a particular position or gaining a particular effect". For instance, the first-person singular pronoun *I* may be used by a speaker to express self-reference and, in fact, it constitutes the default pronoun choice to this end (individual self-reference). However, a speaker may also use the first-person plural pronoun *we* (collective self-reference), which allows them to index a social group and thereby establish "solidarity as well as group identity" (Fetzer and Bull 2008: 276). At the same time, different shades of meaning can be distinguished in the use of *we*: while it may be used in an inclusive way, comprising both the speaker and the addressee, it may also appear with an exclusive function, referring to the speaker and a third party but excluding the addressee (see e.g. Mühlhäusler and Harré 1990; Duszak

2002). As Pavlidou (2014: 11) points out, the pronoun *we* distinguishes not only between different degrees of plurality, but also between different types of collectivity. Given the manipulative possibilities that pronouns thus offer, their use has been studied extensively in a political context where politicians can mould messages through their choice of pronouns so as to reflect their underlying beliefs and ideologies (see Fetzer and Bull 2008; Jaworska and Sogomonian this volume; Tyrkkö 2016 and this volume).

The third group of reference expressions comprises other definite or indexical constructions, whose referents may shift depending on their deictic position in a specific situational context. They may convey spatial relations, express the speaker's attitude, and give insight into their social identification or potential bias towards a referent (see e.g. Nevala 2012). This is illustrated by Stivers' (2007: 85–87) example in which she compares the use of *that next door neighbour* to default expressions such as using the neighbour's name or a construction such as *our next door neighbour*. The demonstrative *that* in this case signals an instance of negative labelling and thereby conveys a negative impression of the referent 'neighbour'. Moreover, when collocating with a term of reference, *that* may even preface or form part of the speech act of complaining, appearing in utterances which mean to criticise a third party. Referring expressions thus "do more than just refer to a person" in so far as speakers can fit "the precise referring expression to the type of pragmatic action that they are otherwise performing" (Stivers 2007: 73; see also Kehoe and Gee this volume).

All of these different means of person reference may be used to foster in-group or out-group discourse, with individuals identifying as part of a specific in-group which stands in opposition to the out-group. While in-group terms are seen as referring to the norm and usually attract more positive connotations, out-group terms pertain to the 'other' and are usually more negatively connotated. As we already stated in Section 2, pronouns have been studied extensively in this respect, especially with a view to creating an *us* versus *them* discourse and thereby facilitating inclusion or exclusion respectively (see also Wodak 2011). With reference to their use in political discourse, Tyrkkö (2016) mentions that "one of the primary techniques in political rhetoric is the setting up of contrasts between *us* and *them*, or between the in-group and the out-group", a task often accomplished by pronouns (see also Jaworska and Sogomonian this volume). Consequently, person reference also contributes to self-presentation as it may reveal aspects of the speaker's group membership (Braun 1998; see also Dossena this volume). In fact, as Kleinke and Bös (2018: 154) find in their analysis of indeterminate *us* and *them* in public online discussion, it may not only be involved in collectively constructing identities but may even reflect the creation of cultural models. However, person reference has also been distinguished from membership categorisation (see in particular Schegloff 2007). Terms of membership categorisation mainly have a describing or identifying function by conveying what a specific person is 'about'. These may include, for instance, terms pertaining to professional roles, such as *nurse* or *teacher*, which convey certain features relating to the activities people engage in as part of their jobs. While they may also be used for reference purposes, terms of membership categorisation are not necessarily equivalent to reference expressions depending on their context of use.

At the same time, the social, situational or professional role of an interlocutor may influence the referential expressions used. That is to say that the context and the participant's role within that context affect the choice of reference terms. They reflect socio-cultural conventions as well as personal characteristics and reveal insights into variables such as power, authority or distance as they mirror the interpersonal relationship between interlocutors (see e.g. Marcus and Evans this volume; Nevala 2009 and this volume). Shifts in context and corresponding changes in an interlocutor's role can affect the appropriateness of referential expressions. Terms of reference therefore serve a variety of pragmatic functions in that they "contain information about speaker and addressee, about their positions in society and vis-à-vis each other, about their moods, intentions, their actual and desired relationship" (Braun 1998) and they thus open a window on the complex intricacies of language in use.

4. Studies

The contributions to Part I of this volume discuss public discourses in the history of English. In the first chapter of this part, **Minna Nevala** studies evaluative person reference in newspaper articles on a late nineteenth-century murder case. Focusing on the Mannings murder, which involved Frederick and Maria Manning hiding a body under their kitchen floor, she investigates how the couple was represented in the Victorian press. This study is based on a data set comprising 1,158 news reports on the Mannings murder case which were published over a period of three and a half months from the murder on 17 August to a fortnight after the couple's execution on 13 November 1849. The analysis, which focuses on both explicit and implicit person reference, shows that the construction of the Mannings' identity was influenced by several factors. While they were generally represented in negative terms, the time of publication, specific events, such as a suicide attempt or a confession to the priest, and whether they were referred to individually or as a couple influenced the way they were described in the news reports. Thus, Nevala's chapter underlines that the construction of the social identity of criminals in nineteenth-century England was a complex phenomenon that was subject to situational variation and could at times even involve positive evaluation through the expression of compassion or pity.

Marina Dossena's chapter continues further into the theme of group homogeneity, but in respect to national identity. Her study investigates the Jacobite cause and its historical narrations in popular culture. The interest lies in such labels as 'Bonnie Prince Charlie', 'Redcoats' and 'Sassenachs', which are familiar to the modern public from, for example, nineteenth-century songs and twenty-first century fiction and TV drama. Such popular, appealing labels may be, however, downright manipulative in their presentation of biased views. Dossena finds that these narrations can be anachronistic and may affect the reliability of the text. She warns against ignoring the identity mechanisms on which labels are based and reminds the users of such labels of their pragmatic strength.

Moving on to the negotiation of identity through more interpersonal means, **Imogen Marcus** and **Mel Evans** investigate how high-ranking early modern women letter-writers negotiated, firstly, their subordinate position as women, and, secondly, their positions of power. They study four central Tudor women, Kathryn Parr, Mary I, Elizabeth I and Elizabeth Talbot, focussing on how they express self-reference and, in particular, stance in their correspondence. The analysis aims at understanding how indexical meaning arises from the convergence of multiple linguistic forms in particular contexts. The results show that scribal status and recipient rank shape the writers' linguistic choices, and that gender is one of the most significant dimensions of the women's epistolary identities.

In the last chapter of the historical part, **Jukka Tyrkkö**'s study continues the discussion of societal attitudes to gender roles. He examines diachronic trends of kinship reference in British parliamentary debates from 1800 to 2005, using the Hansard corpus. The findings show that the representation of female family members has grown more frequent and nuanced over the years. The trends concerning pronouns and kinship terms suggest that the reference to family members has increased, which cannot be solely explained by the type of public discourse in question or its record-keeping practices.

Part II of this volume is dedicated to the study of public discourses in Present-Day English and the first contribution studies topic-based communities in the comment section of *The Guardian*. **Andrew Kehoe** and **Matt Gee** take a corpus linguistic approach in their analysis of 6.2 million reader comments made on over 100,000 articles on the website of this UK newspaper between 2007 and 2010. Their data mainly come from the blog section of *The Guardian*, Comment is Free, which attracts most reader comments, but also include comments on conventional news articles published in other sections, such as Sport or Law. Their research shows that there is a small group of very active users who comment a lot and mainly do so in specific sections of the newspaper, where sub-communities are formed around particular topics and shared interests, such as music or football. By addressing and referring to each other, these users develop relationships and construct their own online identities in interactions which sometimes span months or even years. In addition to providing new insights into the interactions evolving in online comment threads, Kehoe and Gee demonstrate the methodological possibilities that a large-scale, data-driven approach has to offer to the study of pragmatic and social phenomena.

Susanne Kopf studies terms of reference that are used to denote the European Union on Wikipedia. Rather than studying the outward-facing encyclopaedic entries that Wikipedia is known for, she focuses on their use on its talk pages. These inward-facing discussion pages record the process of collaboratively crafting the respective Wikipedia articles and can, as Kopf discusses, be regarded as a transnational public sphere that allows users to negotiate editorial decisions and facilitates opinion formation. In her analysis, she investigates the debates developing on the Wikipedia talk page that accompanies the article titled "European Union" between 2001 and 2015, with a view to uncovering the terms of reference that were put forward for the EU or discarded by the Wikipedia editors. Her corpus-assisted study shows that some terms (e.g. 'country') were more intensely debated than others and that the higher the degree of sovereignty ascribed to the EU and its constituents by a specific term, the better the chances for it to be supported by other users. Her findings reflect how editors discuss reference choices as well as represent and therefore make sense of the EU, opening up a further dimension onto the conceptualisation of this political institution.

In the third contribution to Part II, **Sylvia Jaworska** and **Tigran Sogomonian** study political bonding and imagined collectives on Twitter in the run up to the EU referendum in the United Kingdom in 2016. This chapter investigates the use of personal pronouns for identity construction in political communication given the affordances and constraints of Twitter, such as the 140-character limit demanding concise messages. The analysis is based on two corpora of 1,456 Leave and 1,458 Remain tweets, comprising influential tweets which were retweeted more than 100 times between 1 April and 22 July 2016 and which contained one of a set of relevant keywords, such as *Brexit* or *EU*. In addition to tweets by politicians and media institutions, tweets by members of parliament were included in the data. The results show that the first-person plural pronoun *we* was used to a different extent in Leave and Remain tweets and that it reflects a diverging range of identities and memberships evoked in the two data sets, which potentially influenced the size and nature of the audiences addressed. By studying its interplay with other features, such as @handles and hashtags, the authors show how the use of *we* contributed to creating a more or less involved or persuasive style and consequently to the political identities constructed.

The contributions to Part III of this volume discuss public discourses around the world. **Barbara De Cock**'s chapter studies self- and other-representation in online discussions of diabetes. In particular, she investigates European Spanish data from three different types of public online interaction, a frequently asked questions page, a chat session, and a forum, which are all dedicated to providing information on diabetes and linked to official healthcare institutions in Spain. The participants asking questions and providing answers in her data include patients, their relatives as well as healthcare professionals. By studying both medical and emotional identity features, De Cock uncovers differences in reference and representation across the three media investigated. Her analysis shows that the degree of personal information provided and of emotional features discussed depends on the more or less generalisable nature of the platform studied as well as on whether it is (a)synchronous in kind. At the same time, healthcare professionals are found to focus on medical identity features and less on taking a personalised approach. The data therefore reflect a more traditional patient-healthcare professional relationship where few shared identities are being constructed between the participants of an interaction, thus putting less emphasis on patients as co-responsible for the management of their chronic illness.

Bettina Kluge studies the different identity values associated with the Spanish diminutive *-ita* in a 2012 TV interview between journalist Ana Pastor and Ecuadorian president Rafael Correa. Although Correa later claimed that his use of the form *Anita* was not intended to show disrespect, the public interpreted it as machismo. Kluge discusses the incident from both sides, taking into consideration two other interviews between Correa and Pastor, as well as the subsequent online discussion in Spain, Ecuador and other countries. She finds that the use of the diminutive *Anita* instead of *Ana* did not seem to challenge Pastor's professional identity. This was mainly because both Correa and Pastor treated the incident rather as a situational power struggle than as a premeditated attack on Pastor's identity as a noteworthy journalist.

The last chapter of the book deals with political criticism through the construction of a fictitious national identity. **Nadine Thielemann** studies the use of the hashtag *#SanEscobar*, which originated as a slip of the tongue made by the former Polish Minister of Foreign Affairs Witold Waszczykowski. Thielemann focuses on how Twitter may function as an arena for the subversion of political discourse as she carries out a multi-modal discourse analysis of tweets constructing the virtual country of San Escobar. The online discourse studied involves elements and frames which make reference to South-America, Pablo Escobar and the national conservative PiS party in Poland. The results show that in the data PiS-governed Poland is strongly associated with a stereotypical South American narco-democracy (the fictitious San Escobar) and the discourse functions as a humorous means of criticising the current political situation in Poland.

As the chapter summaries have shown, the ten studies in this book are united in their interest in reference and its role in representing and interactively constructing identities in both past and present public discourses. They are based on diverse types of data and by asking new research questions, they offer significant empirical results that pertain next to different stages in the history of English (see Parts I and II of this volume) also to Spanish and Russian (see the contributions to Part III). The methodologies used are both (corpus-based) quantitative and qualitative in nature. In fact, the volume serves as an example of combining multiple methodological approaches, including socio-scientific and linguistic frames of analysis, which provides an opportunity for innovative findings. The contributions to this volume provide new baseline evidence for linguistic and pragmatic variation in both past and present public discourse and situate the concepts of reference and identity in diverse synchronic, diachronic and language-specific contexts, thus contributing further to our understanding of the discursive representation and construction of individuals and groups in the public sphere.

Acknowledgements

Minna Nevala's work on this volume is done as part of the project *Democratization, Mediatization and Language Practices in Britain, 1700–1950* (DEMLANG), funded by the Academy of Finland (project number 295381). We would like to thank Katharina Berger for her administrative support in editing this volume.

References

Anderson, Benedict. 1983. *Imagined Communities: Reflections on the Origin and Spread of Nationalism*. London: Verso.

Auer, Peter (ed). 2007. *Style and Social Identities: Alternative Approaches to Linguistic Heterogeneity*. Berlin: Mouton de Gruyter.

Benwell, Bethan, and Elizabeth Stokoe. 2006. *Discourse and Identity*. Edinburgh: Edinburgh University Press.

Billig, Michael. 1995. *Banal Nationalism*. London: Sage.

Braun, Friederike. 1998. "Terms of Address." In *Handbook of Pragmatics* Volume 4, ed. by Jef Verschueren, and Jan-Ola Östman. Amsterdam: John Benjamins. https://doi.org/10.1075/hop.4.ter1

Bucholtz, Mary, and Kira Hall. 2005. "Identity and Interaction: A Sociocultural Linguistic Approach." *Discourse Studies* 7 (4–5): 585–614. https://doi.org/10.1177/1461445605054407

Carlson, Gregory. 2004. "Reference." In *Handbook of Pragmatics*, ed. by Laurence Horn, and Gregory Ward, 74–96. Malden, MA: Blackwell.

Coesemans, Roel, and Barbara De Cock. 2017. "Self-Reference by Politicians on Twitter: Strategies to Adapt to 140 Characters." *Journal of Pragmatics* 116: 37–50. https://doi.org/10.1016/j.pragma.2016.12.005

Cramer, Jennifer. 2010. "'Do *We* Really Want to Be Like *Them*?': Indexing Europeanness through Pronominal Use." *Discourse & Society* 21 (6): 619–637. https://doi.org/10.1177/0957926510381217

De Fina, Anna, Deborah Schiffrin, and Michael Bamberg (eds). 2006. *Discourse and Identity*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511584459

De Oliveira, Sandi Michele. 2013. "Address in Computer-Mediated Communication." In *Pragmatics of Computer-Mediated Communication. Handbooks of Pragmatics* Volume 9, ed. by Susan Herring, Dieter Stein, and Tuija Virtanen, 291–313. Berlin: Walter de Gruyter. https://doi.org/10.1515/9783110214468.291

Duszak, Anna (ed). 2002. *Us and Others: Social Identities across Languages, Discourses and Cultures*. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.98

Fetzer, Anita, and Peter Bull. 2008. "'Well, I Answer It by Simply Inviting You to Look at the Evidence.' The Strategic Use of Pronouns in Political Interviews." *Journal of Language and Politics* 7 (2): 271–289. https://doi.org/10.1075/jlp.7.2.05fet

Georgalou, Mariza. 2017. *Discourse and Identity on Facebook: How We Use Language and Multimodal Texts to Present Identity Online*. London: Bloomsbury.

Gordon, Cynthia. 2013. "Analysis of Identity in Interaction." In *The Encyclopedia of Applied Linguistics*, ed. by Carol A. Chapelle. London: Blackwell.

Grad, Hector, and Luisa Martín Rojo. 2008. "Identities in Discourse: An Integrative View." In *Analysing Identities in Discourse*, ed. by Julía Todolí, and Rosana Dolón, 3–28. Amsterdam: John Benjamins. https://doi.org/10.1075/dapsac.28.03gra

Hogg, Michael A. 2005. "Uncertainty, Social Identity and Ideology." In *Social Identification in Groups*, ed. by Shane R. Thye, and Edward J. Lawler, 203–229. Amsterdam: Elsevier. https://doi.org/10.1016/S0882-6145(05)22008-8

Hogg, Michael A., and Dominic Abrams. 1988. *Social Identifications*. London: Routledge.

Joseph, John E. 2009. "Identity." In *Language and Identities*, ed. by Carmen Llamas, and Dominic Watt, 9–17. Edinburgh: Edinburgh University Press.

Kleinke, Sonja, and Birte Bös. 2018. "Indeterminate *Us* and *Them*. The Complexities of Referentiality, Identity and Group Construction in a Public Online Discussion." In *The Discursive Construction of Identities On- and Offline: Personal – Group – Collective*, ed. by Birte Bös, Sonja Kleinke, Sandra Mollin, and Nuria Hernández, 153–176. Amsterdam: John Benjamins.

Kleinke, Sonja, Nuria Hernández, and Birte Bös. 2018. "Introduction: Identity Construction in Complex Discourse Contexts." In *The Discursive Construction of Identities On- and Offline: Personal – Group – Collective*, ed. by Birte Bös, Sonja Kleinke, Sandra Mollin, and Nuria Hernández, 1–12. Amsterdam: John Benjamins.

Landert, Daniela. 2014. *Personalisation in Mass Media Communication. British Online News between Public and Private*. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.240

Mazzon, Gabriella. 2010. "Terms of Address." In *Historical Pragmatics. Handbooks of Pragmatics* Volume 8, ed. by Andreas H. Jucker, and Irma Taavitsainen, 351–376. Berlin: Walter de Gruyter.

McCrone, David, and Frank Bechhofer. 2015. *Understanding National Identity*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9781316178928

Mühlhäusler, Peter, and Rom Harré. 1990. *Pronouns and People: The Linguistic Construction of Social and Personal Identity*. Oxford: Blackwell.

Nevala, Minna. 2009. "Altering Distance and Defining Authority. Person Reference in Late Modern English." *Journal of Historical Pragmatics* 10 (2): 238–259. https://doi.org/10.1075/jhp.10.2.05nev

Nevala, Minna. 2012. "A Bit of *This* and a Bit of *That*: on Social Identification in Early and Late Modern English Letters." *English Language and Linguistics* 16 (2): 261–280. https://doi.org/10.1017/S1360674312000056

Ochs, Elinor. 1993. "Constructing Social Identity: A Language Socialization Perspective." *Research on Language and Social Interaction* 26, 287–306. https://doi.org/10.1207/s15327973rlsi2603_3

Pavlidou, Theodossia-Soula. 2014. "Constructing Collectivity with 'We': An Introduction." In *Constructing Collectivity: 'We' across Languages and Contexts*, ed. by Theodossia-Soula Pavlidou, 1–19. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.239.03pav

Ratia, Maura, Minna Palander-Collin, and Irma Taavitsainen. 2017. "English News Discourse from Newsbooks to New Media." In *Diachronic Developments in English News Discourse*, ed. by Minna Palander-Collin, Maura Ratia, and Irma Taavitsainen, 1–12. Amsterdam: John Benjamins. https://doi.org/10.1075/ahs.6.01rat

Schegloff, Emanuel A. 1996. "Some Practices for Referring to Persons in Talk-in-Interaction: A Partial Sketch of a Systematics." In *Studies in Anaphora*, ed. by Barbara A. Fox, 437–485. Amsterdam: John Benjamins. https://doi.org/10.1075/tsl.33.14sch

Schegloff, Emanuel A. 2007. "Categories in Action: Person-Reference and Membership Categorization." *Discourse Studies* 9 (4): 433–461. https://doi.org/10.1177/1461445607079162

Schiffrin, Deborah. 1996. "Narrative as Self-Portrait: Sociolinguistic Constructions of Identity." *Language in Society* 25 (2): 167–203. https://doi.org/10.1017/S0047404500020601

Stivers, Tanya. 2007. "Alternative Recognitionals in Person Reference." In *Person Reference in Interaction: Linguistic, Cultural, and Social Perspectives*, ed. by Nick J. Enfield, and Tanya Stivers, 73–96. Cambridge: Cambridge University Press.

Stivers, Tanya, Nick J. Enfield, and Stephen C. Levinson. 2007. "Person Reference in Interaction." In *Person Reference in Interaction: Linguistic, Cultural, and Social Perspectives*, ed. by Nick J. Enfield, and Tanya Stivers, 1–20. Cambridge: Cambridge University Press.

Tajfel, Henri (ed). 1982. *Social Identity and Intergroup Relations*. Cambridge: Cambridge University Press.

Triandafyllidou, Anna. 2002. *National Identity Reconsidered: Images of Self and Other in a 'United' Europe*. Lampeter: Edwin Mellen Press.

Tyrkkö, Jukka. 2016. "Looking for Rhetorical Thresholds: Pronoun Frequencies in Political Speeches." In *The Pragmatics and Stylistics of Identity Construction and Characterisation* (Studies in Variation, Contacts and Change in English 17), ed. by Minna Nevala, Ursula Lutzky, Gabriella Mazzon, and Carla Suhr. Helsinki: VARIENG. Available online at http://www.helsinki.fi/varieng/series/volumes/17/tyrkko/ (accessed 15 April 2019).

Van Dijk, Teun A. 2009. *Society and Discourse: How Social Contexts Influence Text and Talk*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511575273

Verdugo, Richard R., and Andrew Milne. 2016. *National Identity: Theory and Research*. Charlotte, NC: Information Age.

Wodak, Ruth. 2011. "'Us' and 'Them': Inclusion and Exclusion – Discrimination via Discourse." In *Identity, Belonging and Migration*, ed. by Gerard Delanty, Ruth Wodak, and Paul Jones, 54–77. Liverpool: Liverpool University Press. https://doi.org/10.5949/UPO9781846314537.004
Wodak, Ruth, Rudolf De Cillia, Martin Reisigl, and Karin Liebhart. 1999. *The Discursive Construction of National Identity*. Edinburgh: Edinburgh University Press.

PART I

Public discourses in the history of English

Two miserable creatures or *those atrocious criminals?*

Evaluative reference in the Mannings murder reporting

Minna Nevala
Tampere University

This chapter studies the way in which an infamous nineteenth-century couple, Frederick and Maria Manning, was described in the news of a murder case, and how by the use of person reference they were placed into a particular social group. The interest is in whether the social representation of murderers became the way in which the Mannings were described as individuals, and in whether the resulting social identities could be considered stable or variable. The study shows that although evaluative reference was mostly rather neutral in tone, the Mannings were in general referred to as negatively as other nineteenth-century murderers. However, there was also situational variation linked to momentary empathy, which made the couple's social representation less stable than expected.

Keywords: social representation, evaluative language, person reference, Manning, Victorian press

1. Introduction

> The horrors of the gibbet and of the crime which brought the wretched murderers to it faded in my mind before the atrocious bearing, looks, and language of the assembled spectators. […] When the two miserable creatures who attracted all this ghastly sight about them were turned quivering into the air, there was no more emotion, no more pity, no more thought that two immortal souls had gone to judgement, no more restraint in any of the previous obscenities, than if the name of Christ had never been heard in this world, and there were no belief among men but that they perished like the beasts. (*The Times*, 13/11/1849)

https://doi.org/10.1075/pbns.306.01nev

In August 1849, two policemen found a dead body under the floor of Maria and Frederick Manning's kitchen in Bermondsey. The excerpt above is from one of the newspaper articles published during the last stages of this famous murder case in late nineteenth-century Britain. In his *Letter to the Editor*, Charles Dickens writes about the execution of the murderous couple, particularly his disgust with the way people flocked in to see their hanging.

Dickens's style resembles greatly that of the news reportage on the Mannings' case in general. Its way of describing the murderers and the audience, as well as their behaviour and the surroundings, is typical of the Victorian popular press. Reporters generally used melodramatic fiction in presenting the darker sides of social life, including rudimentary psychological analyses and a fascination with shocking detail (Linnane 2003: 280; Conboy 2010: 87; O'Reilly 2014: 222). The newspapers' way of using a variety of forms already familiar from popular culture was also aimed at attracting new audiences (Conboy 2002: 86).

Newspaper language can often reveal hidden attitudes and stereotypes about certain groups in society, whether for or against. The actual content of group behaviour (what people actually think and do as members of a group) is shaped by more macro-level dimensions of social identity processes (Hogg 2005). In other words, groups often define their social identity by their common opposition to some enemy or 'out-group'. This opposition breeds positive and negative evaluations and social stereotypes, i.e. generalisations or assumptions about the characteristics of all members of a group (Hogg and Abrams 1988). Such social stereotypes are typically used for constructing the public identities of, for example, criminals and other members of the social margins.

It is the purpose of this chapter to study the way in which the Mannings were described in the news reports of their murder case, and how by the use of person reference they were placed into a particular, negatively evaluated, social group. The main interest lies in how the general social representation of that group of criminals soon became the way in which the Mannings were described as individuals, and in whether the resulting social identities could be considered stable or variable. The media often nurture the fear of the unknown and want to unite the public against a common enemy (Tabbert 2015: 152). From this perspective, it is interesting to find examples expressing compassion towards the Mannings in the data, but this momentary empathy is often caused by a particular event, like an injury or a confession to the priest. A preliminary analysis (Nevala 2016 and 2017) shows, for example, that the dichotomous representation of criminals and their victims was sometimes done rather randomly.

The main research questions in the current study thus concern (1) the way in which both of the Mannings are referred to in the news reports and (2) possible diachronic changes in the degree of evaluation shown by both the nominal person reference terms and/or their textual context (neutral-negative-positive). Furthermore, the focus is on (3) how the reference to the Mannings reflects the social representation of criminals that was prevalent in general at the time and whether the resulting social identities could be considered stable or variable during the course of the news reporting on the murder case.

In the next sections, I will first discuss the concepts of social representation and evaluative reference, as well as the history of the Mannings murder case. Then I will study the degree of evaluation in the data, and give examples of both explicit and implicit cases of person reference. Finally, I will draw conclusions on the diachronic trends emerging from the news reports, as well as discuss the role of positive and negative evaluation and labelling in building social representation for criminals in nineteenth-century Britain.

2. Social representation and evaluative reference

The notion of social representation has its origins in social psychology. It means those values, ideas, metaphors, beliefs, and practices that are shared by the members of groups and communities (cf. Moscovici 1984). In critical discourse analysis, moral evaluation is "linked to specific discourses of moral value" (Van Leeuwen 2008: 110). Representation of a particular individual or a group can manifest itself by the use of adjectives such as *healthy, normal*, or *natural*. This is what Van Leeuwen (2008: 109) calls "moral legitimization": it consists of the processes of evaluation, abstraction, and comparison. People are categorised on the basis of positive and negative values into different sociocultural groups.

Language can be used to reinforce or undermine categorical differences (Hogg and Abrams 1988: 212). The most typical way of manifesting public images, identities and intergroup relations is in the so-called in-group and out-group discourse. Wodak (2008: 61), in her study of discursive exclusion and inclusion strategies, notes that the meaning of who *we* are varies according to prevailing ideologies and power relations: sometimes *we* means 'all of us reasonable people' and, at other times, clearly defined and restricted groups. When describing *us* and *them*, we tend to use attributes. As Van Dijk (2009: 141) states, giving attributes to the self and others concerns interactional and societal contexts. This means that defining the self and others is not only governed by macro-level norms or shared knowledge, but is also produced in micro-level interactions and situations. Ochs (1993: 289) sees the process of building an identity, and thus also representation, as usually not explicitly encoded in language use, but rather as a social meaning inferred in act and stance meanings. It can be seen to evolve and vary in social interaction in response to the acts and stances of other interlocutors, but also according to the speaker's own attitude towards each interactional situation (Ochs 1993: 298).

Similar to Wodak and Ochs, Van Leeuwen (2008: 147) discusses various strategies that can be used either implicitly or explicitly to refer to people as 'others'. In public discourse, people can, for example, be depicted as inferior, deviant, criminal, or even evil, or shown to be members of homogeneous groups and, in this way, they are denied their individual characteristics and differences (i.e. "they are all the same"). Other exclusion tactics include the strategy of negative cultural connotations and the strategy of negative racial stereotyping.

Often the values and norms of a particular group are manifested in the reference practices used by other groups or members of those groups, who create and maintain – mostly – negative impressions, which can be aided or achieved through the use of "labels of primary potency" (Allport 1986). Consequently, certain characteristics, like male/female, white/black, or law-abiding/criminal, carry more perceptual potency than others, and signal difference from what is considered mainstream. In this study, for instance, characterising the Mannings couple as murderers might not only prevent characteristics such as a good spouse or an honourable citizen, but also emphasises their social representation at the time as deviant and criminal.

Neutral, positive, and negative evaluation then reflects a more prevalent, societal attitude either in favour or against particular group memberships, as e.g. in the case of criminals and law-abiding citizens. By creating and using negative terms and attributes, 'respectable' people place criminals in a 'detestable' outgroup. Studies have shown that there are, for example, specific naming strategies for criminals (Mayr and Machin 2012: 57). Clark (1992: 224) calls this process of extreme negative labelling "fiend naming": when criminals are referred to as monsters, they are depicted so evil and alien that they cannot fit into humankind and society. Negative labelling then becomes a strategy based on the notion that no 'normal' person would be capable of such a 'monstrous' crime. In comparison, victims are often labelled positively with what could be called 'angel naming', respectively (cf. Nevala 2016). Tabbert (2015: 152) has found that, in the press, victims are often described in terms of their social network and environment, as well as their character. For example, a victim whose identity is constructed as being cared for and loved by others evokes more empathy and concern, because readers think they must be a 'good person' (2015: 104). The more people are affected, the greater the impact of the crime.

Linguistic evaluation, therefore, carries an element of judgement, since when expressing, for example, our reluctance to be associated with a particular outgroup, we simultaneously assess the characteristics of that group and all its members as negative. According to Bednarek (2006; see also Bednarek and Caple 2012), the judgement of whether something or someone is good or bad falls under the core parameter of emotivity. It is concerned with the expression of a person's approval or disapproval, and it can manifest itself in language in either positive (e.g. a polished speech) or negative (e.g. a rant) descriptions. It also contains an element of force from low intensity (a crime) and medium intensity (an assassination) to high intensity (a slaughter). Bednarek (2006: 45–47) thus associates the scalarity of emotivity with that of objectivity and admits that emotive meaning is difficult to recognise and categorise in general. In Martin and White's (2005) appraisal theory, emotions both for and against are categorised under "attitude". "Attitude" is in turn divided into three semantic dimensions, "affect" (positive/negative feelings), "judgement" (positive/negative attitude to behaviour),[1] and "appreciation" (positive/negative evaluation of semiotic and natural phenomena).

The notion of judgement is most central to the present study, since when describing, for example, a murderer, we first and foremost evaluate that person based on their behaviour and actions. Judgement also concerns ethic evaluation, what is right and what is wrong, and as a dimension of social sanction, it covers praise (descriptions like *good, law-abiding, respectful*) and condemnation (e.g. *evil, corrupt, indecent*). The division between positive and negative evaluation clearly reflects a prevalent societal attitude of being either in favour or against particular group memberships, as e.g. in the case of crime and criminals.[2] Not all evaluation clearly functions on the good-bad, or the positive-negative, parameter, nor is it necessarily signalled by specific linguistic elements, but the distinction can still be found to exist in language use (Thompson and Hunston 2000: 23; Hunston 2007: 39–40). When linguistically visible, evaluative reference can manifest in actual labels and attributes, such as in personal pronouns and terms of reference. In addition, there is a variety of other discursive ways in which interlocutors' social representation is encoded. The juxtaposition of different groups and their members can occur, for example, through acts of criticism and praise or through variation in stance-taking (cf. Alba-Juez and Thompson 2014: 9–14).

1. Martin (2000: 155) describes judgment as "the institutionalization of feeling", meaning the expression of the norms of how people should or should not behave.

2. Martinez Caro (2014: 322) uses the term "value" when she talks about positive-negative evaluation (good-bad).

3. Crime news in nineteenth-century Britain: The Mannings murder in focus

Although the Victorians were fascinated by news about social and moral degradation, they were at the same time very conscious of the respectable vs unrespectable dichotomy. The growing middle class in Britain led the change in moral sensibility, forwarding and reinforcing the Victorian values in newspapers and other public texts (Linnane 2004: 176–177). In the latter half of the nineteenth century, the press was increasingly driven by popularisation, which meant that the production of news was directed towards a larger number of readers (Conboy 2010: 79). Daily papers were directed towards the middle classes, whereas the Sunday ones were mostly read by the working classes and they included more sensational and radical views. Whether or not the Sunday papers' using emotionally loaded language was a covert strategy to influence the reading audience with regard to the reporter's point of view remains uncertain. Highlighted subjectivity and focus on entertainment were, however, characteristics strongly associated with the popular press.

Both then and now, the medium of newspapers naturally affects the language used. Whereas men are, for example, usually portrayed as being the agents of their crimes, women are often seen as objects of male influence or coercion (Mayr and Machin 2012: 135). Present-day news features on crimes committed by women tend to focus on the incapability of women to commit a crime on their own, and although much of the nineteenth-century reporting on crime still worked on the evil vs good dichotomy (i.e. 'social attitude' or 'judgement'), the representations of female criminals particularly were also based on their type of behaviour and level of attractiveness.

The reporting of the Mannings murder case can be seen to deviate somewhat from the reporting 'norm' in this respect: Maria Manning was hardly ever portrayed as being an object of anything or anyone, least of all of her husband, who was increasingly described as being pitiful. Perhaps it was the fact that she was a foreigner, or her supposedly fickle heart and greedy nature, but she was clearly seen as agent to the crime. Both she and her husband also came from the lower classes, which qualified as a type of 'from deprived to delinquent' story. The press were obsessed with Mrs Manning and her proud demeanour, particularly since she was the perfect example of a Victorian murderess, the likes of whom often surfaced in fiction or drama at the time. This is not surprising: it has been estimated that the Mannings case sold 2.5 million copies of various broadsides (Diamond 2003: 165).

So, what actually happened? On August 17th 1849, two policemen found a dead body under the floor of Maria and Frederick Manning's kitchen in Bermondsey, London. The story leading to its deadly end starts when Maria de Roux came to London to serve as a lady's maid for Lady Blantyre. In 1846 she went across the

Channel to Boulogne with her employer and met Patrick O'Connor, a 50-year-old Irishman, who worked as a customs officer in the London docks. Maria was much taken with Mr O'Connor but she was also involved with Frederick Manning, who worked as a guard on the Great Western Railway. Both men proposed to Maria, and she had to decide which of the two would lead her to a more prosperous marriage. Eventually, Frederick was the one chosen, and the couple married in 1847. Maria soon felt she had married the wrong man and was determined to have O'Connor's money by killing him. She invited O'Connor to dinner and when he arrived, she shot him in the head with a pistol. The bullet wound did not kill him, however, and Frederick finished O'Connor off by battering his head with a ripping chisel. The two of them then buried the body below the kitchen flagstones, covering it with quicklime.

The following day Maria went to O'Connor's lodgings and managed to con her way into his rooms where she took everything of value, including his share certificates. Afraid of O'Connor's colleagues asking too many questions, the Mannings fled soon afterwards, both in separate directions. The couple was right in suspecting the colleagues, who then reported O'Connor missing. The police searched the Mannings' house and found the body of O'Connor under the flagstones. A hunt after the Mannings started, and eyewitnesses were interviewed.

It was the impunity of the Mannings that excited the popular press after they had been caught, Maria first in Edinburgh and Frederick later in Jersey (Diamond 2003: 160). Both of them were brought back to London, where they were tried at the Old Bailey on the 25th and 26th of October. Frederick blamed his wife for the murder, claiming first that he had fainted and did not know the body was buried under the kitchen floor. Eventually, he confessed to taking part in killing O'Connor, saying, "I never liked him very much" (Diamond 2003: 161). Both Mannings were charged with murder and remanded to Horsemonger Lane Gaol.

The couple's executions took place in the morning of the 13th of November, when, as Charles Dickens described in the introduction to this chapter, "there was no more emotion, no more pity". Even three years later, he described how he had seen "those two forms dangling on the top of the entrance gateway – the man's a limp loose suit of clothes, as if the man had gone out of them; the woman's a fine shape, so elaborately corseted and artfully dressed, that it was quite unchanged in its trim appearance as it swung slowly from side to side" (*Household Words*, 30/10/1852).[3] This account reflected the above-mentioned manner in which the representation of female criminals was often based on the contradiction between their (positive) attractiveness and their (negative) behaviour. It also

3. Dickens also based the character of Hortense in *Bleak House* (1852–1853) on Maria Manning.

echoed the general view of the popular press at the time: the proud Maria and the unimpressive Frederick, despised for trying to hide behind his wife.

4. The data and method

All the newspapers in the *19th Century British Library Newspapers Database* were searched with the terms 'Manning' and 'murder', in order to avoid hits for other people named Manning as far as possible. The time span was set between August 17th and November 30th, 1849, and the content was restricted to 'news'. All the resulting data were skimmed for false references and hits, as well as double or multiple hits from the same publication (i.e. same issue, same page, same article). The total number of news articles on the Manning murder case extracted from this database amounted to 1,108 articles. In addition to the *19th Century British Library Newspapers*, I searched *The Times Archive* for a small reference dataset of 50 news articles. This was done in order to check the evaluative scale of reference used for the Mannings in the popular press against that used in *The Times*, which was already then considered a more mainstream and traditional newspaper. Although no noticeable stylistic, register-specific differences could be identified between these two data sets in the end, the *The Times* news pieces were included in this study regardless. Altogether, the data for the current study therefore consist of 1,158 news articles.

In order to track the diachrony of the reporting on the Mannings case, the data were divided into three news cycles of different length according to the main events. The beginning of each cycle showed a peak in the number of news articles on the murder. The first cycle covers the period from August 17th to October 24th, i.e. from the date of the murder to the day before the first trial day. The second cycle spans from the trial dates of October 25th/26th to the day before the execution, November 12th. The third cycle reaches from November 13th to November 30th, which covers the execution day and about a fortnight after. After the end of November, the news about the case radically decreased and repeated contents very similar to previous reporting. The number of articles per cycle is given in Table 1.

Table 1. The number of articles in the three news cycles

Dataset	First cycle	Second cycle	Third cycle	Total
C19 BLN	649	259	200	**1,108**
The Times	29	12	9	**50**
Total	**678**	**271**	**209**	**1,158**

As the data were only available as images rather than text files, the resulting figures could not be normalised. Also for this reason, the close analysis of the reference used for the Mannings is mainly of a qualitative nature. As for the qualitative analysis, the news articles were read closely in order to confirm the objects of reference. The main referents and all person reference relating to them were studied along with their linguistic context in order to confirm the positive-negative evaluation. The quantitative part of the study concerns the instances of direct nominal reference terms and contextual evidence, where the emphasis is on the possible diachronic change from mainly negative to more positive or empathetic evaluation in the course of the three news cycles (for 'solidarity', see Nevala 2016 and 2017).

When deciding on the degree of evaluation (whether, for example, a term can be classified according to different degrees of neutrality, positivity, or negativity on the evaluative scale), the context is a crucial factor. As there is not yet a method scientifically agreed upon for the exact analysis of evaluative expressions, scaling is rather difficult especially when based on the semantic properties of the terms alone (White 2001: 13; Bednarek 2006: 47; Hunston 2007: 39). For example, in the present data, expressions like *poor, unfortunate, miserable*, and *wretch(ed)* proved not only to be highly dependent on the textual context in which they occurred, and by whom they were used, but also on the socio-historical background of the time when they were used. All those instances of person reference which did not contain either explicit or implicit negative and positive evaluation were placed in the neutral evaluation category.

The nominal references to the Mannings were divided into two different types, leaving out all pronominal references. Firstly, I used the category of "explicit evaluation" for explicit reference to an actual person, which usually means the use of proper names and other terms of reference. The term "implicit labelling", secondly, is used here for implicit reference to the person in question by referring to connected events or other text-contextual factors. In Section 5, I will discuss both types of evaluation by giving examples from the newspaper data.

5. Evaluative reference to the Mannings

5.1 Degree of evaluation

Altogether 32,947 instances of nominal person reference to the Mannings, either separately or jointly, were found in the data. The exact number of references to Frederick Manning alone was 13,998, and that to Maria Manning a little lower, 13,782. Instances where they were mentioned together (e.g. *the Mannings, the murderous couple*) totalled up to 5,167. Figure 1 shows the degree of evaluation in the entire data during the three news cycles, divided into neutrally, negatively, and positively evaluative references. All in all, the percentage of neutral evaluation in the data is 42.7%, negative evaluation 38.5%, and positive evaluation 18.8%.

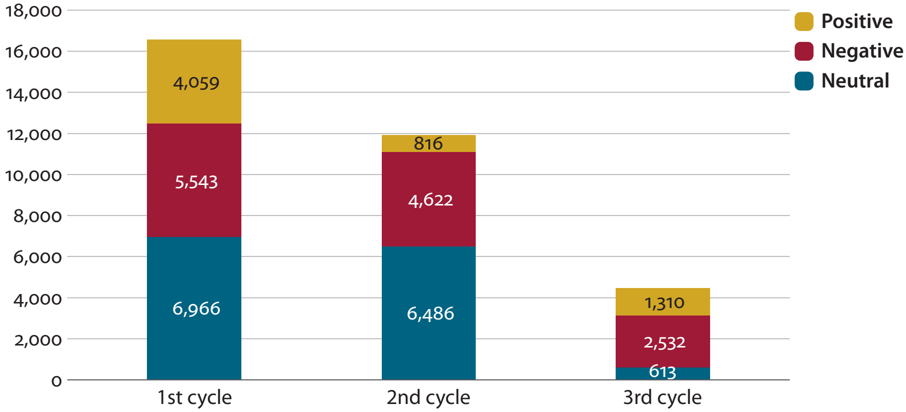

Figure 1. The degree of evaluation in the entire data

We can see that during both the first and the second news cycles, the newspapers mostly use neutrally evaluative references to the Mannings. This can partly be explained by the great number of news pieces which included long stretches of courtroom speeches or police reports of the arrest, in which the culprits were mentioned explicitly by name but without any implicit contextual evaluation. Negatively evaluative references slightly increase (from 33.5% to 38.8%) between the first and the second cycle, whereas positive references radically decrease (from 24.5% to 6.8%). The third cycle, which is the shortest, shows a steep decrease in neutral evaluation (from 54.4% to 13.8%) compared to the second cycle. However, both negative and positive evaluation increase during the last cycle, adding up to 56.8% and 29.4% respectively. This might be explained by the tone or topic of the news reports, which concentrated on either describing the moral degradation of the couple and/or the murder they committed or sympathising with the Mannings' repentance at the time of their hanging.

Figure 2 shows the individual evaluative reference to Frederick Manning. At the beginning of the murder news reporting, reference to Frederick is mostly neutral (39.8%), partly because of the matter-of-fact type and style of accounts of the incident. During the second cycle neutral reference even increases to 56.5%, only to decrease to 13.4% in the third cycle, while positive reference is at its lowest at 9.3%. In general, the last cycle shows an increase in both negative (46.7%) and positive (39.9%) reference, similar to Maria's degree of evaluation, which is displayed in Figure 3.

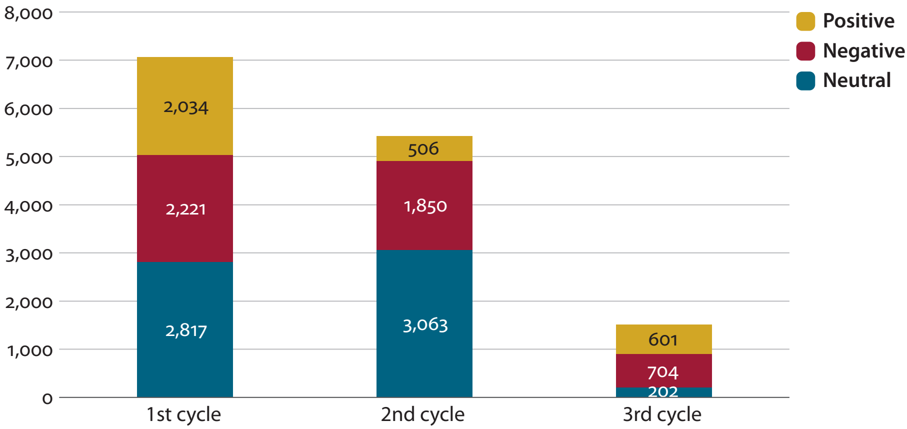

Figure 2. The degree of evaluation in reference to Frederick Manning

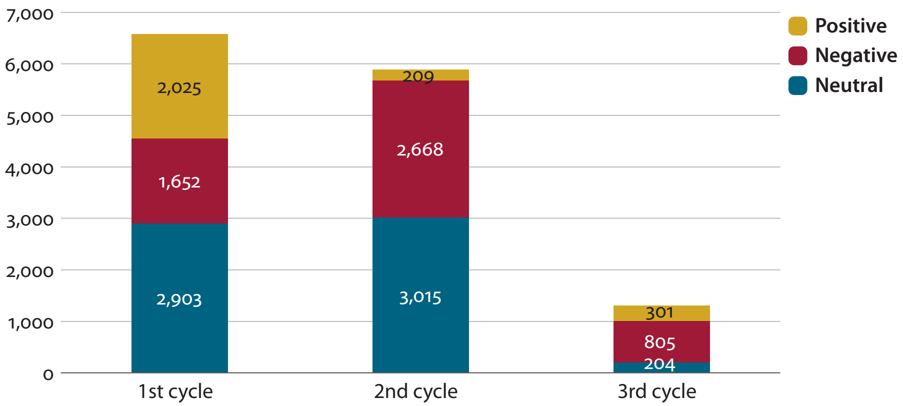

Figure 3. The degree of evaluation in reference to Maria Manning

When we look at the person reference to Maria Manning, we notice a general trend opposite to Frederick's. Neutral reference aside, she is referred to more positively than her husband at the beginning of the murder reporting (30.7%), but from the second cycle onwards she is evaluated more negatively than him. One of the explanations to this development is the increased number of news reports in which her foreign origin is emphasised as a negative influence on her husband or in which her cool behaviour or appearance is somehow taken as incapability of emotion. Also, the fact that Frederick's health deteriorated after he had been imprisoned (second news cycle) may have had an effect on the more extensive negative reference to Maria. There is a relatively high increase in negative reference from the second to the third cycle (45.3 to 61.5%), although an even more considerable one in positive reference from 3.5% to 22.9%. Neutral reference to both Maria and Frederick as individuals is rather similar in numbers during all three news cycles.

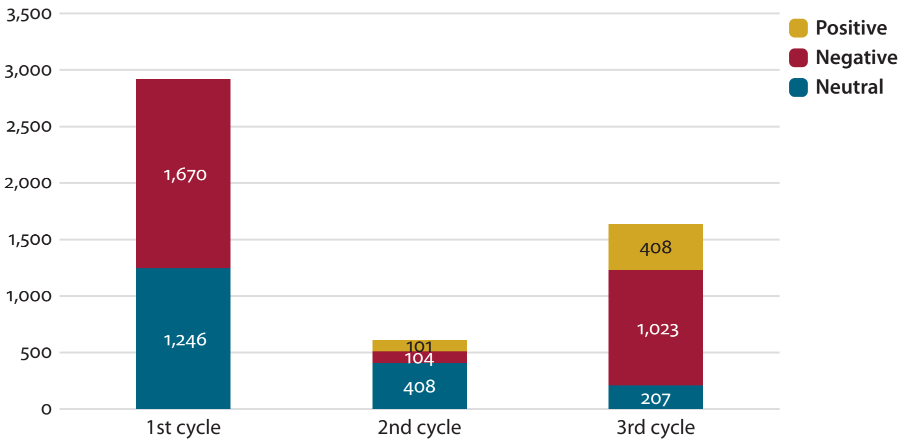

Figure 4. The degree of evaluation in reference to both of the Mannings

Finally, joint reference to the Mannings varies rather considerably between the three cycles, as can be seen in Figure 4. During the first cycle, there are no instances of positively evaluative reference, whereas in the second cycle the percentage of positive evaluation is 16.5% and it increases up to 24.9% in the third cycle. The instances of neutral reference first show an increase from 42.7% to 66.6%, and then a drop down to 12.6% during the last news cycle. The majority of instances appear in the negative category, starting at 57.3% of all references and, after a decrease to 16.9% during the second cycle, showing an increase to 62.5% in the last cycle. The drop in negatively evaluative reference in the middle cycle may be owing to Frederick's health problems, and hence more positive evaluation, mentioned above. This development is interesting, since it appears that when the Mannings are referred to as a couple, the evaluation is more negative than when they, and particularly Frederick Manning, are referred to as individuals. In the next two Sections, 5.2 and 5.3, I will take a closer look at specific examples of the reference to the Mannings, both separately and together.

5.2 Explicit evaluation

As discussed in Section 4, the category of "explicit evaluation" includes terms of direct reference to Frederick and Maria Manning. I have included here nominal phrases, such as proper names and other referential labels, which directly show evaluative stance, often in the form of adjectival attributes or appositives.

The relatively short diachrony of the Bermondsey murder case entails the fast pace at which the reference to the Mannings changed according to the major events, or news cycles, i.e. the murder, the imprisonment, and the hanging. When the body of O'Connor had been found under the kitchen floor and the murder thus revealed, the newspapers were quick to report every single detail of the murder. Reporters speculated about the events which had resulted in the crime as well as about the characteristics of the culprits. Example (1), an extract from one of the first reports, gives a basic account of the background to the murder case. Frederick Manning is referred to as "a man named Manning", with appositives explaining his former occupation ("a discharged railway servant") and previous involvement in criminal acts. Although the explicit reference forms could be categorised as neutral in tone, the implicit description gives a more negative impression. The couple's dubious morale is also stressed by mentioning how the victim had often visited Maria Manning.

(1) The house in question is at present unoccupied, but belongs to Mr. Coleman, a builder, residing at No. 1. About six months ago he let it to **a man named Manning**, *a discharged railway servant, who was dismissed by the directors of the Great Western Railway Company some time ago for being concerned with Nightingale and others in committing a series of robberies on that line*. He resided with **his wife** in Miniver-place, and the murdered man was frequently seen at their house, and especially in the company of **Mrs. Manning**. (*The Derby Mercury*, 22/8/1849)[4]

In Example (2), the attitude towards the Mannings is extremely hostile. The couple is referred to as "these butcher assassins", a phrase which carries the strong implication that the deed was both violent and intentional. It was indeed, and speculations as to whether the driving force behind the deed was passion or self-complaisance were very typical of news writing at the time. What the language in this example is also typical of is that of the newspaper in question, *Daily News*, which represented one of the more radical papers in the nineteenth century.[5] Their word choices could often be somewhat more colourful than those of their contemporaries, particularly in crime news.

(2) One feature in **these butcher assassins** is too remarkable to be unnoticed: they mostly seem to do *their bloody work in a passion* – and perhaps otherwise it could not be done – or, at least, with *a degree of self complaisance*, as if they were more the injured than the injuring party. (*Daily News*, 10/9/1849)

Both the style and tone of reporting were, however, subject to change very quickly according to turns of events. Example (3) was written after the trial, in which

4. In all examples the direct and explicit reference form is given in bold and the implicit evaluative reference or description is in italics.

5. *The Daily News* was founded by Charles Dickens himself in 1846.

it took the jury only 45 minutes to decide the couple was guilty of the murder. Frederick Manning is portrayed here with a more empathetic, even positive, evaluation as an "unhappy relative" of his brother and "[t]he miserable man", most probably owing to his seemingly poor health at the time. He was also pressured by his wife to confess to the murder so she might survive. This he refused to do, and also tried to convince everyone that he was totally innocent.

As already noted in 5.1, reference to Maria Manning started changing towards a more negative evaluation during the second news cycle dating from the trial to the execution. Maria is here deferentially referred to as "Mrs. Manning", but the context tells us that she did not let the prison affect her in any way: she kept dressing up, eating with good appetite, and sleeping well. She also seemed more like a cold person, since she refused to become friendly with the prison staff, unlike Frederick.

(3) After a long pause the brother urged **his unhappy relative** to make his peace with God, who would receive his soul if he was, as he said, innocent of the awful crime. […] **The miserable man** perseveres in his assertion that **his wife** committed the murder, and threatened to take his life also unless he became her accomplice. **Mrs. Manning** still clings to the hope that Lady Blantyre or the Duchess of Sutherland will intercede for her and save her life. *She continues to dress with great care, eats heartily, and sleeps soundly. She attends chapel every morning, and gives very little trouble to those who watch her*. (*Glasgow Herald*, 9/11/1849)

Negativity can be found in the news right after the Mannings' day of execution on November 13. Extreme negative labelling in Example (4) suggests that not much compassion was shown for Maria Manning; instead, she is compared to Jezebel, the wicked seductress from the Bible, who was a foreigner like her. The adjectives *daring, profane, ready, greedy, painted,* and *attired* are all used as appositives to give emphasis to the negatively connotated nouns, most of which relate to negative values (e.g. *unbeliever, arguer, forger*). Because, in general, no drastic differences could be found in the style of reporting between London's tabloid-like papers (apart from *Daily News*, whose style was highly colloquial and sensational) and the national ones, it is interesting to notice the degree of negative evaluation in this particular news piece from *The Times*, which could be classified among the more 'respectable' newspapers.

(4) It is Ahab and Jezebel to the life; – **Jezebel the daring foreigner**, **the profane unbeliever**, as **Maria Manning** now seems to have been; **the ready arguer**, **the greedy aggrandizer**, **the forger**, **the intriguer**, **the resolute, the painted and attired even to death**. (*The Times*, 14/11/1849)

A few days later, Maria is referred to as "[t]he unhappy woman" and "the wretched woman" after she had tried to hang and strangle herself, while her husband is still "the miserable man" (see Example (5)). Regardless of the unfortunate context of her attempted suicide, the reference to the wife is not nearly as empathetic as that to the husband, who appears to have been met with compassion by the newspapers and the reading public throughout most of the ordeal. Even the fact that Maria had let her finger nails grow long was taken as proof of her intentions of getting away with her crime, since she apparently tried to puncture her windpipe with her nails while strangling herself. Again, the repentant Frederick, who was in good terms with the guards, was pictured in a more positive light as he was told to have shown forgiveness by wanting to see his wife.

(5) **The unhappy woman** *attempted self-destruction* in the course of Sunday morning. […] At about four o'clock in the morning she endeavoured to strangle herself. […] As soon, however, as **the wretched woman** saw that the watchers were by her side she seemed to rally herself, and put her head under the clothes. […] While the chaplain was endeavouring to excite a repentant spirit in the heart of his wife, **Manning** at a [sic] about quarter past 7 o'clock breakfasted, and then asked permission to walk in the yard, which was granted. […] It was then about a quarter past 8 o'clock, and while there **the miserable man** expressed to the turnkey who accompanied him *a strong desire to see his wife*. In a few minutes his wish was gratified.
(*Hampshire Advertiser & Salisbury Guardian*, 17/11/1849)

(6) The final exit of **those atrocious criminals**, is so suggestive, that, in common with all our contemporaries, we must make a few more remarks in reference to them. It were well had the public evinced less curiosity than it has manifested in respect to **those wicked people**; […] we would fain impress upon our readers the serious impropriety of regarding the history of **those great offenders** as merely interesting. (*The Era*, 18/11/1849)

The last example of explicit evaluation, (6), is from another account of the hanging. Many newspapers published what could be called 'moral opinion pieces' of the murder case throughout the autumn, but these theoretical ponderings were particularly popular during the third news cycle after the execution. Some of them were extremely negative, as is this one, in which the Mannings are referred to as "those atrocious criminals", "those wicked people", and "those great offenders". The description containing attributes like *atrocious* and *wicked* is typical of any criminal in nineteenth-century Britain, where immoral and illegal acts equalled improper behaviour, something to abhor (cf. Emsley 2005).

5.3 Implicit evaluation

The category of "implicit labelling" includes instances of implicit reference to the Mannings, either by way of referring to connected events (e.g. murder, arrest, imprisonment, execution) or by describing other text-contextual factors, such as the culprits' behaviour or appearance. Implicit evaluation often overlaps with or supports the explicit one, but in some cases the two evaluative strategies may even contradict each other. One such example can be found in (7), in which Maria Manning is first reported as "a remarkably fine woman" who has connections to people "above her own station in life". The seemingly neutral reference term (*Mrs. Manning*) suddenly changes its connotation when the text adopts a more negative tone. Although being a "fine woman", Maria is also very passionate and violent. The account cleverly foregrounds her murderous character, although starting out as seemingly objective.

(7) **Mrs. Manning** is described to be *a remarkably fine woman*, and it is said that *correspondence discovered in her boxes discloses familiarity with persons far above her own station in life. She is a woman of very violent passions*, and Manning's brother states that while she lived in Taunton *she more than once attempted to stab him with a knife during a family quarrel.*
(*The Morning Post*, 23/8/1849)

(8) It appears that the unfortunate deceased cohabited with **Manning's wife** with the knowledge and consent of **the husband**, and that he was frequently in the habit of visiting her when **her husband** was at home; the crime must, therefore, have been committed not from jealousy, but from cupidity. **Manning** is described as *"a fellow fit to do anything."*
(*The Newcastle Courant etc*, 24/8/1849)

In Example (8), we see a description of Frederick Manning, who was labelled more negatively than his wife during the first news cycle right after the murder. At this point, he was said to be "fit to do anything", and the newspapers speculated about his reasons for committing the crime. Later on, the difference in the way Maria Manning was portrayed as opposed to her husband is striking, only then (in the second news cycle) was she more negatively evaluated than her husband. In Examples (9) and (10), imprisoned Frederick is once again described as a feeble, depressed man who walks with "rather a trembling step". His wife, on the contrary, does not falter at all, but walks with "a firm and brisk step". Like in Example (3), Maria appears strong and indifferent to the surrounding people and proceedings, and "asserts to her guardians that she is innocent".

(9) At quarter past two o'clock **the prisoners** were placed in the dock. **Manning** took his usual position, *walking with rather a trembling step, his countenance much more depressed than formerly*. He was followed by **his wife**, who *walked with a firm and brisk step, dressed as on the last examination. She appeared to be as indifferent to the proceedings as upon the preceding occasions, and had a strong flush upon her cheeks, indicative of the mental emotion with which she was inwardly contending.*
(*Freeman's Journal and Daily Commercial Advertiser*, 8/10/1849)

(10) **Manning**, however, *wrote a very spirited letter to his solicitor* (Mr. Binns) on Saturday morning. *He, however, is much depressed in spirits, and watched strictly by the governor and two turnkeys*. **Mrs. Manning** *still keeps up her spirits, and asserts to her guardians that she is innocent, and is sure to be acquitted.* (*The Standard*, 15/10/1849)

As the imprisonment continues, and particularly the shift over to the dreaded Horsemonger Lane Gaol approaches, even Maria's nonchalance begins to disappear. During and after the trial, she starts to behave erratically and in a fury shouts curses at people around her. Example (11) reports of her changing mood from rage to being "quite jocular", which emphasises the popular picture of her as a mad murderess. Frederick, on the other hand, shows "the quiet and submissive demeanour", and makes a full confession. In another report (12), Maria is the weak one surprisingly. She is depicted as bursting into tears and stamping her feet in grief, and, as indifferent as she has been in previous accounts, the extract ends with a more positive evaluation: "The woman, though hardened, was not entirely deadened to feeling".

(11) While being removed to Newgate, preparatory to being shifted to Horsemonger-lane gaol, and during their stay in the prison, **Mr.** [**sic**] **Manning's** *excitement rose to an ungovernable pitch*. On being handcuffed to a turnkey, *so great was her fury, that she exclaimed, "Damnation cease you all!"* The storm was too violent to last long, however; and *by-and-bye she became quite jocular, joked about her manacies, boasted of the resolution she had displayed before the Judges, and took leave of Newgate in regular Jack Sheppard style!* […] **Manning** continued to observe *the quiet and submissive demeanour* maintained during the trial; and paid attention to the exhortations of the Rev. Mr. Roe, the chaplain of Horsemonger-lane gaol, to whom, it is understood, *he has made a full confession*.
(*North Wales Chronicle*, 30/10/1849)

(12) **Mrs Manning**, attended by the principal female turnkeys of the prison, was conducted to one of the cells set apart for the condemned, upon entering which *she burst into tears – sobbing piteously and exciting the commisseration of all who heard her. She stamped the floor with her feet, not in rage, but in grief*; her feelings no doubt being overcome as she reflected upon the great change that had taken place in her destiny since she last saw the officers by whom she was surrounded. **The woman**, *though hardened, was not entirely deadened to feeling.* (*Caledonian Mercury*, 1/11/1849)

Interestingly, the victim, Patrick O'Connor, was described in negative terms in many of the articles about the murder. Several character witnesses testified in the trial that he had been no angel himself, instead, he was – as referred to in Example (13) – "a sordid, abandoned wretch", whom nobody liked or missed.[6]

(13) O'Connor was a sordid, abandoned wretch, for whom it is impossible to entertain the least sympathy. He seems to have left behind him few or none to regret him. He will, therefore, be little missed, and the world is perhaps as well without him; but this affords no extenuation of *the horrible crime of* ***the Mannings****, which in its appalling and cold blooded details – a woman figuring as the chief actor – is almost without a parallel in the annals of crime.* (*Dundee Courier*, 7/11/1849)

It is noticeable, however, that when, at the end of the passage in (13), the topic returns to describing the deed ("the horrible crime", "appalling and cold blooded", "almost without a parallel in the annals of crime"), the emphasis is again more on Maria, "the chief actor". No matter if the victim was "little missed", and what a criminal he himself had been, it did not warrant him being cold-bloodedly murdered.

(14) **Manning**, *apparently unable to control his feelings longer*, leaned forward towards **his wife**, and *in the most imploring accent* said: – "I hope you are not going to depart this life with feelings of animosity toward me." *The appeal was too much for her*, and leaning toward him she said, "I have no animosity towards you." He said, "Will you not kiss me, then?" **The female convict** said "Yes," and **both parties** having arisen, *they shook hands and kissed each other several times.* (*Daily News*, 14/11/1849)

6. An interesting detail concerning the reference to the victim: Irish newspapers like *Freeman's Journal and Daily Commercial Advertiser* overwhelmingly used O'Connor's name in the headlines to the articles about the murder, presumably because he was Irish (e.g. "The Murder of Patrick O'Connor at Bermondsey", 8/10/1849). English newspapers used the name *Manning* instead.

Many newspapers feasted on eye-witness accounts of the couple's behaviour right before the execution as well as during it. One of the things that the reporters wanted to know about was not only how the Mannings felt about their sentence and the fact that they were going to die for their crime, but also the state of Frederick and Maria's relationship. The couple had been blaming each other for the murder from the start, and now people wanted to know whether they could reconcile. In some accounts, the reading audience was granted a more positive end to the story; in Example (14), emotional Frederick approaches his wife, not being able to "control his feelings longer", and asks her to kiss and make up. Maria shakes his hand and kisses him, confirming that she has "no animosity" towards him.

6. Conclusions on the Mannings murder reporting

6.1 General trends

Returning to the research questions at the beginning of this chapter, we can answer the first two by looking at some diachronic trends emerging from the data. During the first and the second news cycles the newspapers mostly used neutrally evaluative references for the Mannings. Negative evaluation showed a slight increase between the first and the second cycles, whereas positive references decreased. The third cycle then entailed a steep decrease in neutral evaluation from the second cycle, which might be the result of a more involved and emotional style of reporting in a higher number of opinion and social discussion pieces. This is probably also the reason why both negative and positive evaluation increased towards the end of the last cycle, since the tone and topic of the news reports often described such issues as the moral degradation of the Mannings or sympathised with the couple's deteriorating health or their repentance before their hanging.

If we look at Frederick and Maria Manning separately, we notice certain differences in how reference to them changed from the first to the third cycle. The degree of evaluation to Frederick was slightly more negative than that of Maria during the first cycle. From the second cycle onwards, however, the positive references to Frederick increased. At the beginning of the murder news reporting, reference to Frederick was mostly neutral, and, during the second cycle, neutral reference increased, probably owing to reports of the trial. The last cycle showed an increase in both negative and positive reference to Frederick (see Figure 2). In general, reference to Maria (see Figure 3) showed rather opposite trends compared to Frederick. She was referred to more positively at the beginning of the murder reporting, but from the second cycle onwards the situation changed and she appeared to be evaluated more negatively.

It appears that when the Mannings were referred to as a couple, the evaluation was more negative than when they, and particularly Frederick, were referred to as individuals. Joint reference to the Mannings varied rather considerably between the three cycles (see Figure 4). During the first cycle, there were no instances of positively evaluative reference, whereas the percentage of positive evaluation then increased towards the third cycle. The instances of neutral reference first showed an increase, only to drop again in the end. The majority of examples referring to the Mannings together appear in the negative category.

6.2 Culprits, creatures, or criminals?

If we go back to what Van Leeuwen (2008: 109) calls "moral legitimization", we can say that the Mannings seem to have been categorised negatively when judged against the moral and social values of respectable society. Committing a murder was reason enough for them to be cast aside and put into the group of unwanted margins. Their social position, and hence their social representation, was not very enviable to start with, and, through the public judgement and condemnation shown by the newspapers feasting on their crime, it turned into a more excluded one. In other words, when evaluating the Mannings by increasingly using negative terms and attributes, the 'good' in-groupers placed them in a 'bad' out-group. The main hypothesis of this article, whether the main social representation of criminals at the time became the way in which the Mannings were described as individuals, appears to have proved correct. In explicit evaluation, attributes like *atrocious, wicked, evil*, and *revolting* were used by the press, alongside such terms as *butcher, assassin, creature*, and *monster*. The attributes given to murderers reflected in general the value tapestry of Victorian society, in which negative imagery was often biblical and wrong-doers were depicted as animals, devils, or monster-like creatures. In implicit evaluation, the Mannings, the murder, and its repercussions were described as *cold-blooded, horrible*, and *appalling*. Although an extensive share of evaluative reference was rather neutral in tone, the Mannings were in general referred to as negatively as other murderers during the nineteenth century.

Two questions about the style of reporting arise from the data. Firstly, why is there so much neutral reference in the data? For the most part, this appears to pertain to the matter-of-fact reporting either of what happened right after the murder had been discovered or later of what people said in the courtroom during the trial. Secondly, does the style of reporting also dictate how and how much positive and negative evaluation is used? This may partly be true, since the most colourful language or, in other words, negative evaluation appeared in social discussion pieces that were increasingly more typical towards the end of the third news cycle, during and after the execution. Also, as already mentioned in Section 5.2, although no big differences between various newspapers (at least between *The Times* and others) were found in the data, there is an indication that some of the most negative references to the Mannings appeared in tabloid-like London broadsheets and were then copied to some provincial newspaper, often verbatim. Apart from the tone and style of writing, the details which were pointed out in the news were crucial tools for placing the wrong-doers outside respectable society. In the Mannings case, emphasising that Maria Manning was of foreign origin or that she seemed emotionless and cold-hearted, while her husband was of poor disposition and repentant, surely influenced the public judgement of the reading audience.

An earlier study (Nevala 2016) shows that the dichotomous representation of criminals and their victims was sometimes done quite randomly in nineteenth-century news media. Therefore, the third research question about whether the social identities of the Mannings could be considered stable or variable is difficult to answer. If we looked at the number of negative references to the couple alone (see Figure 4), we could say that the social representation of the Mannings in the newspapers clearly reflected the way in which they were shunned into the out-group of detested criminals. As the analysis also shows, however, there was situational variation. On closer scrutiny, I found many examples expressing compassion towards the Mannings in the data. This momentary empathy was often caused by a particular event, like a suicide attempt, an injury, a confession to the priest, and an apology to the accomplice. All these incidents evoked pity towards the culprits, but also meant that their social image shifted closer to a more positive one for a while – from totally atrocious to slightly acceptable.

Acknowledgements

My work on this chapter is done as part of the project *Democratization, Mediatization and Language Practices in Britain, 1700–1950* (DEMLANG), funded by the Academy of Finland (project number 295381).

References

Alba-Juez, Laura, and Geoff Thompson. 2014. "The Many Faces and Phases of Evaluation." In *Evaluation in Context*, ed. by Geoff Thompson, and Laura Alba-Juez, 3–23. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.242.01alb

Allport, Gordon. 1986. "The Language of Prejudice." In *Language Awareness*, ed. by Paul Escholz, Alfred Rosa, and Virginia Clark, 261–270. New York: St. Martin's Press.

Bednarek, Monika. 2006. *Evaluation in Media Discourse: Analysis of a Newspaper Corpus*. London: Continuum.

Bednarek, Monika, and Helen Caple. 2012. *News Discourse*. London: Bloomsbury.

Clark, Kate. 1992. "The Linguistics of Blame: Representations of Women in the Sun's Reporting of Crimes of Sexual Violence." In *Language, Text and Context: Essays in Stylistics*, ed. by Michael Toolan, 208–224. London: Routledge.

Conboy, Martin. 2002. *The Press and Popular Culture*. London: SAGE.

Conboy, Martin. 2010. *The Language of Newspapers: Socio-Historical Perspectives*. London: Continuum.

Diamond, Michael. 2003. *Victorian Sensation, or the Spectacular, the Shocking and the Scandalous in Nineteenth-Century Britain*. London: Anthem Press.

Dickens, Charles. 1852–1853. *Bleak House*. London: Bradbury & Evans.

Emsley, Clive. 2005. *Crime and Society in England, 1750–1900*. Harlow: Longman.

Hogg, Michael A. 2005. "Uncertainty, Social Identity and Ideology." In *Social Identification in Groups*, ed. by Shane R. Thye, and Edward J. Lawler, 203–229. Amsterdam: Elsevier. https://doi.org/10.1016/S0882-6145(05)22008-8

Hogg, Michael A., and Dominic Abrams. 1988. *Social Identifications*. London: Routledge.

Hunston, Susan. 2007. "Using a Corpus to Investigate Stance Quantitatively and Qualitatively." In *Stancetaking in Discourse: Subjectivity, Evaluation, Interaction*, ed. by Robert Englebretson, 27–48. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.164.03hun

Linnane, Fergus. 2003. *London: The Wicked City*. London: Robson Books.

Linnane, Fergus. 2004. *London's Underworld: Three Centuries of Vice and Crime*. London: Robson Books.

Martin, James R. 2000. "Beyond Exchange: Appraisal Systems in English." In *Evaluation in Text*, ed. by Susan Hunston, and Geoff Thompson, 142–175. Oxford: Oxford University Press.

Martin, James R., and Peter R. R. White. 2005. *The Language of Evaluation: Appraisal in English*. Basingstoke: Palgrave Macmillan. https://doi.org/10.1057/9780230511910

Martinez Caro, Elena. 2014. "The Expression of Evaluation in Weekly News in English." In *Evaluation in Context*, ed. by Geoff Thompson, and Laura Alba-Juez, 321–343. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.242.16mar

Mayr, Andrea, and David Machin. 2012. *The Language of Crime and Deviance: An Introduction to Critical Linguistic Analysis in Media and Popular Culture*. London: Continuum.

Moscovici, Serge. 1984. "The Phenomenon of Social Representations." In *Social Representations*, ed. by Robert M. Farr, and Serge Moscovici, 3–69. Cambridge: Cambridge University Press.

Nevala, Minna. 2016. "Solidarity in Evaluation: The Portrayal of Murderers and Their Victims in Late Nineteenth-Century Newspapers." In *The Pragmatics and Stylistics of Identity Construction and Characterisation* (Studies in Variation, Contacts and Change in English 17), ed. Minna Nevala, Ursula Lutzky, Gabriella Mazzon, and Carla Suhr. Helsinki: VARIENG. Available online at: http://www.helsinki.fi/varieng/series/volumes/17/nevala/ (accessed 15 April 2019).

Nevala, Minna. 2017. "The Public Identity of Jack the Ripper in Nineteenth-Century Newspapers." In *Diachronic Developments in English News Discourse*, ed. by Minna Palander-Collin, Maura Ratia, and Irma Taavitsainen, 199–216. Amsterdam: John Benjamins. https://doi.org/10.1075/ahs.6.11nev

Ochs, Elinor. 1993. "Constructing Social Identity: A Language Socialization Perspective." *Research on Language and Social Interaction* 26: 287–306. https://doi.org/10.1207/s15327973rlsi2603_3

O'Reilly, Carole. 2014. "'Dirt, Death and Disease': Newspaper Discourses on Public Health in the Construction of the Modern British City." *Journal of Historical Pragmatics* 15: 207–227. https://doi.org/10.1075/jhp.15.2.04ore

Tabbert, Ulrike. 2015. *Crime and Corpus: The Linguistic Representation of Crime in the Press*. Amsterdam: John Benjamins. https://doi.org/10.1075/lal.20

Thompson, Geoff, and Susan Hunston. 2000. "Evaluation: An Introduction." In *Evaluation in Text*, ed. by Susan Hunston, and Geoff Thompson, 1–27. Oxford: Oxford University Press.

Van Dijk, Teun A. 2009. *Society and Discourse: How Social Contexts Influence Text and Talk*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511575273

Van Leeuwen, Theo. 2008. *Discourse and Practice: New Tools for Critical Discourse Analysis*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195323306.001.0001

White, Peter R. R. 2001. "Appraisal Outline." http://www.grammatics.com/appraisal/ (accessed 15 April 2019).

Wodak, Ruth. 2008. "'Us' and 'Them': Inclusion/Exclusion – Discrimination via Discourse." In *Migration, Identity, and Belonging*, ed. by Gerard Delanty, Ruth Wodak, and Paul Jones, 54–78. Liverpool: University of Liverpool Press.

The Prince and the Sassenach

Constructing group homogeneity through labels (and anachronisms) in Late Modern times and beyond

Marina Dossena
Università degli Studi di Bergamo

In political communication group identity construal is often oversimplified: the supposed homogeneity of individual groups is emphasized, while their mutual antagonism is highlighted. This phenomenon also has considerable time depth and even influences the historical (re)construction and (re)interpretation of events. In this respect, the Jacobite cause is a significant example: narrations that have circulated for centuries in popular culture have contributed to the popularity of labels like 'Bonnie Prince Charlie', 'Redcoats' and 'Sassenach' well beyond the historical context in which they could be viable. In this contribution I aim to outline how different labels pertaining both to people and to social groups are seen to occur in present-day and in Late Modern English uses, in order to analyse how certain items have become so persistent as to (almost) exclude alternatives from general usage, not least in public discourse and especially in the tourist industry.

Keywords: Late Modern English, popular culture, onomastics, identity, argumentation

1. Introduction

The Naming of Cats is a difficult matter, […]
First of all, there's the name that the family use daily, […]
But I tell you, a cat needs a name that's particular, […]
But above and beyond there's still one name left over,
And that is the name that you never will guess;
The name that no human research can discover – […]

T. S. Eliot, *Old Possum's Book of Practical Cats* (1939)

https://doi.org/10.1075/pbns.306.02dos

Political communication is a prototypical example of clear-cut identity construal: especially in texts meant to encourage solidarity within groups, and indeed meant to persuade through more or less emphatic propaganda, the presupposed homogeneity of the groups themselves is a crucial element. As a result, in a perennial game of 'Us' vs 'Them' (see Van Dijk 1995: 22), the beliefs and behaviour of opponents are stigmatized and/or ridiculed, while one's own party is praised and set up as an example of bravery, probity, and any other positive quality that may be highlighted concerning it. This implies that any differences within groups are levelled out, while differences between groups are stressed.[1] In addition, this process of group homogeneity construal may be shown to be at work both synchronically and diachronically; not only are differences described as key features of the groups under discussion at the time of the events and as the debates rage: they are also important elements in their historical (re)construction and (re)interpretation.

In this respect, the Jacobite cause is a significant example: nowadays even users whose familiarity with Scottish history is minimal will recognize and understand labels like 'Bonnie Prince Charlie', 'Redcoats' and 'Sassenachs', mostly thanks to narrations of historical events in popular culture, whether in nineteenth-century songs and ballads or in twenty-first century fiction and TV drama. Such narrations, however, can be anachronistic, because they are subjective reconstructions of events based on what is known and, at least occasionally, on what is imagined; as a result, this may decrease the reliability of the text for the sake of its empathetic appeal. This process, in turn, can be shown to have an impact on public discourse, especially when labels become crystallized, because any bias they may convey could become less and less perceptible as their circulation spreads among non-specialists. For this reason, in this contribution I will attempt to assess labels used in different Late Modern English documents pertaining (mostly, though not exclusively) to the Jacobite rebellions, in order to study the origins of some persistent labels identifying both events and people. Within this framework, special attention will be given to ego documents and popular culture, on account of their viability across a wide range of language users.

After an overview of the ways in which names, pseudonyms and nicknames are seen to occur in various domains, I will concentrate on how these labels are employed in present-day public discourse for informative or argumentative purposes. More specifically, I will consider how labels can be manipulated to the point of oversimplification in materials aimed at the general public, such as in the case of promotional discourse in the tourist industry both in Britain and in the US.

1. Beyond political communication, the studies by Nevala (2016 and 2017) provide an interesting account of how these mechanisms operate in the linguistic construal of identities of both criminals and victims in the Victorian press.

The case of "The Forty-five" will be shown to be emblematic of how audiences are provided with representations that reflect their expectations, but which do not necessarily correspond to a reliable reconstruction.

The materials under investigation will be analyzed within a Discourse-historical approach (Wodak and Meyer 2001), so as to assess their argumentative quality. At the same time, it should be noted that, within this framework, no quantitative data could be offered in any reliable way, because labels are an open-class category, and therefore the representativeness of the inventory under investigation can hardly be guaranteed. However, in Section 3 some quantitative findings will be provided in relation to some key items concerning Charles Edward Stuart, the popularity of whose labels are of specific interest for the purposes of my study.

Finally, some concluding remarks will be offered. First of all, however, an outline of how labels may serve different pragmatic purposes, especially in popular culture, is given in Section 2 below.

2. What's in a name? Nicknames, memorability and pragmatics

In their very recent book on socio-onomastics, Ainiala and Östman (2017: 1–12) present an overview of the approaches that so far have been taken in linguistic studies of names, not least in relation to deixis and politeness theory, but also from a diachronic point of view, as onomastics was a very significant branch of early philological investigations. Attempts to identify the origins of place names, surnames, and indeed personal names have informed many studies of language varieties since Late Modern times, as we see in various glossaries collected in the *Salamanca Corpus*.[2] In addition, Ainiala and Östman (2017) draw attention to the importance of name variability and the ways in which it is employed to convey stance in different sociological contexts. The choice of names, nicknames, pet names, and even slang and abbreviated forms of toponyms, whether rural or – perhaps more frequently – urban, reflects varying degrees of familiarity, and may thus be indexical of in-group affiliation. Their adoption or rejection may signal greater or lesser cultural acknowledgement of that affiliation – a choice that places users within or outside the network to which the labels refer. Indeed, this is seen

2. See for instance Bannister (1869–1871) and Ellwood (1895). *The Salamanca Corpus*, currently comprising more than 450 texts, i.e. more than 12m words, is a digital archive of English dialect texts spanning more than four centuries (c.1500–c.1950), thus supplementing the *English Dialect Dictionary* (1898–1905). Although for some areas no materials have been traced yet, it nonetheless offers a very valuable collection of both dialect literature and literary dialect, glossaries and other texts on folk traditions and uses, thus shedding light on Early and Late Modern views of popular culture and linguistic features.

in the creation and usage of derogatory labels, such as in the case of xenophobic slurs, which stress the distance existing between the social group indicated by the label, and the one using it.

A consequence of this socio-pragmatic feature of labels is that they can be used in so many different and complex ways that in their study we should attempt to identify categories which can convey their pragmatic values in relatively fine-grained terms. To this end, it may be useful to employ "label" as a superordinate, regardless of its positive, neutral or negative connotation. Within this broad category, we can distinguish between actual names, pseudonyms, and nicknames, as outlined below:

- actual names comprise personal names, both in their full and hypocoristic forms (e.g. *John, Johnny, Jack*); family names, e.g. *Phillips*; toponyms, e.g. *Chester*; and general social category names, e.g. *kids, students*;
- pseudonyms may replace a person's name in full or in part: e.g. *George Eliot* for *Mary Anne Evans* (English novelist, 1819–1880); *Tony Curtis* for *Bernard Schwartz* (American actor, 1925–2010); *David Tennant* for *David McDonald* (British actor, 1971–); *Marilyn Monroe* for *Norma Jean Baker* (American actress, 1926–1962) etc.; finally,
- nicknames replace the actual name (of a person or a place) relying on lexical alternation which may also comprise pre- or post-modifiers: e.g. *Albion* for *England*; *the Big Apple* for *New York City*; *Auld Nick*, a popular Scots euphemism for the devil;[3] *Malleus Scotorum* (Hammer of the Scots)[4] for *Edward I of England* (1239–1307).

In usage these groups can occur interchangeably, depending on the communicative purpose of the text or utterance for which they are selected. A pseudonym may replace the actual name systematically, to the point that outsiders may not be aware that the label is in fact a fabricated one and that it may have little if anything in common with the name it hides. Nicknames, instead, may occur with varying frequency and do not necessarily replace the actual name throughout the text in which they are employed – in fact, they may serve the double purpose of avoiding

3. In folklore, nicknames are often used euphemistically in order to avoid names that are considered unlucky or to show superstitious respect, such as in the case of fairies, who are called *the good people* so as not to hurt their sensitivity and thus prevent their revenge – see the entry for *guid* in the *Scottish National Dictionary*, s.v., 7. (22).

4. This label is actually in the Latin epitaph carved on his tomb in Westminster Abbey (see www.westminster-abbey.org/our-history/royals/edward-i-and-eleanor-of-castile); his other nickname was a more descriptive (and less politically charged) *Longshanks*. Note that all websites mentioned in this chapter were last accessed in April 2019, unless otherwise indicated.

repetition and creating variety in the text: it is the case, for instance, of *Auld Reekie* replacing *Edinburgh* in materials in which the equivalence of the two labels can be recognized easily either because the author has placed them in such a context that one may function as endophoric reference for the other, or because readers are assumed to be familiar with Robert Fergusson's 1773 poem from which the nickname is borrowed, if not with the poet's works more in general.

Interestingly, pseudonyms and nicknames are typically found in such apparently distant domains as the arts, sports, politics and history, signaling an intriguing contiguity between academic subjects and popular culture.[5] In what follows, some instances of this phenomenon are discussed, starting from the language of entertainment, i.e. pertaining to sports, music, (popular) literature and the film industry – see Section 2.1 below. In Section 2.2, instead, I will focus on the use of labels in representations of historical figures and periods, so as to provide a background overview to my analysis of labels related to the Jacobite risings of the eighteenth century.

2.1 Labels in the language of entertainment

Entertainment is an overarching term comprising different domains in which labels are seen to occur in various ways, depending on the generic specificity of the domains themselves. In sports, pseudonyms are perhaps less frequent than nicknames, although both cases can be observed: on the one hand, football enthusiasts know all about the games in which the protagonist was Edson Arantes do Nascimento, aka *Pelé* (1940–); on the other hand, they may remember Diego Armando Maradona (1960–) as *El pibe de oro* ('the Golden Boy'), employing the Spanish version of a nickname that, in English, had also applied to Italian footballer Gianni Rivera (1943–). Nor is football unique in the use of nicknames: in international motorbike racing Valentino Rossi (1979–) is also known as *Il dottore* ('The Doc'), on account of an honorary degree in Promotional Communication he was awarded in 2005; Marco Pantani (1970–2004), an Italian cycling champion, instead, was called *Il pirata* ('The Pirate') because of his look, which typically comprised earrings and a bandanna.

Perhaps not unexpectedly, these labels are often culture-bound: it is the case, for instance, of *Minnesota Fats*, a fictional pool hustler whose international fame is due to a 1961 film starring Paul Newman (1925–2008).[6] Indeed, popular culture has contributed greatly to the memorability of labels like *The Sundance Kid* (i.e. Harry Longabaugh, 1867–1908), *Billy the Kid* (i.e. Henry McCarty, 1859–1881) or

5. An early study of historical nicknames is provided by Oswald (1895).

6. The film was *The Hustler*, directed by Robert Rossen and based on Walter Tevis' 1959 novel with the same title.

'Buffalo' Bill Cody (i.e. William F. Cody, 1846–1917), but this phenomenon is not restricted to literature, films and shows. Also in music pseudonyms are frequently found: one example among countless is *Elton John*, whose actual name is Reginald Kenneth Dwight (1947–). Moreover, nicknames can contribute to the complexity of an artist's public image: for instance, David Bowie's (1947–2016) popularity is famously connected with at least two of his best-known personae, *Ziggy Stardust* and *the Thin White Duke*. In this respect Olsen (2006) discusses the importance of nicknames in jazz music in her article on how memorials record or ignore labels which are in fact very familiar to connoisseurs and to more general audiences. Some instances are found in the Woodlawn Cemetery in the Bronx, where one cannot find the grave of *Duke Ellington* (1899–1974), but of Edward Kennedy Ellington, while Miles Davis' sarcophagus bears the inscription "In Memory of *Sir Miles Davis* 1926–1991", although the title only refers to the fact that he was a member of the Knights of Malta, not an MBE like Sir James Paul McCartney (1942–) or Sir Richard Starkey, aka *Ringo Starr* (1940–).

Sometimes, however, the adoption of new names signals a change of political, religious and cultural affiliation: Cassius Clay (1942–2016) and Cat Stevens (i.e. Steven Demetre Georgiou, 1948–) changed their names to Muhammad Ali and Yusuf, respectively, when they converted to Islam. Similarly, Malcolm X (1925–1965) famously made a political statement when he renounced his Anglo-American surname (Little). The significance of onomastic labels in historical and political discourse therefore deserves to be considered in its own right, and will be the object of the next section.

2.2 Labels in the language of (political) history

In politics, both individuals and social organizations can be identified on the basis of pseudonyms: everybody knows Vladimir Ilych Uljanov (1870–1924) as *Lenin* and Iosif Vissarionovič Džugašvili (1878–1953) as *Stalin*, while at the beginning of the twentieth century the Industrial Workers of the World were known as *Wobblies* and the Pennsylvania miners who sabotaged several plants in the second half of the nineteenth century were called *Molly Maguires*, employing a generalizing pseudonym which both reversed gender identification and evoked the Irish origin of many workers.[7]

7. Although the etymology of *Wobbly* is unknown (see the official site of the Industrial Workers of the World, www.iww.org/da/history/icons/wobbly), it is nonetheless in current use also in scholarly investigations – e.g. Cartosio (2007). As for *Molly Maguires, Encyclopædia Britannica* states that "the group named itself after a widow who led a group of Irish antilandlord agitators in the 1840s" (see www.britannica.com/topic/Molly-Maguires).

Indeed, references to origin or other intrinsic features have been listed by Abel (2004) as underpinning the nicknames attributed to American Civil war generals; predictably, such labels could suggest hostility or affection, depending on which side created and circulated them – for instance, Confederates called Union General Ulysses Grant (1822–1885) *Butcher*, while George Armstrong Custer (1839–1876) was nicknamed *the Boy General* by his own (Union) side. Moreover, opposing armies or groups often have specific labels; these can be self-attributed, such as in the case of *Green Berets*, the US Army Special Forces, or they may reflect outsiders' views: it is the case, for instance, of names attributed to Native American Nations that are in common use today, but which do not correspond to their autonym. Among such cases we find the Niimíipuu nation, called *Nez Percé* by French explorers, and the Dineh nation, which is normally called *Navajo*.

As a matter of fact, even historical periodization relies on a series of labels which function as shorthand terms for a specific time frame, although this may span centuries: it is the case of the *Middle Ages*, the *Renaissance*, the *Victorian Era*, or the *Gilded Age*. Sometimes these labels are again culture-bound: the Italian word *Risorgimento* cannot actually be translated because of its historical uniqueness; *Resistenza*, instead, does have translations on account of its international dimension. In a similar vein, in historical narratives the use of eponyms and nicknames is a widespread phenomenon: not only do historical figures acquire postmodifiers that identify them univocally and somewhat 'officially', albeit always in evaluative terms (e.g. *Alexander the Great, William the Conqueror, Richard the Lionheart, Ivan the Terrible*), but they may also be identified by means of qualifiers or indeed nicknames which immediately convey the image that those who created those labels wanted to establish:[8] extreme cruelty in the case of *Bloody Mary* (Mary I Tudor, 1516–1558), chastity in the case of *The Virgin Queen* (Elizabeth I Tudor, 1533–1603), strong-willed power in the case of *The Iron Lady* (Margaret Thatcher, née Roberts, 1925–2013) or fierce and undiscussed leadership in the case of *Il duce* (Benito Mussolini, 1883–1945) and *Führer* (Adolf Hitler, 1889–1945).

As we see, these labels do not necessarily include the name of the person being identified. When the name is omitted, the presupposition concerning the recognizability of the subject is stronger, thus also emphasizing the in-group solidarity that derives from shared background knowledge. However, the viability of these

8. In this sense it would be practically impossible to separate labels into neat categories according to their polarity: as we will see in Section 3 below, the same referent could be represented by a positive or a negative label, depending on the point of view of the subjects attributing the labels. This, in turn, implies that usage of one label as opposed to the other defines the in-group affiliation of the users themselves, who thus qualify as supporters or opponents, depending on what linguistic choice they make.

labels may be relatively limited in time and/or space – for instance, *The Great Helmsman* (Mao Zedong, 1893–1976) was certainly more transparent a few decades ago than it is today. Unless the labels become somehow 'lexicalized', and thus begin to be used in texts which have a certain degree of permanence, such as history books, but also successful popular documents, such as songs, ballads, and even literary texts, they are bound to become increasingly opaque as their frequency of usage decreases.

Indeed, labels can become almost empty signifiers, to the point that they can become the object of spoof and satire: the best-known examples of this are probably in *1066 and All That* (Sellar and Yeatman 1930: passim), where we find references to the "Venomous Bead (author of the Rosary)", "the great St Bernard (originator of the clerical collar), […], St Kit and St Kin", and various "Egg-Kings", such as "Eggberd, Eggbreth, Eggfroth, etc." The humorous conflation of names, modifiers, historical references and literary adaptations results in a *pastiche* which aims to entertain the reader both with its preposterous malapropisms and with the game of hide-and-seek that the readers are invited to play (Dossena 2017a); the text challenges its readers to recognize the correct forms, thus flattering their supposed historical competence. An emblematic case is the section in which King Alfred and King Arthur are mentioned alongside "Good King Wenceslas", the protagonist of a popular Christmas carol, while in Chapter 7 we read that

> ALFRED had a very interesting wife called Lady Windermere (The Lady of the Lake), who was always clothed in the same white frock, and used to go bathing with Sir Launcelot (also of the Lake) […]. It was also in King Arthur's time that the Anglo-Saxon Chronicle was published: this was the first English newspaper.

References to modern literature (Oscar Wilde's *Lady Windermere's Fan*), Arthurian romances, and actual history, albeit in a distorted form (the Anglo-Saxon Chronicle can hardly be said to have been a newspaper), contribute to a humorous representation of history perceived as somewhat confusing, but certainly entertaining[9] – a mechanism that is still at work in Terry Deary's *Horrible Histories* series, begun in 1993, and comprising titles like *The Vicious Vikings, The Groovy Greeks*, or *The Terrible Tudors*: titles in which alliteration is meant to make the labels both memorable and amusing.

In historical narrative, however, nicknames for the protagonists appear to have a more complex and argumentative function. The so-called *Robber Barons*, i.e. capitalists of the Gilded Age (Sauers 2006), are often identified by means of labels

9. Along similar lines, Monty Python, a well-known British comedy group, lampooned Arthurian romances in their 1975 film *Monty Python and the Holy Grail* and the New Testament in *Monty Python's Life of Brian* (1979).

that summarize their cavalier and/or aggressive approach to business: for instance, James Fisk Jr "was dubbed the 'Barnum of Wall Street'" (Sauers 2006: 106), while Jay Gould was nicknamed the "Mephistopheles of Wall Street" (Sauers 2006: 159). The pattern identified in these labels is <Definite article> + <(generally disparaging) label> + <of> + <placename> and would be repeated in the title of a 2013 film, *The Wolf of Wall Street*, in which the animal metaphor signals both aggressiveness and relative isolation.[10]

Metaphorical uses also appear in the names with which Native Americans have been identified since Late Modern times and which are seldom reported in their original form. An early twentieth-century travelogue, Simpson (1903), gives a very clear example of the way in which names are not only translated, but actually replaced altogether, when he describes the proceedings in a reservation in the White Rolling Bluffs of Missouri. In the local school he meets children with supposedly amusing exotic names, and readers are promptly informed that they also have 'modern' names; however, nothing is said of their native names – see the quotation below:

> When the school had assembled the first thing done was to call the roll: a list of names such as Maneater, Young Snake, Little Hawk, Good Chief, Sawn Egg, White Breast, Big Bear, Tabow, Armbroke, Sportman, Snowball, Rainbow, White Wolf, etc. The children were given modern names such as Washington, Victor Hugo, Daniel Webster, Oliver Goldsmith, Napoleon Bonaparte, etc.
>
> (Simpson 1903: 170)

A few years later, in Dixon's 'account' of the last Great Indian Council, held in the valley of the Little Horn, Montana, in September 1909, the protagonists are presented "with their English, tribal, and Indian designations", such as in (1).

(1) CHIEF PLENTY COUPS, Chief of the Crow Nation, bearing the Indian name of Aleck-shea-Ahoos, signifying Many Achievements. […] (Dixon 1913: i)

10. The film, starring Leonardo DiCaprio in the title role, was directed by Martin Scorsese and was based on the true story of the protagonist, Jordan Belfort. Interestingly, the title also echoes the use of labels that is found in titles like *The Bride of Lammermoor* (authored by Walter Scott in 1819), *The Master of Ballantrae* (authored by Robert Louis Stevenson in 1889) or *The Man Who Would Be King* (authored by Rudyard Kipling in 1888), in which the protagonist is identified by means of a post-modifying phrase meant to elicit curiosity through its simultaneous specificity (introduced by the definite article) and vagueness (no proper name is given). Indeed, the replacement of names with generally qualifying labels seems to have been a recurrent strategy throughout the history of Western movies: see *The Virginian* (versions of which circulated in 1914, 1923, 1929 and 1946, based on a novel published with the same title in 1902), *The Good, the Bad and the Ugly* (directed by Sergio Leone in 1966), *The Magnificent Seven* (directed by John Sturges in 1960), *The Hateful Eight* (directed by Quentin Tarantino in 2015) and *Revenant* (directed by Alejandro González Iñárritu also in 2015).

Throughout Dixon's book, however, only the English names are employed, possibly to facilitate reading, but definitely showing little concern for native identity – see Dossena (2015).

Following a similar line of reasoning, slaves were also attributed Anglo-European names, both to facilitate their owners' usage and to elide their identity. Examples can be seen both in the slave manifests listing the people that ships were carrying, in which names are provided so as to prove that they had not been imported into the US illegally,[11] and in adverts concerning fugitive slaves.[12] In both cases only first names are provided, often in their shortened, colloquial form – see Examples (2) to (5) (original emphasis).

(2) A NEGRO WENCH, named **Cloe**, about thirty years old, pretty stout made, but not tall; speaks English and French, the latter not fluently. […]. Berthier, 21st July 1791. J. Joseph, *Quebec Gazette*, 28 July 1791 (https://earlycanadianhistory.ca/2016/02/29/canadian-fugitive-slave-advertisements-an-untapped-archive-of-resistance/)

(3) Oct. 10, 1767. Prince George, Sept. 28, 1767. RUN away from the subscriber, the 22d of this instant, three slaves, viz. **JUPITER, alias GIBB**, a Negro fellow, about 35 years of age, about 6 feet high, knock kneed, flat footed, the right knee bent in more than the left, has several scars on his back from a severe whipping he lately had at Sussex court-house, having been tried there for stirring up the Negroes to an insurrection, being a great Newlight preacher. **ROBIN** about 25 years of age, a stout fellow, about 6 feet high, has a film over one of his eyes, a sore on one of his shins, and is brother to Gibb. **DINAH**, an old wench, very large, near 6 feet high; she has a remarkable stump of a thumb, occasioned by a whitlow, by which the bones of the first joint came out and is mother to the two fellows.

(4) Oct. 13, 1774. Run away from the Subscriber, last Friday, a likely Virginia born Negro Man called **JOHNNY**, about 22 Years of Age, five Feet eight Inches high, has a down Look, and Waiter.

(5) June 16, 1774. RUN away from the subscriber, about the 20th of April, a Mulatto Fellow named **Peter Brown**, by Trade a Painter, but can do Carpenters Work, about 35 or 40 Years of Age, […]. As he had a Wife at Mr. Benjamin Hubbard's, it is likely he may be lurking in that Neighbourhood, or near Petersburg, where he was raised.
(*Virginia Gazette*, www.history.org/history/teaching/runaway.cfm)

11. See *Slave Manifests for Charleston* at www.archives.gov/atlanta/finding-aids/slave-manifests.

12. See *Colonial Williamsburg* (n.d.), Nelson (2016) and Eddins (2017).

When surnames do appear, as in (5) above, they are stereotypical and generalizing identity markers, which certainly do not correspond to the subject's own cultural identification.

In recent years, genealogists and historians have also paid attention to how and when the names of immigrants to the USA who landed at Castle Garden first and Ellis Island later could be anglicized, so as to facilitate transcription in the registers or because mistakes were made in the registrations (Cannato 2009). However, it appears that name change was a later phenomenon, typically meant to avoid discrimination on the basis of alien status: indeed, an early twentieth-century English grammar for Italian immigrants (Anon. 1905) systematically translates first names in all the model letters it provides (see Dossena 2008).

Although many of these examples may seem of marginal interest, their function is to show that, at least from a Eurocentric point of view, naming and, more in general, labeling has never been a neutral activity. It is a conscious practice which always conveys attitude in relation to the subject under discussion and to the social and cultural context in which the subject is seen to operate, as aptly shown also by McConnell-Ginet (2003) in relation to twentieth-century gender practices.

In Section 3 below these preliminary considerations will be applied to one of the most significant phases in British history: the last attempt on the part of the Stuart family to regain the throne after Queen Anne's death in 1714 and the coronation of George I Hanover as a result of the 1701 Act of Settlement.

3. The case of "The Forty-five": *Young Pretender, Chevalier*, or *Bonnie Prince*?

In the analysis of names, nicknames, and labels outlined in this contribution, the Jacobite risings of 1715 and, even more significantly, of 1745 are very interesting cases in point. Indeed, the second rising has come to be identified with a noun phrase, "the Forty-five", indicating its date, although its final act was the battle of Culloden, fought in Scotland on 16th April 1746.[13] In addition, all the main protagonists are identified by varying labels, depending on which side is looking at them.

First of all, supporters of the House of Hanover identified the Stuart descendants as *pretenders*, thus drawing attention to the supposed illegitimacy of their claim to the British throne.[14] The so-called *Old Pretender*, James Francis Edward

13. The label was apparently created by Sir Walter Scott: according to the *Oxford English Dictionary*, it was first employed in 1832, in *Redgauntlet* II. xi. 247: "Ye have heard of a year they call the *forty-five*."

14. See the entry in the OED for *pretender*, n., 3.a.

Stuart (1688–1766), son of James II, claimed the throne as James III of England and Ireland and James VIII of Scotland, but was defeated in 1715. His attempts were taken up by his son, Charles Edward Stuart (1720–1788), nicknamed *the Young Pretender*, who was also defeated, but who went on to become an iconic figure in Scottish history. For his supporters, who toasted *the king over the water*,[15] he could be *a certain young gentleman* (thus imposing a kind of self-censorship on the name, which was nonetheless obvious to anyone in the know), or *the Young Chevalier* (his father being of course *the Old Chevalier*), using a French noun that evoked the geographical area where the court first lived in exile.[16] However, nowadays the most frequently-occurring label is *Bonnie Prince Charlie* – see Table 1, based on several corpora available in the Brigham Young University suite[17] and on ordinary Google searches.

Table 1. Labels identifying Charles Edward Stuart

Label	**Hits (23/6/18)**	**Frequency**			
	Google	**iWeb**	**NOW**	**GloWbE**	**Wikipedia Corpus**
Bonnie Prince Charlie	619,000	947	335	61	447
Charles Edward Stuart	292,000	323	74	12	467
Young Pretender	216,000	284	256	56	156

As shown in Table 1, it is only in the *Wikipedia Corpus* that *Charles Edward Stuart* is a little more frequent than *Bonnie Prince Charlie*; in all the other (web-based) cases the predominance of the latter label is often striking.

The success of the noun phrase *Bonnie Prince Charlie* is probably due to the fact that various elements contribute to its phonaesthetic quality: symmetry in word stress (the first and third items are stressed on the first syllable, while the second item is monosyllabic); the hypocoristic suffix in the proper name; and the familiar, not without emotional connotations, overtly Scots qualifier *bonnie*.

However, diminutive forms could be used sarcastically, as in the Jacobite song which referred to George I Hanover as *The wee, wee German Lairdie* (see

15. This indicated the exiled Stuart dynasty, whose court was at Saint-Germain-en-Laye, west of Paris, between 1689 and 1718 (Cruickshanks 1995), before moving to Italy.

16. These labels occur in numerous literary works on the Jacobite risings and their aftermath, such as Robert Louis Stevenson's *Kidnapped* (1886) and *Catriona* (1893), which combine fiction with historical reconstruction and have undoubtedly contributed to the creation of long-standing images of the Jacobite world and its opponents.

17. See the full list and descriptions of individual corpora at https://corpus.byu.edu/.

Dossena 1998, 2012 and 2013). Much more strongly disparaging labels were used for William Augustus, son of George II, who led the Hanoverian army at Culloden: his harsh treatment of the defeated side caused him to be called *the Butcher* or even *Stinking Billy*, reversing the connotative polarity of the name of a flower that in England is called *Sweet William*.[18]

Another reason why *Bonnie Prince Charlie* has become such a successful label is that it seems to have originated in a song, *Bannocks o' Barley* (Cromek 1810: 192–193 and Gilchrist 1815: II, 356).[19] The drama of historical events, especially concerning battles, has survived for centuries in popular lore, often thanks to the fact that oral history can rely extensively on song and poetry evoking battles, lamenting losses or exalting victories. Indeed, many Jacobite tunes stress the positive qualities of the defeated party, and the works of Robert Burns (1759–1796), Alasdair McMhaighstir Alasdair (c. 1698–1770), and Carolina Oliphant (Lady Nairne, 1766–1845) have coexisted with less literary compositions which circulated widely as ballads sung and sold in the street.[20]

Among these sources of varying quality and popularity, the idealization of the Jacobite saga is epitomized in what is perhaps the best-known tune to have been associated with the Young Chevalier's escape, *The Skye Boat Song*; it is based on a Gaelic air, but it was first published in 1884 by Harold Boulton and it was actually re-written by Robert Louis Stevenson in his *Songs of Travel*, of 1895. When the two sets of lyrics are compared, the more existential dimension of Stevenson's text emerges quite clearly, as the protagonist does not reflect on the outcome of the battle of Culloden, but on his youth, a fleeting time which nobody can recover.[21] However, this has not prevented Stevenson's text from being adopted (and adapted) to become the title song of *Outlander*, a TV series based on Diana Gabaldon's (1952–) books.[22] In the new version the label identifying the protagonist, *lad*, is

18. This is *Dianthus barbatus*, see www.plant-lore.com/4684/sweet-william/.

19. Note that Johnson (1839: V, 419–420) challenges Cromek's attribution of these lyrics to Robert Burns.

20. The National Library of Scotland has collected and digitized numerous instances of such broadsides; the project is called "The Word on the Street" and as many as thirty-three items are listed under the 'Jacobites' entry in the subject index (see https://digital.nls.uk/broadsides/subject.cfm/key/Jacobites).

21. Both the traditional lyrics and Stevenson's version can be retrieved from the collection of Scottish songs at www.rampantscotland.com/songs/blsongs_index.htm.

22. The series is produced by Starz and first aired in 2014 (see www.starz.com/series/outlander/episodes); as for the historical fiction on which it is based, background information is provided by the author herself at her official website, www.dianagabaldon.com/books/outlander-series/.

changed into its feminine equivalent, *lass*, so as to reflect the main storyline about a time-travelling Englishwoman who finds herself in the 1740s, while on a journey in mid-twentieth-century Scotland,[23] and whose eighteenth-century Scottish husband affectionately calls her *Sassenach*, although in Late Modern times this label could be used in a disparaging sense (see Dossena 2018).

In any case, what has long been seen as a traditional Jacobite song is in fact a literary artefact in which the contribution of popular culture is restricted to the tune, while the lyrics are the product of educated authors. A similar phenomenon is seen in other Jacobite songs,[24] such as *Charlie's my darling*, or *Will ye no come back again*, authored respectively by James Hogg (1770–1835)[25] and Lady Nairne. The creation of these songs and pibroch tunes, many of which sounded like ordinary love ones, but which (often barely) hid a political message, helped that "invention of tradition" (Hobsbawm and Ranger 1983: Titlepage), in which the Jacobite cause was reinterpreted as a Highlands vs. Lowlands, Scotland vs. England, opposition, while in fact loyalties were much more divided.

As for authentic documents, a very good starting point for the investigation of labels is *The Lyon in Mourning* (Forbes 1746–1775 [Paton 1895–1896]). This is a collection which comprises familiar and official letters, dying speeches, journals and accounts, in addition to songs and verse: i.e. text types in which subjectivity and emotional involvement inevitably occur alongside strategies meant to stress factuality and reliability. Although it is presented as a contemporary 'account of events',[26] the compiler's stance in favour of the Jacobite cause is undoubted and

23. The composer, the director and the leading actors discussed this lexical change during an EMMY show in 2016: see www.emmys.com/video/outlander-origins-skye-boat-song; the title sequence itself is available at www.youtube.com/watch?v=ledHVF1ZtDI. As mentioned in the EMMY show by Sam Heughan, whose role is that of the male protagonist, after the battle of Culloden Charles Edward Stuart was smuggled out of the country in women's clothes, so the pronoun change might actually be justified also on historical grounds. While this is certainly true (Charles Edward Stuart went under the name of *Betty Burke*, Flora MacDonald's maid – see Macdonald 1839), what the discussion does not highlight is that the adaptation of the lyrics is in fact also a subtle move of literary appropriation aiming to establish a continuum between Stevenson's narrative and the show's storyline, thus upgrading the latter to a more prestigious level.

24. Useful studies of Jacobite songs are provided by Donaldson (1988) and Pittock (2009).

25. Indeed, in 1817 James Hogg compiled a collection of Jacobite songs (*The Jacobite Relics of Scotland*, published in Edinburgh in 1819) on commission from the Highland Society of London.

26. For a discussion of how fact and argument coexist in the collection see Dossena (2017b).

as many as 44 documents refer to "the Prince" in their titles. Below are a few examples from Volume I:

- Captain O'Neil's Journal of the Prince's retreat
- Journal of Captain Malcolm MacLeod, second cousin to Malcolm MacLeod, Laird of Raaza, as to the Prince's escape, his own sufferings, and some other incidents of the Rebellion
- Remarks and particular sayings of some who were concerned in the Prince's preservation
- A short but genuine account of Prince Charlie's wanderings from Culloden to his meeting with Miss MacDonald, by Edward Burke
- Journal of the Prince's embarkation and arrival, etc., taken from the mouth of Aeneas MacDonald, a banker in Paris, and brother of Kinlochmoidart, when he was in a messenger's custody in London, by Dr. Burton of York

The use of the official title is meant to show both respect and acknowledgement of the actual royal status of the figure under discussion: even to this day, Jacobite websites and publications call members of the Stuart dynasty kings and queens, following Papal usage, although this may be somewhat confusing for users unaccustomed to these lexical choices because of their greater familiarity with British (i.e. anti-Jacobite) labels – see Example (6), which describes the Church of the Holy Apostles in Rome.

(6) It was the parish church of King James III and VIII when he lived in the nearby Palazzo Muti (now Palazzo Balestra). The body of Queen Clementina lay in state here before her funeral in the Basilica di San Pietro.
(www.jacobite.ca/gazetteer/Rome/SSXIIApostoli.htm)

The coexistence of varying and often biased labels can thus be shown to have considerable impact on usage in present-day public discourse, especially as far the tourist and entertainment industries are concerned. This phenomenon will be discussed further in Section 4 below.

4. Labels, bias, and public discourse

From the analysis that I have conducted so far, it appears that the most familiar label in use today is probably *Bonnie Prince Charlie*. Although it does not quite match earlier labels, but rather conflates them in its invariable form, it has become a set phrase for huge numbers of people both within and outside Scotland, regardless of their knowledge of Scottish history. In practice, it has become a very useful short-hand term in the popularization of historical events, to the point that an exhibition held at the National Museum of Scotland in Edinburgh between 23 June and 12 November 2017 actually used it in its title, *Bonnie Prince Charlie and the Jacobites*.[27] In this respect, the (mis)representation of people and events in public discourse is a very interesting phenomenon, as it may occur in situations in which the memorability of specific uses may be such that the labels get accepted as reliable items in spite of any bias they may convey. This is the case, for instance, of tourist trails, in which the visitors' experience 'in the footsteps of' one or the other historical or literary figure is enhanced through the presentation of what is offered as accurate background information.

There is no doubt that, like all fraught issues, Jacobitism has been represented in highly polarised ways, but with time an increasingly romanticized image has taken shape (see Clyde 1994).[28] This view is certainly related to the success of the numerous literary works authored by Sir Walter Scott, James Hogg and Robert Louis Stevenson long after the Jacobites' final defeat. Nowadays fiction is supplemented with the extraordinary means made available by film and TV adaptations of texts, as the visual element takes audiences into the story itself, especially when filming is on location or in places that are meant to reflect the expectations of both authors and audiences. In addition, authenticity can be conveyed through skilful use of accents, and empathetic participation can be elicited through an evocative soundtrack. This may lead to such a suspension of disbelief that viewers can actually be encouraged to visit the places where the action is supposed to take place, regardless of historical accuracy.

The latest addition to these attempts to re-create history through the eyes of romance is in *Outlander* trails, which originate in the success of the TV series I mentioned in Section 3 above. Agencies organize guided tours of filming locations, which can also be visited following a map on the website of the Scottish Tourist Board.[29] In the boxes outlining what the various places represent in the series, the conflation of fact and fiction is undoubted; in Examples (7) and (8) the first names of the protagonists (Claire and Jamie) are employed, signaling their familiarity for

27. See www.nms.ac.uk/national-museum-of-scotland/whats-on/bonnie-prince-charlie-and-the-jacobites/.

28. In relation to the romanticization of Scotland, Calloway (2008) has compared the predicament and representation of Highlanders with that of Native Americans: in both cases their defeat underpinned cases of 'imperialist nostalgia' that led to the idealisation of a lost past in which former foes were depicted as brave warriors belonging to a world that in the meantime had become irretrievable – see also Cartosio (2016 and 2018). On the progression from 'Scotophobia' to 'Celtomania' see Sorensen (2000).

29. See www.visitscotland.com/see-do/attractions/tv-film/outlander/.

readers, and the inevitable reference to *Bonnie Prince Charlie* also occurs, so as to enhance the readers' involvement through its immediate recognizability.

(7) The charming Glencorse Old Kirk near Edinburgh, nestled in the grounds of the Glencorse House in the foothills of the picturesque Pentland Hills, is where Claire and Jamie tie the knot. With its lush parkland, lake and splendid garden, it's a wonderfully atmospheric venue, if not one of the most romantic wedding venues imaginable for an Outlander fan.

(8) This magnificent 17th century castle [Drumlanrig Castle] boasts over 40 acres of beautiful gardens set near Thornhill in Dumfries and Galloway. The exterior, living rooms and bedrooms became Bellhurst Manor, including one bedroom once slept in by Bonnie Prince Charlie on his way north to Culloden. Explore the Renaissance castle to see spectacular collections of silver, French furniture and art.

Nor is this phenomenon quite new, although it can unfold in somewhat different ways. In Chapter 17 of Stevenson's novel *Kidnapped*, we find the murder of *the Red Fox*, i.e. Colin Campbell of Glenure, who owed his nickname to the colour of his hair and his sly personality. Far from being a diversion in the main storyline, this captivates present-day readers even more because it is still an unsolved mystery,[30] and hikers can follow in the footsteps of Alan Breck and David Balfour, i.e. the two literary protagonists, in an intriguing blend of fact and fiction.[31]

Also American literature and popular culture have re-invented the story of real-life characters like *Billy the Kid* in countless dime novels, actual novels and, later, films (see Rosso 2016). Similarly to what happens in Scotland, nowadays "Travelers can track the Old West's most notorious legends" (Boardman 2015), according to the subtitle of an article meant to outline the possibilities of heritage travel in the USA. On these itineraries participants are invited to enjoy a vicarious experience of what is supposed to have been authentic, and the identification of present-day tourists with the well-known (and clearly labelled) outlaws of the past is evoked through verbs that indicate the equivalence of actions. An example of this is in the exhortation "Let's mount up and hit the trails", in Boardman's presentation of the Outlaw Trail (2015: homepage), as if twenty-first century visitors still travelled on horseback like nineteenth-century adventurers. In the same style, readers come across the frequent expression "Put your boots up" as the various points on the trail are outlined. This introduces paragraphs providing suggestions

30. On the so-called 'Appin Murder', named after the area in which it occurred, see MacGrigor (2002), Gibson (2003), Nimmo (2005), and Dossena (2010).

31. Details on the itinerary and other additional information can be found at www.stevenson-way.org.uk/.

for accommodation, food and entertainment, and again a certain image of the visitor is implicitly evoked: a (presumably male) boot-wearing subject with an interest in food, alcohol and historical venues like jails and graveyards. The examples from Boardman (2015) concern the well-known gunfight at the O. K. Corral (9) and Pat Garrett and Billy the Kid (10), people and events that are supposed to be well-known to contemporary audiences thanks again, first and foremost, to western films.[32]

(9) **EARPS AND McLAURYS: FROM IOWA TO TOMBSTONE** […].
The stock-yards museum features numerous exhibits on the history of the city "Where the West Begins." *FortWorthStockyards.org*
And Tombstone? Well, the entire town is history, isn't it? Including Boot Hill, where Tom and Frank McLaury (and fellow combatant Billy Clanton) are buried. *BoothillGraves.com*
Put Your Boots Up
The legendary Bird Cage Theater opened its doors on Christmas 1881 – right in the middle of the Cowboy/Earp problems. The place provided music and dancing and booze and more. It's still open for visitors, a showcase of Tombstone history and legend. *TombstoneBirdcage.com*

(10) **BILLY AND PAT: FROM LINCOLN TO FORT SUMNER** […]
In mid-July, Sheriff Pat Garrett and two deputies made their way to Fort Sumner. The final act, in the bedroom of Pete Maxwell, was something of an accident. Billy and Garrett never expected to run into each other there. And Garrett initially wasn't sure it was the Kid who'd taken the fatal bullet. Garrett left the place in triumph. Billy stayed forever. Both became legends.
Off the Trail
Lincoln (60 miles west of Roswell) is one of the most authentic Old West towns anyplace. Check out the old courthouse where the Kid was held and killed the two jailers. *LincolnCountyNM.org*
And at Fort Sumner (158 miles southeast of Albuquerque), the Old Fort Sumner Museum has a number of Billy-related items. And just behind it is the fenced-in gravestone of Billy and two of his gang members. *FortSumnerChamber.com*

32. *Gunfight at the O. K. Corral*, starring Burt Lancaster and Kirk Douglas, was directed by John Sturges in 1957, but other films also represented the story, such as *Wyatt Earp*, with Kevin Costner in the title role and directed by Lawrence Kasdan in 1994. *Pat Garrett and Billy the Kid*, starring James Coburn and Kris Kristofferson, was directed by Sam Peckinpah in 1973; the cast also included Bob Dylan, who authored the soundtrack and included a song that would become a hit in its own right, *Knockin' on Heaven's Door*.

> **Put Your Boots Up**
> Just across the street from the old courthouse is the Wortley Hotel. Deputy Bob Olinger was eating there when Billy the Kid made his escape – it was Olinger's last meal, since the Kid killed him in the process of breaking out. It is still a wonderful hotel and restaurant. *WortleyHotel.com*

In both cases the coexistence of 'legend' and authenticity is pervasive in the text: as Groves (2018) suggests, readers should "Walk into an Old West hotel or saloon and let history embrace you. […]. Step in and relive the past."

The current success of trails in the modern tourist industry appears to derive both from the appeal of (mostly) literary works and from the ways in which they enable non-expert visitors to embark on a journey of discovery between fact and (more or less accurate) reconstruction.[33] It is as if visitors, authors and their creations could share an experience made more accessible thanks to the relatively simplified overviews that familiar labels convey.

5. Concluding remarks

Nowadays Italian schoolchildren no longer study Garibaldi as the 'Hero of the Two Worlds', but tourists from all over the world learn of the lad who sailed "over the sea to Skye". Heritage trails, museums, and materials aimed at the general public build on a distinctive fascination with a romanticized past to allure and inform at the same time. In this respect, labels and their undeniable evaluative function (as outlined in Section 2 above) may contribute to a certain distortion of the way in which the past is put on display and understood. Indeed, popular, appealing labels may be manipulative: far from favouring knowledge, they may contribute to the crystallization of relatively biased views. It is therefore important that the mechanisms underpinning labels should be understood, and greater awareness is required of their potential pragmatic strength.

33. The present-day tourist industry offers countless trails both in Britain and in the US: as far as literature is concerned, in addition to the trails we have already mentioned, we could cite the Bronte Way in Yorkshire (see www.bronte-country.com/bronte-way.html); as for non-literary figures who have inspired trails on both sides of the Atlantic, the most significant case is probably that of John Muir: in Scotland the John Muir Way is a coast-to-coast trail running from Helensburgh, from where the Muir family emigrated to the United States, to Dunbar, where Muir was born (see http://johnmuirway.org/); in California, instead, there is the John Muir Trail, which starts in Yosemite National Park, runs through the Ansel Adams Wilderness, Sequoia National Park, King's Canyon National Park, and ends in the Sierra Nevada (see www.pcta.org/discover-the-trail/john-muir-trail/).

Such mechanisms can range from the use of nicknames to pseudonyms to familiarizing forms, or they may rely on more elliptical, but equally recognizable, emblematic forms, like in the phrase <Definite article> + <noun> + <of> + <toponym>. However, what they all have in common is that recipients are expected to acknowledge and understand them, thus implicitly sympathizing with the positive or negative stance that they convey. On the other hand, frequent usage may result in desemanticization, i.e. in an increasing opacity of the evaluative quality of the label: consequently, over time users can assume it is a 'standard' label, and fail to perceive its intrinsic subjectivity. Of course evaluations can be shared, but – like everything else in pragmatics, or indeed in life – their validity as identity markers should not be taken for granted.

Acknowledgements

This study has benefited from research I conducted within a project partly funded by the Italian Ministry of Education (PRIN grant prot. No. 2015TJ8ZAS).

References

Primary sources

Bannister, John. 1869–1871. *A Glossary of Cornish Names*. Truro: Netherton.

Colonial Williamsburg, n.d. Available online at www.history.org/history/teaching/runaway.cfm (accessed 15 April 2019).

Cromek, Robert H. 1810. *Remains of Nithsdale and Galloway Song*. London: Printed for T. Cadell.

Dixon, Joseph K. 1913. *The Vanishing Race*. Garden City, NY: Doubleday, Page & Co.

Eddins, Crystal. 2017. On the Lives of Fugitives: Runaway Slave Advertisement Databases. Available online at www.hastac.org/blogs/crystal-eddins/2017/03/30/lives-fugitives-runaway-slave-advertisement-databases (accessed 15 April 2019).

Ellwood, Rev. Thomas. 1895. *Lakeland and Iceland; Being a Glossary of Words in the Dialect of Cumberland, Westmoreland and North Lancashire Which Seem Allied to or Identical with the Icelandic or Norse, Together with Cognate Place-names and Surnames, and a Supplement of Words used in Shepherding, Folk-lore and Antiquities*. London: Published for the English Dialect Society by Henry Frowde.

Forbes, Robert (compiler). 1746–1775. *The Lyon in Mourning*, ed. by Henry Paton for the Scottish History Society (1895–1896). Available online at http://digital.nls.uk/print/transcriptions/index.html (accessed 15 April 2019).

Gilchrist, John. 1815. *A Collection of Ancient and Modern Scottish Ballads, Tales and Songs*. Edinburgh: Printed for William Blackwood.

Johnson, James. 1839. *The Scottish Musical Museum*. Edinburgh: Printed for William Blackwood. Available online at https://digital.nls.uk/special-collections-of-printed-music/archive/91519813 (accessed 15 April 2019).

Macdonald, Alexander. 1839. *An Interesting and Faithful Narrative of the Wanderings of Prince Charles Stuart and Miss Flora Macdonald, after the Battle of Culloden. With a Memoir of his Life, and Several Jacobite Poems and Songs*. Glasgow, n.p.

Nelson, Charmaine A. 2016. Canadian Fugitive Slave Advertisements: An Untapped Archive of Resistance. Available online at https://earlycanadianhistory.ca/2016/02/29/canadian-fugitive-slave-advertisements-an-untapped-archive-of-resistance/ (accessed 15 April 2019).

Oxford English Dictionary. Second edition, online. Oxford: Oxford University Press. http://www.oed.com

Salamanca Corpus: Digital Archive of English Dialect Texts. www.thesalamancacorpus.com/index.html (accessed 15 April 2019).

Scottish National Dictionary. www.dsl.ac.uk

Sellar, Walter C., and Robert J. Yeatman. 1930. *1066 and All That: A Memorable History of England*. London: Methuen & Co.

Simpson, Alexander G. 1903. *The Life of a Miner in Two Hemispheres*. New York: The Abbey Press.

Slave Manifests for Charleston. www.archives.gov/atlanta/finding-aids/slave-manifests (accessed 15 April 2019).

Stevenson, Robert Louis. 1886. *Kidnapped*. London: Cassell.

Stevenson, Robert Louis. 1889. *The Master of Ballantrae*. London: Cassell.

Stevenson, Robert Louis. 1895. *Songs of Travel*. London: Chatto & Windus.

Secondary sources

Abel, Ernest L. 2004. "Nicknames of American Civil War Generals." *Names* 52 (4): 243–285. https://doi.org/10.1179/nam.2004.52.4.243

Ainiala, Terhi, and Jan-Ola Östman (eds). 2017. *Socio-Onomastics. The Pragmatics of Names*. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.275

Boardman, Mark. 2015. "The Outlaw Trail." *True West: History of the American Frontier*. Available online at https://truewestmagazine.com/the-outlaw-trail/ (accessed 15 April 2019).

Calloway, Colin G. 2008. *White People, Indians, and Highlanders. Tribal Peoples and Colonial Encounters in Scotland and America*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195340129.001.0001

Cannato, Vincent J. 2009. *American Passage: The History of Ellis Island*. New York: Harper.

Cartosio, Bruno (a cura di). 2007. *Wobbly! L'Industrial Workers of the World e il suo tempo*. Milano: Shake.

Cartosio, Bruno. 2016. "American Artists Look West." In *Knowledge Dissemination in the Long Nineteenth Century: European and Transatlantic Perspectives*, ed. by Marina Dossena, and Stefano Rosso, 9–26. Newcastle u.T.: Cambridge Scholars.

Cartosio, Bruno. 2018. *Verso Ovest: Storia e mitologia del Far West*. Milan: Feltrinelli.

Clyde, Robert. 1994. *From Rebel to Hero: The Image of the Highlander, 1745–1830*. Edinburgh: John Donald.

Cruickshanks, Eveline. 1995. *The Stuart Court in Exile and the Jacobites*. Hambledon: Continuum.

Donaldson, William. 1988. *The Jacobite Song. Political Myth and National Identity*. Aberdeen: Aberdeen University Press.

Dossena, Marina. 1998. "Diminutives in Scottish Standard English: A Case for Comparative Linguistics?" *Scottish Language* 17: 22–39.

Dossena, Marina. 2008. "Prescriptivism a Century Ago: Business Correspondence Taught to Emigrants – A Case Study." In *Perspectives on Prescriptivism*, ed. by Joan Beal, Carmela Nocera, and Massimo Sturiale, 41–58. Bern: Peter Lang.

Dossena, Marina. 2010. "Reading with Murder: Stevenson's *Kidnapped* and the Popularisation of Scottish History." In *The Knowledge of Literature IX. Literature and History*, ed. by Angela Locatelli, 57–76. Bergamo: Bergamo University Press.

Dossena, Marina. 2012. "Vocative and Diminutive Forms in Robert Louis Stevenson's Fiction: A Corpus-based Study." Special issue on A New Approach to Literature: Corpus Linguistics, ed. by Ángela Almela Sánchez Lafuente, and Irina Keshabyan. *International Journal of English Studies* 12 (2): 1–17. https://doi.org/10.6018/ijes/2012/2/161721

Dossena, Marina. 2013. "'And Scotland will march again'. The Language of Political Song in 19th- and 20th-century Scotland." In *After the Storm: Papers from the Forum for Research on the Languages of Scotland and Ulster triennial meeting, Aberdeen 2012*, ed. by Janet Cruickshank, and Robert McColl Millar, 141–165. Aberdeen: Forum for Research on the Languages of Scotland and Ireland.

Dossena, Marina. 2015. "(Re)constructed Eloquence. Rhetorical and Pragmatic Strategies in the Speeches of Native Americans as Reported by Nineteenth-century Commentators." *Brno Studies in English* 41 (1): 5–28. https://doi.org/10.5817/BSE2015-1-1

Dossena, Marina. 2017a. "Mrs Malaprop Goes to Hastings: History, Parody, and Language in *1066 and All That* (1930)." Special issue on Errors: Communication and its Discontents, ed. by Paolo Caponi, Giovanni Iamartino, and David Newbold. *Altre Modernità*: 219–232.

Dossena, Marina. 2017b. "'Tis (almost) Three Centuries Since. Linguistic Explorations in *The Lyon in Mourning*." In *Before the Storm: Papers from the Forum for Research on the Languages of Scotland and Ulster triennial meeting, Ayr 2015*, ed. by Janet Cruickshank, and Robert McColl Millar, 127–144. Aberdeen: Forum for Research on the Languages of Scotland and Ireland.

Dossena, Marina. 2018. "'Sassenach', eh? Late Modern Scottish English on the Borders of Time and Space." *Token: A Journal of English Linguistics* 7: 53–76.

Gibson, Rosemary M. 2003. *The Appin Murder: In Their Own Words*. Edinburgh: National Archives of Scotland.

Groves, Melody. 2018. "History, Heritage, and Hospitality." *True West: History of the American Frontier*. Available online at https://truewestmagazine.com/history-heritage-hospitality/ (accessed 15 April 2019).

Hobsbawm, Eric, and Terence Ranger (eds). 1983. *The Invention of Tradition*. Cambridge: Cambridge University Press.

MacGrigor, Mary. 2002. *Grass will not Grow on my Grave: The Story of the Appin Murder*. Dalkeith: Scottish Cultural Press.

McConnell-Ginet, Sally. 2003. "'What's in a Name?' Social Labeling and Gender Practices". In *The Handbook of Language and Gender*, ed. by Janet Holmes, and Miriam Meyerhoff, 69–97. Hoboken, NJ: Wiley-Blackwell. https://doi.org/10.1002/9780470756942.ch3

Nevala, Minna. 2016. "Solidarity in Evaluation: The Portrayal of Murderers and Their Victims in Late Nineteenth-century Newspapers." In *The Pragmatics and Stylistics of Identity Construction and Characterisation* (Studies in Variation, Contacts and Change in English

17), ed. by Minna Nevala, Ursula Lutzky, Gabriella Mazzon, and Carla Suhr. Helsinki: VARIENG. Available online at: www.helsinki.fi/varieng/series/volumes/17/nevala/ (accessed 15 April 2019).

Nevala, Minna. 2017. "The Public Identity of Jack the Ripper in Late Nineteenth-century British Newspapers." In *Diachronic Developments in English News Discourse*, ed. by Minna Palander-Collin, Maura Ratia, and Irma Taavitsainen, 199–216. Amsterdam: John Benjamins. https://doi.org/10.1075/ahs.6.11nev

Nimmo, Ian. 2005. *Walking with Murder. On the Kidnapped Trail*. Edinburgh: Birlinn.

Olsen, Susan. 2006. "Nicknames and Remembrance: Memorials to Woodlawn's Jazz Greats." *Names* 54 (2): 103–120. https://doi.org/10.1179/nam.2006.54.2.103

Oswald, F. L. 1895. "Historical Nicknames." *The North American Review* 161 (465), Aug. 1895: 254–256.

Pittock, Murray G. H. 2009. *The Myth of the Jacobite Clans: The Jacobite Army in 1745*. Edinburgh: Edinburgh University Press. https://doi.org/10.3366/edinburgh/9780748627561.001.0001

Rosso, Stefano. 2016. "The Winning of the Western: Early Dissemination of a Literary Genre." In *Knowledge Dissemination in the Long Nineteenth Century: European and Transatlantic Perspectives*, ed. by Marina Dossena, and Stefano Rosso, 27–44. Newcastle u.T.: Cambridge Scholars.

Sauers, Richard A. 2006. "Wall Street's Gilded Age: Nicknames of Her Capitalists and Robber Barons." *Names* 54 (2): 147–172. https://doi.org/10.1179/nam.2006.54.2.147

Sorensen, Janet. 2000. *The Grammar of Empire in Eighteenth-Century British Writing*. Cambridge: Cambridge University Press.

Van Dijk, Teun A. 1995. "Discourse Analysis as Ideology Analysis." In *Language and Peace*, ed. by Christina Schäffner, and Anita L. Wenden, 17–33. Aldershot: Dartmouth.

Wodak, Ruth, and Michael Meyer (eds). 2001. *Methods for Critical Discourse Analysis*. London: Sage. https://doi.org/10.4135/9780857028020

“Right trusty and well-beloved”

The socio-pragmatics of gender, power and stance in sixteenth-century English letters

Imogen Marcus and Mel Evans
Edge Hill University / University of Leicester

The paper investigates how high-ranking early modern women letter-writers negotiated the social tension between their subordinate position as women, and their positions of power, in the linguistic construction of their epistolary identities. We focus on the letters of four Tudor women: Kathryn Parr, Mary I, Elizabeth I and Elizabeth Talbot. We consider how each woman's use of self-reference compares with contemporary epistolary theory and practice, interpreted using the concept of stance which offers a way of understanding how indexical meaning arises from the convergence of multiple linguistic forms in particular contexts. Our analysis finds that scribal status and recipient rank shape the women's linguistic choices, and that gender is a significant, but not the only, dimension of their epistolary identities.

Keywords: stance, gender, early modern women, correspondence, self-reference, address forms, pronouns, register

1. Introduction

Whilst theoretically distinct, the differentiation between private (individual, sentimental, self-interest) and public (community, rationality, solidarity) discourse is often more complex in application (Gal 2002). In the early modern period (1500–1750), letters – the data source for this paper – represent one such case. For early modern letter-writers, private was understood to mean ‘secret’ or ‘familial’ (Akkerman 2018: 12). Whilst the communication structure was prototypically dyadic (i.e. between writer and the named recipient) – letters were used in a range of official, professional contexts, with the processes of composition, dispatch and reception fundamentally collaborative. Letters were written by scribes and a

https://doi.org/10.1075/pbns.306.03mar

named-author, typically read aloud upon reception, and copies made and disseminated within the social networks of both writer and recipient (Daybell 2012: 175–176). Thus, 'private' familial letters were also inherently public, in terms of the range of parties involved; messages were also vulnerable and insecure, subject to third-party eyes and ears (Daybell 2012: 110).

Within this private-public context, an early modern letter-writer makes a written declaration of their experience of the world, transforming their internal emotions or beliefs into a linguistic expression to be shared with third-parties and preserved on the page. Communicative acts of this kind therefore have implications for that writer's persona, from the viewpoint that identity is an intersubjective negotiation between interlocutors (Bucholtz and Hall 2005). How a letter-writer signals their evaluation and understanding of the world (e.g. their subject-matter, their recipient), that is, their stance, shapes their identity-in-interaction.

This paper examines the construction of identity in sixteenth-century correspondence using stance as an analytic framework (Kiesling 2009). Stance denotes the ways in which individuals use language and other semiotic resources to demonstrate their identity with regards to that of their addressee, subject-matter and wider context of communication (Du Bois 2007; Kiesling 2009). As a concept, stance has proved highly productive in sociolinguistics and pragmatics. It enables analysts to focus on the construction of identity-in-interaction as achieved through diverse linguistic forms, avoiding an over-simplistic correlative reading between language and monolithic social categories (e.g. age, gender). Kiesling goes so far to suggest that stance is the "primitive" of the relationship between language and social meaning. Repetitive acts of stance-taking lead to the association of linguistic forms with social group identities (Kiesling 2009: 172). Stance, in this view, offers the potential to identify the intersection between "the everyday use of language variation in discourse" (that is, the micro-level) and "patterns on larger social scales", including those of language change (the macro-level) (Kiesling 2009: 172), an approach that has been productive in historical as well as present-day sociolinguistic and pragmatic work (e.g. Nevalainen and Raumolin-Brunberg 2003; Moore and Podesva 2009; Nurmi et al. 2009).

Stance is a profitable concept for exploring historical, written communication (see Fitzmaurice 2004; Dossena 2006; Dossena and Fitzmaurice 2006). Our chapter considers how language and material features of early modern correspondence signal stance, in a genre that was characterised by formulaic (and prescribed) language elements, and assesses how these features may index higher-level social meanings, linked to (meso-level) personas or (macro-level) social groups (Moore and Podesva 2009). The study focuses in particular on how stance-taking in epistolary contexts negotiated, and related to, female conventions of language use and femininity.

The study analyses the letters of four high-status sixteenth-century women: Queen Katherine Parr, Queen Mary I, Queen Elizabeth I and Elizabeth Talbot, Countess of Shrewsbury (known as Bess of Hardwick). Each woman occupied an exceptional position within the patriarchal structures of their society, holding institutionalised positions of power that yielded financial, material and legal rights, which contrasted with the subordinate position of their sex; a situation perhaps most recognisable in relation to the "Virgin Queen", Elizabeth I.

Whilst style is a matrix of linguistic features (Moore and Podesva 2009: 448), self-reference forms are a key resource "where the identity and relational functions of language become visible" (Palander-Collin 2011: 84) and are explicitly bound up with stance, as such expressions "overtly position the writer and express his or her point of view" (Palander-Collin 2011: 84). Direct address (e.g. second-person pronouns, nominals), also provide a valuable way-in to investigate the subjectivity and relationality of the language choices in women's epistolary communication. The analysis concentrates on the distribution and co-textual environment of self-reference and address forms (see Mühlhäusler and Harré 1990 for foundational work in this area) and evaluates how such features mark epistemic and affective stance, assessing their role in the construal of their identity as high-status, exceptional women. We argue that the demonstration of epistolary stance relies on the circularity of indexicality (micro- to macro-level meanings) (Snell 2010), which encompasses not only linguistic forms but also material facets of a letter (cf. Barnes 2018).

2. Background

2.1 Stance and subjectivity

Stance encompasses "the linguistic mechanisms that convey a speaker or writer's personal attitudes and assessments" (Gray and Biber 2012: 15), and is typically split into two categories: epistemic stance (belief or knowledge e.g. *I know, I don't understand*) and affective stance (emotional orientation e.g. *I am so pleased*) (Du Bois 2007; Kiesling 2009). Within present-day sociolinguistics, stance is used to explain how linguistic forms acquire social meanings. Stance-taking draws on the direct indexicality of linguistic forms and their connection with particular interactive objectives. Repetitive and/or high profile use of these forms results in subsequent associations between particular groups and stance repertoires; that is, the forms come to indirectly index a social identity (e.g. Ochs 1992). For example, hedges (e.g. *kinda*) are used by speakers to signal their epistemic stance (uncertainty, caution) towards a topic and/or addressee, but have come to indirectly index femininity due to wider social structures and ideologies, even though empirical evidence of their typicality in women's language is mixed (Leaper and Robnett 2011). Indexicality does not simply evoke monolithic social categories (e.g. male or female), but has more nuanced realisations within a community (cf. e.g. Duszak 2002), with the relationship between linguistic form, stance-taking and social categories argued to be a circular process (see Eckert 2000; Moore and Podesva 2009; Snell 2010).

Studies of stance in written discourse have tended to focus on linguistic constructions such as adverbials, complement clause constructions and modal verbs (Gray and Biber 2012: 19), perhaps because these forms are transparent and searchable using corpus approaches (Gray and Biber 2012: 20). In multi-dimensional analyses (Biber and Finegan 1989), present-day letters showed a greater proportion of properties associated with affective marking, in contrast with, say, academic discourse. In general, stance is more prominent in spoken-language genres, and those with persuasive and interpersonal (rather than informational) functions (Gray and Biber 2012: 24).

For earlier periods of language use, the approach provides an important perspective for our understanding of the relationship between micro-level activities of individuals and macro-level processes of language change in private and public discourses. Studies of stance in historical texts have considered which stance-marking features characterise a particular genre/time-period, and/or how these features have developed over time. Dossena's (2006) survey of stance in nineteenth-century business correspondence finds "that the perceived social status of encoders and recipients appears to be a far stronger conditioning factor than age, gender, and actual status" (Dossena 2006: 252). The perceived relationship between writer and addressee shapes the use of address formulae, the construction of directives (with varying degrees of mitigation), epistemic modality, and self-reference strategies. Of the latter, Dossena notes that some letters shift to the first-person plural suggesting an early notion of 'corporate identity', but the singular form remains dominant when marking the identity, authority and agency of the writer. In fifteenth-century medical charms, Alonso-Almeida identifies multiple features that mark "the presence of the writers in their texts, as well as [...] their positions of power in relation to their audience" (Alonso-Almeida 2009: 14). Lexical words within the field of religion, e.g. *honour, love*, carry "a heavy semantic load of affectiveness", as they convey the writers' beliefs and social position towards a proposition (such as a man sick with 'evel' fever). These occur alongside more prototypical grammatical stance markers such as adverbials and modality. Alonso-Almeida speculates that the rarity of stance adverbials in the charm corpus illustrates the shift of the charm towards a "more formal written tradition" in the Middle English period (Alonso-Almeida 2009: 14). Fitzmaurice (2004) explores the functional development of stance-marking formulae in Late Modern English correspondence and drama. She observes that stance expressions can acquire, firstly, intersubjective, and secondly, interactive (discourse marking) functions over time using evidence from the ARCHER corpus and focussing on the verb formulae *you know, you say* and *you see*. Within correspondence, many of these forms appear motivated as much by conventions of the genre as they do by the specific interpersonal requirements of a given interaction (Fitzmaurice 2004: 436).

2.2 Historical correspondence

Historical correspondence is a rich resource for the study of language variation and change, including stance-marking, due to its status as a dynamic dialogic genre. Whilst all language acts involve stance (Jaffe 2009), the dialogic nature of correspondence entails that stance-taking is a particularly prominent feature (Biber and Finegan 1989). Self-reference is a key resource for stance-marking, with first-person pronouns and nominal labels attuned to the "social and interpersonal dimensions" of the interaction (Palander-Collin 2011: 84), typically in combination with genre-specific formulae. Palander-Collin's (2011) study of early modern gentlemen's letters finds formulaic expressions (e.g. opening salutations) and address terms combined with self-reference are used to signal their position regarding a letter's content and its recipient. Thus, self-reference is "an outcome of larger societal processes and structures and local interactional negotiation that is partly conventional and habitual and partly less than fully conscious" (Palander-Collin 2011: 86). Institutional hierarchies of power pre-exist, and thus pre-configure, epistolary interaction. In early modern society, such hierarchies were arguably more influential, and less negotiable, than in the present-day era of "conversationalization" (Fairclough 1994).

As noted in the Introduction, early modern letters share features of private and public discourse. Their subject-matter can be intimate and personal, perhaps shared or relevant to only the named interlocutors, and efforts were made to keep letters secret and secure, such as through seals and folds, and linguistic ciphers (see Dambrogio and Starza-Smith 2017). Yet, the processes of consumption and reception involved parties outside the named author and recipient. At the royal court – relevant to three of the four women analysed in this chapter – letters sent to the monarch would usually be read first by their secretary, and then (in the case of scribal letters) read aloud to the king or queen, in the presence of other members of the court (Allinson 2012). Similar practices operated within noble households; the servants and secretaries of Bess of Hardwick, for example, assisted their mistress in the reading of her correspondence (Wiggins 2017: 94). Any writer of a letter would be aware of its multiple, public audience, and therefore the need to attend to the social relationship between writer and recipient, in order to present an epistolary identity that would best ensure the letter's objectives were met.

The social significance of identity construction in correspondence, and the integral role of language in the process, is reflected in the prescriptive pressures circulating around epistolary discourse. The forms of communication were stipulated, either explicitly through didactic manuals such as Angel Day's *The English Secretorie* (1592) and Fulwood's *Enemie of Idleness* (1568), or more implicitly through exposure to contemporary practices, and combined medieval systems, e.g. *ars dictaminis*, attentive to social status, with classical models and Humanist sensibilities (Magnusson 1999: 61–62). Letters were categorised into types (e.g. demonstrative, familiar) and their structural components adjusted accordingly. Manuals reveal the social import of a letter-writer's language choices. Erasmus's *De Conscribendis Epistolis* (1522) stressed the importance for a writer to adjust their language choices according to their relationship with the addressee, and the subject-matter, highlighting particular address and self-reference forms as critical to the success of the epistle. He offers different advice according to whether the addressee is superior, equal, or inferior to the writer, as well as the degree of familiarity within that relationship (Erasmus [1522] 1985); for instance, endorsing the use of "honorary or adoptive names" such as using *son* when writing to "a young man who is dear to us" (Erasmus 1985: 57). Later in the sixteenth century, Fulwood offers similar (derivative) prescriptions in his English-language guide, noting that "[y]f we speake to our inferiour, we must vse a certayne kynde of modest and ciuill authoritie, in giuing them playnely to vnderstand our intent and purpose", avoiding overly ornate or hyperbolic styles (Fulwood 1568: A7v). The manuals stress the interpersonal work of correspondence; that is, how a letter-writer establishes their stance with the potential for indexing higher-level meanings evoked by (co-occurring) linguistic features, such as a learned persona. This invites consideration of the bi-directional relationship between stances, personas and social categories.

The manuals also stress the significance of material choices, and work on epistolary materiality (e.g. Daybell 2012) has observed the social resonance of handwriting and layout. Placing one's signature at the bottom-right corner of the page signalled deference; the top left margin would show the letter-writer's superiority. A significant material feature was whether the letter was written in the author's hand (a holograph text, generally perceived to be more intimate and direct) or that of scribes (typically more formal and socially distant). Meta-pragmatic comments, such as apologies for a scribal letter on personal matters, indicates the interpersonal significance of such choices.

We suggest that letters are a particularly good example of how language combines with the visual and material to produce social meaning. The use of a holograph versus a scribal letter may indicate the letter-writer's attitude towards subject-matter and recipient. Whilst contextual events may conspire to produce a scribal letter when a holograph would have been the norm (for example, Elizabeth

I apologises for scribal letters sent to James VI, due to her ill health), the materiality of the letter would shape a recipient's interpretation of the linguistic contents.

As with other kinds of prescriptive manuals (Yáñez-Bouza 2015), it is difficult to establish a direct causation between the epistolary publications and the letter-writing practices (linguistic, material) of individuals. However, their effects may have been greatest for their target male audience. Day's preface uses the masculine pronoun throughout to denote his reader, who may, if they follow his advice, assume "the office of a secretorie" (a male profession), and states that his book is provided for "the generall and friendlie regard of all men" (Day 1592: 3–4). Fulwood's letter to the reader offers less explicit characterisation of the social demographic, but the verse dedication describes the value of a letter "Vnto our frend, though distance farr […] And talke with him as face to face" (Fulwood 1568: Aiii). As Palander-Collin observes "the normative ideas from manuals express the expectations placed on a gentleman's verbal behaviour" (Palander-Collin 2011: 88).

Whilst this focus on a male audience is unsurprising given how low female literacy was in the early modern period (see Hubbard 2015), women from gentry or aristocratic backgrounds were often literate in the period, with letters playing an important role in their daily lives. It is therefore highly likely that the women of the EWL corpus would have been aware of contemporary epistolary conventions.

However, these women were still setting-up their subjectivity and interpersonal relationships using linguistic and material systems developed primarily for the needs of men. Palander-Collin (2011) focuses her study of self-reference on the letters of male gentry, and finds that self-reference is sensitive to the "private and public roles of the letter writers" (Palander-Collin 2011: 90). The relationship between stance, indexicality and gender has been extensively discussed in twentieth-century and present-day linguistic studies (see e.g. McConnell-Ginet 2011). The historical context of the early modern period presents a valuable counterpoint to this work, providing another perspective on how women negotiate societal pressures appertaining to communicative (epistolary) conventions and the social practices associated with their social roles and their sex.

3. Methodology

Our investigation of self-reference in the letters of 'exceptional' early modern women uses a corpus-based approach combining quantitative and qualitative analysis. Corpora have proved a productive resource for the investigation of stance and style (e.g. Hunston 2007; Palander-Collin 2011), providing valuable evidence of the relationship between specific interactional meanings, and patterns and trends in social practice. Whilst stance is the accumulation of multiple linguistic features, one profitable method for corpus analysis is to identify linguistic features that have a functional association and therefore co-occur with stance (Hunston 2007: 28).

Self-reference forms, namely pronominals and nominal categories, index speaker identity, and are closely bound-up with stance-taking. Our analysis considers how 'exceptional' women letter-writers use self-reference forms, focussing on their frequency, co-textual and collocational environment. The Exceptional Women's Letters corpus (EWL) includes letters by four sixteenth-century women: (Dowager) Queen Katherine Parr (b. 1512–d. 1548), Queen Mary I (b. 1516–d. 1558), Elizabeth Talbot, Countess of Shrewsbury (b. c.1521–d. 1608) and Queen Elizabeth I (b. 1533–d. 1603). These women were selected due to their position at the (near) pinnacle of English society at some point in their lives: Parr, Mary and Elizabeth were part of the royal elite, and Talbot rose from humble beginnings to become a very powerful woman in England, on account of her wealth, property and social connections (Wiggins 2017: 13). Letters for all four women survive from various points in their lives, including when they were at their most institutionally powerful, either as monarch or landowner. The corpus includes both holograph and scribal letters to allow us to explore the potential stance-marking functions of materiality and language.

EWL (see Table 1) is relatively small in scale, although attempts have been made to provide a similar quantity of holograph and scribal material for each woman. There are two caveats. Firstly, there are only holograph letters by Parr; her status and the unusually intimate subject-matter of some of her correspondence (written to her suitor and later husband, Thomas Seymour) justified their inclusion. Secondly, the holograph letters for Mary I date primarily from the years before her accession to the throne in 1553, as few survive (in English) from her years as queen.

The corpus was compiled from various sources. Elizabeth Talbot's letters were obtained from bessofhardwick.org. The royal letters of Parr, Mary and Elizabeth were collected from existing transcripts in the Calendar of State Papers or published editions, and/or transcribed from the original manuscripts. All spelling was normalised using VARD 2.5.4 (Baron 2017) to improve the accuracy of the concordance software AntConc (Anthony 2018).

The corpus is sub-divided by author, materiality (holograph and scribal) and recipient relationship (Tables 2 and 3). The categorisation by recipient relationship sub-divides addressees according to "the variables of power and distance" (Nevala 2004: 78), separated into socially superior, equal and inferior recipients, respectively, although we recognise that this schematization cannot account for the fluidity of interpersonal relations in interaction. Previous socio-pragmatic studies have used similar categorisations in a similar way, with profitable results (e.g. Nevala 2004; Marcus 2017).

Table 1. Corpus details for EWL

Author	Holograph letters: (number; word count (dates))	Scribal letters: (number; word count (dates))	Total letters: (number; word count)
Katherine Parr (b. 1512–d. 1548)	13; 3,640 (1544–1548)	N/A	13; 3,640
Queen Mary (b. 1516–d. 1558)	30; 7,845 (1536–1554)	34; 10,289 (1553–1558)	64; 18,134
Elizabeth Talbot (*alias* Bess of Hardwick) (b. c. 1521–d. 1610)	18; 4,819 (1552–1587)	39; 12,699 (1582–1603)	57; 17,518
Queen Elizabeth I (b. 1533–d. 1603)	95; 34,519 (1560–1603)	60; 35,237 (1559–1600)	155; 69,756
OVERALL	**156; 50,823**	**133; 58,225**	**259; 109,048**

Table 2. EWL by material status and writer contexts

Author	Holograph letters	Scribal letters
Katherine Parr	Letters to superiors, equals and inferiors. Written as Queen, and Queen Dowager; sent to husband King Henry VIII, her fiancé Lord Admiral, Thomas Seymour and female acquaintances of lower rank.	N/A
Queen Mary	Letters to superiors, equals and inferiors. Written as princess/daughter of the king (when excluded from succession); sent to father King Henry VIII, Thomas Cromwell and male/female acquaintances and servants of lower rank.	Letters to inferiors. Written as reigning queen; sent to councillors, nobility and local officials.
Elizabeth Talbot (*alias* Bess of Hardwick)	Letters to superiors, equals and inferiors. Written as wife of Sir William Cavendish (m. 1549–1557), Sir William St Loe (m. 1559–1564/5) and George Talbot, Earl of Shrewsbury (m. 1568, separated in the 1580s). Sent to Queen Elizabeth I, courtiers, noble acquaintances, fourth husband George Talbot and servants.	Letters to superiors, equals and inferiors. Written as wife to George Talbot, Earl of Shrewsbury. Sent to Queen Elizabeth I, courtiers, noble acquaintances, family members and acquaintances of lower rank.
Queen Elizabeth I	Letters to equals and inferiors. Written as reigning queen; sent to other monarchs, favoured courtiers and noblemen and woman.	Letters to equals and inferiors. Written as reigning queen; sent to other monarchs, courtiers, councillors, nobility and local officials.

Table 3. EWL by materiality and recipient-type

Holograph	**Authors represented**	**Word count**
Superiors	Parr, Mary I & Talbot	5,409
Equals	Parr, Mary I, Talbot & Elizabeth I	33,778
Inferiors	Parr, Mary I, Talbot & Elizabeth I	11,636
Scribal	**Authors represented**	**Word count**
Superiors	Talbot	5,181
Equals	Talbot & Elizabeth I	10,711
Inferiors	Mary I, Talbot & Elizabeth I	42,333

The size of the corpus is relatively small, at just over 100,000 words, although this should not be seen as an impediment to quantitative analysis, which should focus on the typicality of an occurrence, not whether the token counts are 'huge' (Hunston 2007: 28). However, Elizabeth I's letters dominate both the holograph and scribal parts of the corpus, and this is taken into account in our analysis. Given her standing as perhaps the most successful, and definitely best known, 'exceptional' woman, her epistolary practices perhaps warrant the extra emphasis.

The results are contextualised through a comparison with male-authored contemporary practices, based on a sub-set of the *Parsed Corpus of Early English Correspondence*. This subset, henceforth PCEEC-16RN, includes all letters in the socially representative corpus written between 1510–1610 by male letter-writers of royal and noble rank. The spelling has again been normalised using VARD 2.0.

The analysis adopts a circular approach, from macro- to micro-level comparisons and back again, analogous with the theoretically circular chain of indexicality (Snell 2010: 650). We identify quantitative differences in self-reference in EWL and PCEEC-16RN, before comparing self-reference between the holograph and scribal correspondence. Finally, our analysis explores the interactional choices according to each woman's relationship with the recipient and the primary communicative function (e.g. reportage, petition) of the letters under investigation. We adopt the term "*primary* communicative function" because early modern letters often have what Palander-Collin calls "mixed purposes" ((2010: 652); see also the discussion of mixed letters in contemporary manuals (e.g. Fulwood 1568: 80–81)), but it is usually possible to identify a primary function that the letter is aiming to fulfil. We conclude by reflecting on the pragmatic significance of self-reference and associated features within the letters, and consider how the stance markers in the letters engage with and/or negotiate their social status and roles.

4. Analysis

4.1 Pronominal distribution: EWL and PCEEC

The frequencies of first- and second-person pronoun forms provide an initial measure of the self-reference trends in EWL and PCEEC-16RN (see Table 4). First-person singular forms occur at similar rates in the two corpora, with occurrences slightly higher in PCEEC (39.4) than EWL (37.8). There is a greater difference between the other pronoun groups. First-person plural forms are far more frequent in EWL than PCEEC, which primarily reflects the linguistic practices of royal scribal correspondence. The women letter-writers also use *you* 1.5 times more often than their male contemporaries. This could reflect register differences: the PCEEC correspondence includes official letters addressed to recipients with whom the senders had more formal, distant relationships, which may also be a factor here, affecting strategies of relationship construction.

Table 4. 1PP frequencies in EWL and PCEEC-16RN

	1PP singular (N.)	Per 1,000 words	1PP plural (N.)	Per 1,000 words	2PP (N.)	Per 1,000 words
EWL	4,118	37.8	2,746	25.2	4,545	41.7
PCEEC-16RN	6,552	39.4	1,163	7.0	4,882	29.4

The similarity in frequency in the first-person singular, which may reflect general generic similarities, masks interesting differences in co-textual environment. Collocate analysis of *I* (R1 position) provides a succinct indication of the immediate contexts of self-reference (Table 5).

Table 5. R1 collocates of I in EWL and PCEEC-16RN

RANK	EWL	Freq	Per 1,000 words	PCEEC-16RN	Freq	Per 1,000 words
1	have	194	1.8	have	412	2.5
2	am	124	1.1	am	232	1.4
3	will	92	0.8	shall	160	1.0
4	shall	74	0.7	will	142	0.9
5	would	67	0.6	would	140	0.8
6	beseech	58	0.5	think	130	0.8
7	pray	54	0.5	trust	102	0.6
8	can	54	0.5	pray	99	0.6
9	may	51	0.5	do	99	0.6

(*continued*)

Table 5. (*continued*)

RANK	EWL	Freq	Per 1,000 words	PCEEC-16RN	Freq	Per 1,000 words
10	do	46	0.4	can	90	0.5
11	trust	45	0.4	know	74	0.4
12	should	41	0.4	had	71	0.4
13	know	38	0.3	may	68	0.4
14	doubt	38	0.3	should	60	0.4
15	think	35	0.3	did	59	0.4
16	hope	35	0.3	was	53	0.3
17	must	34	0.3	doubt	53	0.3
18	had	34	0.3	must	43	0.3
19	was	33	0.3	beseech	43	0.3
20	see	28	0.3	commend	40	0.2

The collocates include verb forms that can be connected to epistemic stance (the expression of belief or understanding). Forms such as *know, doubt, think* and *see* explicitly articulate the writer's comprehension of events. *See* collocates more frequently with *I* in the women's letters. Ranked 20th in EWL, it occurs as the 30th most-frequent collocate in PCEEC-16RN. (The verb is absent from the 3-gram clusters of gentry letters from CEEC discussed in Palander-Collin (2011)). Closer examination reveals that the verb signals both reported and hypothesised literal perception e.g. "shall thank long till I see you both" (Talbot to her children) and metaphorical statements of understanding and comprehension e.g. "but I see out of sight out of mind with your unkind none" (Talbot to her husband). Both assert the relationship between author and recipient, with an emphasis on the affective dimensions. The collocate appears to be less central to the male-authored subject-matter and expressions of comprehension.

Other collocates, such as *trust* and *hope*, signal the writer's desires (also affective stance markers). Notably, these forms often have an interpersonal focus; for example, an expression of desire is oriented towards the recipient's attitude or physical well-being e.g. "all such error I hope you will impute to affection not my curiosity" (Elizabeth I to James VI, 1590). *Hope* ranks higher in EWL than PCEEC-16RN (36th), and its use as a stance marker to frame a request or expression of the writer's wants underlies this difference in frequency.

The collocate list includes several verbs marking verbal processes (Simpson 1993), such as *pray, beseech* and *commend*. These provide a meta-pragmatic frame for the communicative act, and typically co-occur with requests (that is, expressions of desire). *Commend* is one verb more frequent in PCEEC-16RN than EWL (ranked 72nd), used in opening and closing formulas e.g. "My good lord and brother in my heartiest manner I commend me unto you" (Henry Percy, Earl of Northumberland to Henry, Earl of Cumberland, 1528; Clifford_043). These formulae establish the author's potential indebtedness towards the recipient, requesting the latter's "kindly remembrance" (OED Online 2018), or favour. Although wide-spread in male-authored letters, and despite the deferential stance, the formula occurs only twice in EWL; once as a closing expression in a letter from Talbot to Sir Francis Walsingham, and once in the opening of a pre-accessional letter from Mary to Sir Thomas Cromwell. The former letter sees Talbot reporting to Walsingham about Mary Queen of Scots, who was at that time under Talbot's care. She states that her husband George "pronounced the contents of the articles you sent to this queen" and that Mary has grown "lean and sickly". It is therefore a letter of reportage written to a social superior. The letter from Mary to Cromwell is a letter of request, asking Cromwell to show favour to the bearer of the letter in a suit he has put forward to him. In both cases, the relationships between recipients could be classed as relatively formal and distant, similar to the relationships mediated by the (often official) letters penned by male writers in the PCEEC. This could be why the female senders use a phrase more typical of male-authored epistolary language, which may also have indirectly indexed a social persona (professional, official) in which gender, specifically masculinity, was a relevant attribute.

Keyword analysis provides a useful complement to the quantitative overview of pronoun frequencies. Keywords are lexical items occurring at an unusual frequency (here, words with a log-likelihood score > 17.0, equivalent to $p < 0.001$) when compared to their distribution in a reference corpus (in this case, PCEEC-16RN) (Scott and Tribble 2006: 59). The test identified 104 keywords; Table 6 (the first two columns) reproduces the first thirty for EWL (L-L score > 30). A second keyword analysis flips the analysis, and identifies the keywords of PCEEC-16RN using EWL as the reference corpus (columns 3 and 4), to provide an indicative comparison of the male and female-authored correspondence. The discussion first examines the EWL keywords, before turning to those from PCEEC-16RN.

The first-person plural pronouns (*our, we* and *us*) and the second-person pronoun *you* are the top keywords in EWL. Although keywords are usually content words linked to the content and subject-matter, function keywords indicate stylistic characteristics (Scott and Tribble 2006: 60). In the present case, the pronominal forms can be understood to fulfil both of these criteria. The pronouns indicate that subjective and interpersonal reference are prominent aspects of the correspondence; that is, the women letter-writers talk about themselves, and their recipients, extensively. And, stylistically, they use particular forms (1PP plural and 2PP forms) as (one of) the primary mechanisms to achieve this, more so than their male counterparts in PCEEC. Interestingly, *I, me* and *my* do not occur in the keyword list, suggesting that the similarity in normalised frequencies (Table 6) reflects parity in usage across the two corpora.

Table 6. Top 30 keywords by log-likelihood scores in EWL and PCEEC-16RN

EWL keywords	L-L Score	PCEEC-16RN keywords	L-L score
our	596.4	majesty	343.6
we	529.9	ye	272.4
you	475.4	her	239.4
us	191.0	mr	217.6
dear	121.9	the	185.9
elizabeth	119.6	lordship	112.6
not	74.2	lord	109.4
than	67.8	he	109.2
suppose	66.3	majesty’s	100.6
too	65.0	here	89.2
realm	64.8	very	87.6
such	62.9	there	77.4
sister	61.9	k	74.8
brother	57.2	grace	72.7
ever	56.8	lordship’s	72.4
mary	52.8	at	72.2
counsel	51.9	i	68.6
e	51.5	ij	66.6
hardwick	42.1	duke	64.3
care	41.2	grace’s	59.3
just	38.0	written	59.1
so	35.4	men	55.9
happy	35.2	addressed	54.8
whose	34.5	highness	53.8
regard	34.4	hearty	50.6
whom	33.7	emperor	50.1
arbell	33.4	fare	49.5
life	33.2	burghley	48.7
affectionate	32.3	money	45.7
never	32.1	cromwell	45.6

The other EWL keywords indicate the subject-matter of the women's correspondence, and therefore the co-textual environment in which the negotiation of subjectivity and inter-personal meaning is undertaken. Proper names (e.g. *Elizabeth, Mary, Shrewsbury, Arbell*) are expected in a keyword list, as names are textually specific. That said, the presence of third-party reference, *Arbell* (Arbella Stuart, Bess of Hardwick's grand-daughter), suggests a recurrent topical focus, within Bess's letters, on family and domestic concerns, and sheds light on how her first-person pronominal usage may link to her social roles as grand-mother, wife and householder.

Adjectives (*dear, affectionate, care, happy*) are linked to author-recipient positioning within a positive, affective frame. For example, *dear* and *affectionate* occur in phrases that construct the recipient's identity in terms of the author, e.g. "my dear sweetheart" (Talbot) and the author's depiction of their identity with regards to the recipient, e.g. "your most affectionate sister and cousin" (Elizabeth I). These adjectival forms are therefore less frequent in PCEEC-16RN. *Affectionate* occurs only five times in PCEEC (0.03/1,000 words), four of which are found in letters by the Earl of Arundel writing to his wife in the early 1600s. This suggests that EWL authors are more likely to express an affective stance (in these terms) than their male contemporaries, although the type of letter should also be taken into account here. The preponderance of official letters in PCEEC-16RN entails that these affective adjectival forms are arguably less likely to occur. Tanskanen (2003) points out that the male-authored Hutton correspondence contains affectionate subscriptions and superscriptions, indicating that letter type, as well as recipient, play a significant part in the expression of affective stance. However, the Hutton letters date from 1770 to 1823, much later than the EWL corpus, suggesting that, as part of the increasing personalisation of letters over the course of the early modern period (Nevala 2004), writers were able to express more affective, intimate interpersonal meanings in the eighteenth century than was conventional in the sixteenth century.

Nominal keywords denoting both writer and recipient are also relevant for the marking of affective stance. Elizabeth I uses kinship terms, *sister* and *brother*, when writing to James VI, following the conventions for inter-royal communication. The forms' meanings are layered: they index sibling affection and sanguine proximity (as understood within a Protestant moral frame), whilst also signalling the author and recipient's specific membership of the royal elite, which legitimises such usage with non-kin referents. Her letters show a persistent adherence to this technique across twenty years of correspondence, suggesting that it provided a valuable frame for their epistolary conversation. The kinship terms, through their broader social meaning, as well as their more restricted, elite usage, promote alignment between Elizabeth and James at the start and end (at least) of every letter, framing – in a positive manner – the instruction and advice they contain.

Table 6 also includes function words indicative of stylistic strategies. Polarity adverbs e.g. *not, ever* suggest the construction of rescinded and superlative realities, often in reference to the author's own activities and behaviours, e.g. "I fear danger that I was not able for my life to withstand"; "I am assured, that none living, could be more dutiful, true, faithful, and careful to a husband, than I have ever been to you" (Talbot). *Suppose* is a verb of cognition, primarily indicating the beliefs, opinions or ideas of the author (in EWL, primarily Elizabeth I), but also projecting those of the recipient, and third parties; e.g. "I suppose his own conscience will never accuse me of any over great partiality" (Elizabeth I).

The keywords for PCEEC-16RN offer a contrastive perspective. The third-person pronouns (*he, her*) suggest that the male letter-writers more commonly discuss entities and events external to the author-recipient interaction, and the presence of titles (*lord, lordship, mr* and *majesty*) could indicate a greater use of nominal address, and/or third-party reference. The presence of *ye* in the PCEEC list complicates the self-referential picture, and prompts a closer examination of the second-person pronoun forms, here split into *you/thou* and *ye/you* variables (Table 7). Firstly, early modern writers had a choice between *you* and *thou* forms, each of which had distinct pragmatic meanings. *You* was the decorous form, used between social equals and those in polite society, with *thou* the marked pronominal, used in intimate and emotionally-charged exchanges (Nevala 2004: 177–178). *Thou* occurs infrequently in both corpora, with only six examples in EWL, a trend that fits with previous studies of correspondence (e.g. Nevala 2004). These occur in a letter by Elizabeth I. As discussed in Evans (2013), this letter from the queen to her godson, John Harrington, is an affectionate message enclosing a transcript of Elizabeth's parliamentary speech. Her "boy Jack" was to ponder his godmother's "poor words". The letter's general positioning of intimacy through wordplay and informal address (e.g. nicknames) could explain the use of *thou*, which contrasts with Elizabeth's typical pronominal practice; several decades later, the daughter of James I, Elizabeth of Bohemia, combined *thou* and familiar nicknames, such as *Honest Thom* in her letters to Sir Thomas Roe; see e.g. her letter of 19th November 1629 (Stuart 2015: 791–792). However, the letter between Elizabeth I and Harrington survives only as an eighteenth-century copy, a factor which, given the rarity of *thou* in letters of the period surveyed here more generally, raises questions about the reliability of the transcript.

Table 7. Frequencies of 2PP in EWL and PCEEC-16RN

	2PP (N.)	THOU (all cases) (/1,000 words)	YOU (all cases) (/1,000 words)	YOU (subj. & obj.) (/1,000 words)	YE (subj. & obj.) (/1,000 words)
EWL	4,545	10 (0.01)	4,535 (41.6)	2,551 (23.4)	91 (0.8)
PCEEC-16RN	4,882	14 (0.1)	4,868 (29.3)	1,861 (11.2)	608 (3.7)

Ye, originally the subject form, became increasingly marginal throughout the early modern period (Lass 2000: 154). Its replacement by *you* has a gendered dimension: Nevalainen and Raumolin-Brunberg (2003: 118–119) observe that women lead the change in the early- and mid-sixteenth century, which occurs at a slower rate in the (male-only) upper and middle social ranks (2003: 142). Regionally, *you* emerges in southern and eastern parts of the country (2003: 172). The infrequency of *ye* in EWL and its greater quantity in PCEEC-16RN are congruent with this sociolinguistic profile in terms of gender and also social rank and region (excepting Talbot, who was from Derbyshire). Pragmatic meanings may also be relevant. Nevalainen and Raumolin-Brunberg (2003: 194) note that, in early stages of the change (1500–1540), register variation is significant: *you* occurs more often in letters to family recipients. At the end of the century, male letter-writers at the court "stand out as a conservative force" (Nevalainen and Raumolin-Brunberg 2003: 195), continuing to use *ye* in their letters even when *you* has become the norm.

Of the 91 occurrences of *ye* in EWL, 62 appear in the (near exclusively) male-authored scribal letters of Elizabeth I, Mary I and Talbot, thus fitting the sociolinguistic profile. The connection with scribal correspondence could indicate an association with more formal communicative activities than the holograph equivalents, which mainly use *you*, indicating a possible connection to author stance and interpersonal relations. Within the holograph correspondence, generational differences may be at play: the form occurs in the letters of the two oldest women, Parr, writing to fiancé Thomas Seymour, and Mary when writing to her father's Principal Secretary Thomas Cromwell; examples which offer little support for the association of *ye* with socially distant recipients.

The distribution of *you* and *ye* represents a relatively minor linguistic difference between EWL and PCEEC. However, the pronoun forms co-occur with other linguistic features, and cumulatively the second-person forms may comprise a specific epistolary style, or stance repertoire (Kiesling 2009), indexing scribal (and potentially official) practices, with possible gendered associations.

4.2 Holograph and scribal identities in the EWL

This section explores self-reference forms in scribal and holograph letters in EWL to establish if and how material status indexes stance. The first-person self-reference pronouns (singular and plural combined) occur at a similar frequency in both sub-corpora: 66/1,000 words in the holograph correspondence, and 60.3/1,000 words in the scribal correspondence (both more frequent than PCEEC-16RN: 46.4/1,000 words). Yet the distribution between singular and plural forms is very different (Table 8).

Table 8. 1PP frequencies in holograph and scribal letters

	I (/1,000 words)	**ME (/1,000 words)**	**MY/MINE (/1,000 words)**	**Total 1PP singular (/1,000 words)**
EWL Holograph	1,527 (30.0)	462 (9.1)	1,132 (22.3)	3,121 (61.4)
EWL Scribal	457 (7.8)	182 (3.1)	358 (6.1)	997 (17.1)
	WE	**US**	**OUR**	**Total 1PP plural**
EWL Holograph	74 (1.5)	34 (0.7)	124 (2.4)	232 (4.6)
EWL Scribal	969 (16.6)	334 (5.7)	1,211 (20.8)	2,514 (43.2)

In the holograph letters, the first-person singular dominates, occurring three times more frequently than in the scribal letters, and almost double the frequency identified in PCEEC-16RN (39.4/1,000 words). The plural forms, comprising inclusive and exclusive *we*, are less frequent (4.6/1,000 words), suggesting a focus on individual subjectivity, rather than a collective or shared experience. The R1 collocates of *I* in the holograph letters suggest mitigated directives and requests; e.g. *pray, may, beseech, hope, trust*, which can be linked to the affective stance practices discussed above. Conversely, the scribal letters contain a higher frequency of plural pronouns (43.2/1,000 words). However, this does not indicate a tendency towards a collective stance. The majority of examples are *royal we* forms, a conventionalised self-reference in which the monarch refers to themselves using the plural form; e.g. "we could not but particularly let you understand by our own letters, how well we do think of you for the same" (Elizabeth I to Earl of Argyll, 20 January 1582); for a discussion of the likely origins and modern trajectory of *royal we*, see Wales (1996). In the scribal letters of Mary and Elizabeth, the plural form supplants the singular alternative. Only Talbot's scribal letters contain unambiguous instances of the first-person plural: 1.7/1,000 words, lower than the holograph correspondence.[1] Thus, the holograph letters of the royal writers share the pool of

1. *Royal we* is a highly conventionalised form in royal scribal letters of the period, and examples of referential ambiguity are rare (see Evans 2018). For an overview of the use of 'we' across different languages and contexts, see Pavlidou (2014).

self-reference resources with their noble contemporaries: *royal we* occurs rarely in the holograph letters of Elizabeth I and not at all in the letters (of those in EWL, at least) of Mary and Parr.

The royal scribal letters thus provide a marked example of how self-reference forms and material attributes co-occur to index stance. The letters follow epistolary conventions developed for (male) sovereigns. In terms of their composition, royal letters were a collaborative endeavour, which may include – but did not necessitate – the involvement of the monarch. The secretariats of Mary and Elizabeth worked to a house style, which followed historical epistolary models stemming from the *ars dictaminis* (Richardson 2001). The letters' contents and expression were drafted, edited and approved by multiple individuals prior to their transcription into fair copy, ready for the royal signature and dispatch. Thus, royal scribal letters present a complex case in which the agency of the royal author is generally difficult to establish (see Evans 2016). Nevertheless, the resultant stylistic properties undertake stance-marking and interpersonal work.

We suggest that *royal we* indexes an explicitly royal social role, anchored to the physical and moral order of its referent. It signals the author's authority both as an *a priori* hierarchical, institutional structure, but also within the epistolary text itself. This can be seen in some of the most frequent R1 collocates of *we*: e.g. *have, greet, require, think, do* and *did*. Requests are made and permissions granted that directly affect the recipient e.g. "we do authorise you to give license" (Mary I to Stephen Gardiner). *Greet* is a meta-pragmatic formulaic salutation "we greet you well", which prefaces the royal scribal letters almost without exception. Notably, *greet* is also a top collocate of *we* in PCEEC-16RN. The royal scribal letters use epistolary conventions to undertake communicative activities that, in combination, index both the authors' authoritative interactional stance, and their socio-legal authority as monarch. Given the patriarchal frame of European sovereignty, the self-reference forms and co-textual features of the scribal letters feasibly index an authority that is inherently masculine. By comparison, Talbot's scribal letters use conventional first-person *I* and *we*. The collocates of *I* can be linked to acts of request, but they lack the authoritative stance of the royal letters e.g. *I beseech, I pray*. The collocates of *we* in Talbot's letters typically point to hypothetical scenarios relevant to author and recipient e.g. "if it please you to regard it bind me and mine in bonds that we shall ever bear to do your Lordship thankful service".

The differences between the self-reference forms and associated stance-marking in the holograph and scribal letters are not unexpected: they fit with early modern evaluations of a holograph letter as the more intimate and private epistolary format. The contrast in self-reference is particularly acute in the royal scribal letters, although a modest correlation is observable in the letters of Talbot, and fits with previous discussions of linguistic differences between holograph and scribal letters in the period (e.g. Williams 2014; Marcus 2017).

Keyword analysis comparing the holograph and the scribal letters reveals that first-person pronouns are the most statistically significant keywords (singular forms in the holograph letters, and plural forms in the scribal letters). However, subsequent keywords – all of which are highly statistically significant ($p < 0.001$) – relate to epistolary conventions of each letter type. In the holograph letters, the keywords can be connected to affective stance marking, focussed on interpersonal relations with the recipient e.g. “dear heart”, “my most dear father”. Scribal keywords *trusty* and *beloved* serve a similar function in the scribal letters as part of more fixed address formulae. These genre-specific constructions combine with the material and linguistic features observed above, contributing to the affective and authoritative stance of the holograph and scribal texts, respectively.

Many of the keywords from the holograph/scribal analysis appear in the EWL-PCEEC16RN comparison, whereas some of the aforementioned keywords from the scribal/holograph comparison appear as PCEEC-16RN keywords when PCEEC is contrasted with EWL. This could suggest that EWL holograph letters possess a more individualised epistolary style, keyed to the interactional requirements of the women letter-writers, whereas the EWL scribal letters adhere to a more generic and conventionalised style. For instance, in the royal scribal letters, the status of the writer, indexed throughout the letter text through *royal we*, may entail that more extensive interpersonal work is not required. However, the influence of the holograph/scribal status is not necessarily the only factor governing the relationship between writer and recipient; the primary communicative function of the letter may also play a part as we show in the following section.

Overall, the differences between the holograph and scribal letters suggest that the material status is part of stance-marking, indexing the relationship with recipient and subject; comparable with the use of objects to complement language choices in spoken conversation (e.g. Burkette 2016). To explore further the interactional dimensions of self-reference within their material and pragmatic contexts, the final section examines the impact of stance and self-reference practices.

4.3 Letter recipient and function in EWL

Understood as a component of register, the power and social distance between author and recipient influences the management of subjectivity and interpersonal relations in correspondence. To investigate this, the holograph and scribal letters are split into three recipient groups: superior, equal and inferior. The categorisations reflect *a priori* rankings of social roles. The analytic approach does not dispute the potential for power relations to develop throughout interaction (cf. Locher 2004).

However, the salience of social structures for sixteenth-century English society suggests that interlocutor status has a strong influence on a writer's interactional choices. The quantitative profiles of letters to superior, equal and inferior recipients reveal the more distinctive characteristics of self-reference in these sub-groups. We consider how they relate to identity-in-interaction in the letters themselves, and how such choices may index indirect social meanings, linked to personas and social categories, particularly gender.

Figure 1 shows the frequency of first-person pronouns in the holograph and scribal letters, divided by recipient relationship. For the scribal letters, the frequencies include *royal we*, which is referentially equivalent to *I*.

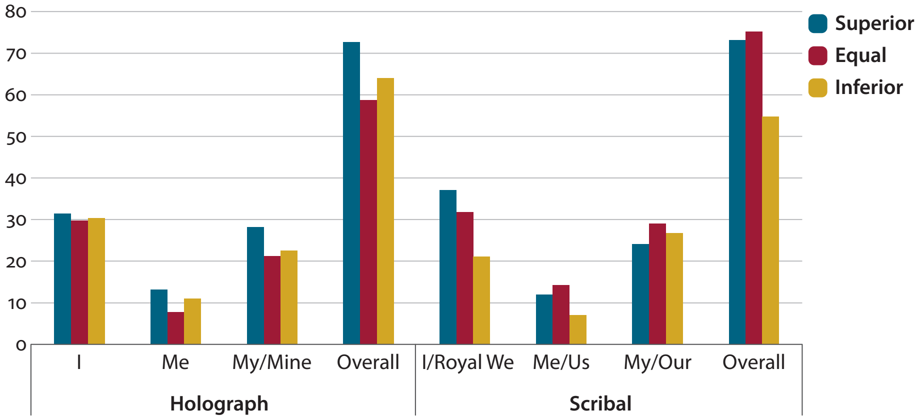

Figure 1. Frequencies (per 1,000 words) of 1PP in EWL by recipient types

In the holograph correspondence, letters to social superiors contain the highest proportion of first-person singular pronouns (72.7/1,000 words). Letters to social equals contain the least (58.7/1,000 words), with inferiors in-between (63.9/1,000 words). This suggests that self-reference is a significant component when writing from a position of deference, and can be understood to reflect the letter-writers' greater efforts in stance-marking and alignment. In the letters to equals and inferiors, this positioning is perhaps less essential for the letters' primary communicative function. Conversely, scribal letters to social equals contain the highest proportion of first-person pronouns (75.2/1,000 words), with letters to social superiors also similar (73.2/1,000 words). Letters to inferiors contain the least instances. Unlike the holograph letters, in which the superior recipient letters have the highest frequency of pronouns for all cases, the distribution for the scribal letters is inconsistent. Superior recipient scribal letters contain the highest proportion of *I/royal we*, whereas object and genitive forms are more frequent in letters to social equals.

Figure 2 compares the distribution of second-person forms. In the holograph letters, frequencies are similar across recipient categories, but with differences between the case forms. This is most striking for superior recipients: there are few subject and object forms (5.2/1,000 words), but *your* occurs very frequently (44.1/1,000 words). This is explained by the use of direct address e.g. "your grace", "your lordship", and properties attributed to the recipient; for example "saving the confidence I have in your blessed nature" (Mary to Henry VIII). The distribution of 2PP in the scribal letters is less differentiated, but mirrors the patterning seen in the holograph letters.

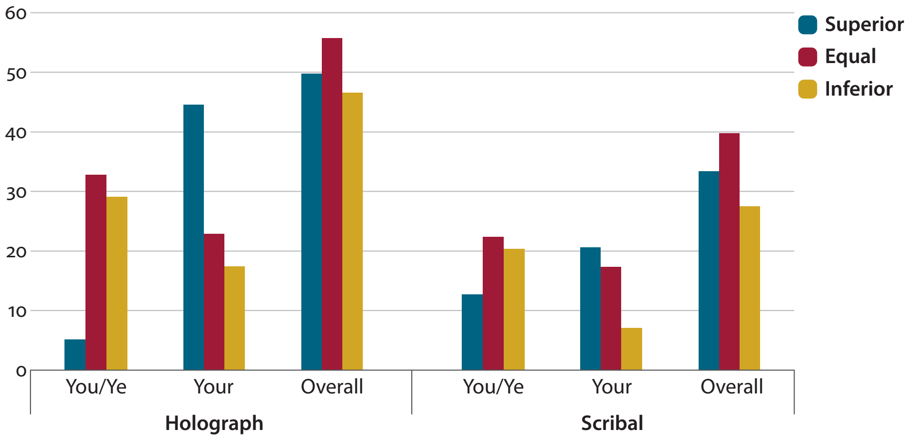

Figure 2. Frequencies (per 1,000 words) of 2PP in EWL by recipient type

Quantitatively, holograph letters to superiors have the most distinctive profile for self-reference and direct address, and qualitative analysis indicates how these pronominal trends are shaped by the letters' primary communicative function and recipient relationship. A letter from Mary to her father, King Henry VIII, provides a useful exemplar. The letter, reproduced below, forms part of a sequence written when Mary had been removed from the line of succession. Mary sought Henry's benevolence and therefore, by implication, her reinstatement. The letter's primary function is therefore one of request.

The holograph letter situates Mary as the recipient's subordinate. She describes herself as being physically prostrate before the king, at his "most noble feet", and reinforces this description with pragmatic features that metaphorically convey a prostrate position. For example, Mary uses the *your* + title address structure eight times (44/1,000 words) employing *synonymia* (elegant variation e.g. highness – majesty – grace) to achieve this (Adamson 2000: 556–559). The direct address combines with positive epithetic descriptions of the recipient, framed through his social roles with Mary: as prince, and as father. This strategy acknowledges his superiority in both capacities, but also triggers moral associations of both roles, encouraging a response that would fit with the character of a benevolent king and responsible father.

> Most humbly prostrate at *your Highness*' most noble feet with like desire of *your Majesty's* daily blessing: *I* do in semblable manner thank *your Grace* for the great and inestimable goodness which your most noble princely heart and father pity does daily extend and show unto *me*, the least piece whereof albeit *I* shall never be able towards *your Highness* to recompense, yet *I* trust with God's help with such faithful obedience redub that *I* have offended, that *your Grace* nor yours shall never have cause to repent any part of your most abundant goodness declared towards *me*; trusting in God that *your Grace* doth so accept *me*, who shall never during my life do or attempt that thing, that may sound to the contrary to win by the same the whole world; as our Lord knoweth to whom *I* do and shall daily pray for the preservation of *your most excellent Highness* in health and felicity, with the granting of your noble and virtuous desires. From Hertford the second day of October. *Your grace's most humble daughter and subject. Mary*.
>
> [emphasis added] (Hearne 1716: 143-144)

The formulaic, superlative address forms reinforce the recipient's high status, while the absolute negator, *never*, provides an emphatic declaration of Mary's loyalty: "shall never have cause to repent any part". The latter expression uses the incoming negative concord construction, which was associated at this time with formal and professional texts (Nevalainen 1998). The emphasis on "daily" blessings of Henry and God contributes to the hyperbolised, superlative expression. The letter also includes components of the Humanist-influenced *ars dictaminis*, such as the closing benediction. The materiality, as a holograph letter, indexes the time and labour taken over composition, indirectly signalling respect for the addressee.

Mary's letter follows prescribed epistolary norms for communicating with a superior. Yet, within this frame, her linguistic choices activate other indexical associations such as her status as a learned, literate noblewoman; a status that, as Henry's offspring, would reflect positively on the king as well. The letter is also an explicitly gendered interaction, e.g. "humble daughter". The self-derogation and declarations of filial loyalty, achieved through the linguistic features discussed above, can be linked to contemporary social models of women as the deferential and subservient sex (Eales 1998: 4). Indeed, the incorporation of gender as a facet of Mary's epistolary subjectivity was a conventional requirement; and yet it was not necessarily so explicit (as Talbot's letters will show). To a degree, therefore, it is a deliberate act. Not acknowledging her subordinate position as a female subject may have negatively impacted her construed identity and intended objectives.

The holograph letters of Parr and Talbot to superior recipients show similar linguistic strategies. Talbot's letters sent to noble superiors include frequent address nominals (e.g. "your Lordship"), and construct a hyperbolic, intensified stance using superlatives, absolute negators and non-assertive forms; e.g. "but will rest ever most faithful as many your most honourable favours may Justly challenge".

However, Talbot represents herself gender-neutrally as a “friend”, probably due to her non-kin relationship with her superior recipients (unlike Mary and Parr). One exception occurs in a petitionary letter to the Earl of Leicester, “my good lord and brother”, written May c.1577, sent from “your lordship’s faithful sister”. Her letter sought a marital match for her daughter. The pseudo-kin terminology sets up an affective stance that draws on the social ideologies of sibling relationships to bolster her request. Other features in the letter are more deferential; for example, there are eighteen address forms *your lordship* (40.4/1,000 words), and superlatives and absolute negators intensify the proclamations of loyalty and gratitude; e.g. “I being of no power nor ability any ways to make appear the duty of thanks”. Materially, the holograph status signals proximity and respect towards the recipient. The kinship terminology therefore allows Talbot to combine the conventional deferential stance with a more intimate depiction of her relationship with the recipient, presumably to increase the chances of her petition’s success.

The contextual meaning of stance is evident when the holograph letters to social equals are compared. Many of the letters share the same primary functions with those to superiors (e.g. petition, request), but the pronominals occur in a different co-textual environment. For example, many of Talbot’s letters to her fourth husband, George Talbot, request various domestic materials, such as bottles of sack. Her requests are bold and assertive. For example, in a reminder to her husband that he promised her financial assistance – “you promised to send me money afore this time too by oxen” – she offers an evaluation that operates on both epistemic and affective dimensions: “but I see out of sight out of mind with your unkind none”. The appeal to proverbial wisdom supports the declaration of her subjective interpretation of the situation “I see”, which can be understood as being given in jest. Her teasing is made clear by ventriloquising G. Talbot’s own diminutive term for her (“none”). The choice foregrounds the interpersonal relationship as one of husband and wife, within an expressly affective stance. The use of *your* with a self-reference nominal, rather than in direct address, is typical of letters to social equals, and underpins the lower frequency of this case-form in the sub-group.

In the scribal letters, the frequency of pronouns in letters to inferior recipients is the lowest across EWL as a whole, suggesting this communicative context is particularly marked in terms of subjectivity and interpersonal relations, representing something of a counterpoint to the holograph letters to superior recipients, the category with the highest frequency of self-reference pronouns.

Scribal letters to inferior recipients are the majority relationship group for royal scribal letters, but are under-represented in Talbot’s corpus (two letters). The royal correspondence, which is mostly official, uses formulaic components, particularly for letter openings and closings, to undertake habitual yet significant stance work. This can be seen in the following letter from Elizabeth I to Thomas Randolph:

> *BY THE QUEEN. Elizabeth R. Trusty and well-beloved*, we greet you well. Understanding by a letter of our cousin, the earl of Bedford, to our Secretary, that he thinketh you to be not out of danger there by malice, in so much as some pistols have been shot into the house where you lodge ; *we have thought it meet*, not only to advise you, if you find any such peril, but also to license you to withdraw yourself for a time to Berwick : pretending the same to be for your own *private business*, in sort, as you may return to your charge upon occasion. *Given under our signet*, at our Palace of Westminster, the third of October, 1565, the seventh year of our reign. [emphasis added] (BL Lansdowne 8 f.99)

The text is framed with referential features that index the royal source of the text, including the opening declarative prepositional phrase "By the Queen", the authenticating signature, and the formulaic salutation "Trusty and well-beloved". It concludes with a similarly formulaic statement commenting on the place and time of composition. These genre-specific and authority-marking features frame the actual message of the text, which shows a more nuanced engagement with the recipient. When granting Randolph permission to withdraw, the mental verb in "we have thought it meet" meta-pragmatically comments on the Queen's engagement with the affairs of the recipient, indexing a caring royal persona. This is extended in the subsequent directive that Randolph pretend he is on "private business", presumably to protect his honour. This implicit step signals an assumed intimacy with the recipient, within the formulaic authoritative frame of the epistle.

Overall, the lower frequency of self-reference forms in royal scribal letters can be understood to arise partly from the material and linguistic conventions of these particular letters. Rather than repeatedly emphasising one's deferential stance (as in Mary's letter to Henry VIII), the institutional power indexed by the linguistic and material components plausibly holds sway over the utterance, without the need for repetition. The royal scribal letters thus signal an authority that, through the consistent reoccurrence of stance-marking features and self-reference, such as *royal we*, reinforces and relegitimises the indexical power of the self-reference forms themselves.

The letters of Talbot provide an important counterpoint to the specific epistolary language of the royal scribal letters. Her letters comprise the only examples to superior recipients in EWL. These texts draw on formulaic components, similar to the royal scribal texts, although with different stance-marking effects. For example, in Talbot's petitionary letter to Sir Robert Cecil requesting support during a family feud, self-reference forms occur in deferential formulae in the letter's opening and closing sections e.g. "I am loathe to be troublesome to you", "I have and do rely thereon", "I beseech god". Letters to Talbot's superiors also draw on the friendship trope, with variations of "your very loving friend" present in a number of these letters. This is an interesting contrast with the scribal letter from Elizabeth, discussed above, which intimated royal friendship (through the protection of Randolph's honour) without making it textually explicit. Talbot's scribal letters are therefore more affectively demonstrative, even whilst adhering to epistolary conventions befitting her social status and communicative objectives.

5. Conclusion

The exploration of self-reference in a corpus of 'exceptional' women's letters reveals the ways in which their strategies of interaction draw on epistolary conventions of the period. Viewed through the lens of stance, the comparative analysis with PCEEC-16RN indicates similarities in the presentation of writer subjectivity, based on the pronominal frequencies and top collocates of *I*, yet with subtle differences relating to the presentation of knowledge and understanding (e.g. *see*) and the framing of relationships with the recipient (e.g. *commend*). The differences are most striking when the material status is considered, particularly in the royal scribal letters, which incorporate epistolary formulae and specific linguistic features, such as *royal we*, to index an authoritative stance, evoking higher-level identities of royal – and potentially masculine – personas. Conversely the holograph letters use self-reference and associated forms to foreground affective stances, such as deference or intimacy. Thus, the material status of an early modern letter is a critical component of stance-marking, drawing on, and reinforcing, social conventions regarding the social distance and power differential in epistolary interaction. Finally, the analysis examined how the larger quantitative patterns relate to linguistic choices on the micro-level, exploring the effect of recipient type upon the women's use of self-reference and forms of address. Self-reference pronouns and nominals intersect consistently with other linguistic features, such as superlatives, and epistolary formulae, as the writers use their correspondence to make requests, express their gratitude or authorise behaviours. The clustering of stance-marking forms in letters written to particular recipient groups, and for particular purposes, supports the proposed relationship between stance and style (Kiesling 2009). Moreover, the linguistic choices can be seen to draw on their macro-level associations, such as their connection to a learned persona (e.g. *synonymia*), or to social groups connected to the Court (e.g. negative concord).

The investigation of EWL shows that the four women work within the conventions of the epistolary genre to achieve their communicative objectives. At times, this requires the explicit invocation of their sex and that of their recipient, such as the allusion to the moral responsibilities of a husband, or a father. The use of an affective stance in the holograph letters could also signal, and reflect, social values of femininity; markers such as *affectionate* and *dear* are notable by their sparsity in

PCEEC-16RN. These features also occur less often in the scribal letters, particularly in the royal correspondence, in which royal authority – and benevolence – are the key indices of stance-marking, but these may themselves signal masculinity on account of their material and generic provenance.

Overall, the study suggests that stance provides a useful framework for identifying and unpicking the relationship between the interactional language choices of a letter-writer and the larger macro-level patterns relating to social values, genre conventions and so on. For the four exceptional women, their social status did not mean that their gender was irrelevant for their presentation of subjectivity and the negotiation of interpersonal relations, yet it was not the only relevant aspect, either. Their letters represent public communicative acts: a declaration of their experience with the world, and therefore their identity within it, which could be read by both known and unexpected parties. Their stance-marking choices, and the resultant epistolary styles, show their assured use and manipulation of the linguistic and material affordances of the epistolary genre.

References

Adamson, Sylvia. 2000. "Literary Language." In *The Cambridge History of the English Language*, ed. by Roger Lass, 539–653. Cambridge: Cambridge University Press. https://doi.org/10.1017/CHOL9780521264778.008

Akkerman, Nadine. 2018. *Invisible Agents: Women and Espionage in Seventeenth-Century Britain*. Oxford: Oxford University Press.

Allinson, Rayne. 2012. *A Monarchy of Letters: Royal Correspondence and English Diplomacy in the Reign of Elizabeth I*. New York: Palgrave Macmillan. https://doi.org/10.1057/9781137008367

Alonso-Almeida, Francisco. 2009. "Stance Marking and Register in Middle English Charms." *Studia Anglica Posnaniensia: International Review of English* 45: 13–30. https://doi.org/10.2478/v10121-009-0002-9

Anthony, Laurence. 2018. *AntConc* (version 3.5.6). Windows. Tokyo: Waseda University. http://www.laurenceanthony.net/software/antconc/ (accessed 15 April 2019).

Barnes, Sonia. 2018. "Copula Variation in Asturian Spanish and the Multidimensionality of Stancetaking in Interaction." *Journal of Sociolinguistics* 22 (1): 29–54. https://doi.org/10.1111/josl.12273

Baron, Alistair. 2017. *VARD 2* (version 2.5.4). Lancaster: UCREL, University of Lancaster.

Biber, Douglas, and Edward Finegan. 1989. "Styles of Stance in English: Lexical and Grammatical Marking of Evidentiality and Affect." *Text – Interdisciplinary Journal for the Study of Discourse* 9 (1): 93–124. https://doi.org/10.1515/text.1.1989.9.1.93

British Library Lansdowne MS Vol. 8.

Bucholtz, Mary, and Kira Hall. 2005. "Identity and Interaction: A Sociocultural Linguistic Approach." *Discourse Studies* 7 (4–5): 585–614.

Burkette, Allison. 2016. "Linguistic and Object-Based Stance-Taking in Appalachian Interviews." *Language in Society* 45 (3): 331–351. https://doi.org/10.1017/S0047404516000063

Dambrogio, Jana, and Daniel Starza-Smith. 2017. *Dictionary of Letterlocking*. http://letterlocking.org/dictionary (accessed 15 April 2019).

Day, Angel. 1592. *The English Secretorie, or, Plaine and Direct Method, for the Enditing of All Manner of Epistles or Letters*. Imprinted at London: By Richard Iones, dwelling at the Rose and Crowne neere Holborne Bridge.

Daybell, James. 2012. *The Material Letter in Early Modern England*. London: Palgrave Macmillan. https://doi.org/10.1057/9781137006066

Dossena, Marina. 2006. "Doing Business in Nineteenth-Century Scotland: Expressing Authority, Conveying Stance." *IEEE Transactions on Professional Communication* 49 (3): 246–253. https://doi.org/10.1109/TPC.2006.880736

Dossena, Marina, and Susan M. Fitzmaurice (eds). 2006. *Business and Official Correspondence: Historical Investigations*. Bern: Peter Lang.

Du Bois, John W. 2007. "The Stance Triangle." In *Stancetaking in Discourse: Subjectivity, Evaluation, Interaction*, ed. by Robert Englebretson, 139–182. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.164.07du

Duszak, Anna (ed). 2002. *Us and Others: Social Identities across Languages, Discourses and Cultures*. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.98

Eales, Jacqueline. 1998. *Women in Early Modern England, 1500–1700*. London: Routledge.

Eckert, Penelope. 2000. *Linguistic Variation as Social Practice: The Linguistic Construction of Identity in Belten High*. Malden, Mass: Blackwell Publishers.

Erasmus, Desiderius. 1985. *Collected Works of Erasmus. […] 4: Literary and Educational Writings De Pueris Instituendis, de Recta Pronuntiatione*. Translated by Charles Fantazzi, and Brian McGregor. Toronto: University of Toronto Press.

Evans, Mel. 2013. *The Language of Queen Elizabeth I: A Sociolinguistic Perspective on Royal Style and Identity*. Chichester: Blackwell.

Evans, Mel. 2016. "By the Queen: Collaborative Authorship in the Scribal Letters of Queen Elizabeth I." In *Women and Epistolary Agency in Early Modern Culture*, ed. by James Daybell, and Andrew Gordon, 36–54. London: Routledge.

Evans, Mel. 2018. "Styling Power: A Corpus Linguistic Approach to the Correspondence of Queen Elizabeth I." In *Elizabeth I in Writing*, ed. by Donatella Montini, and Iolanda Plescia, 59–82. New York: Palgrave Macmillan. https://doi.org/10.1007/978-3-319-71952-8_4

Fairclough, Norman. 1994. "Conversationalization of Public Discourse and the Authority of the Consumer." In *The Authority of the Consumer*, ed. by Nicholas Abercrombie, Russell Keat, and Nigel Whiteley, 235–248. London: Routledge.

Fitzmaurice, Susan. 2004. "Subjectivity, Intersubjectivity and the Historical Construction of Interlocutor Stance: From Stance Markers to Discourse Markers." *Discourse Studies* 6 (4): 427–448. https://doi.org/10.1177/1461445604046585

Fulwood, William. 1568. *The Enimie of Idlenesse Teaching the Maner and Stile How to Endite, Compose and Write All Sorts of Epistles and Letters*. Imprinted at London: By Henry Bynneman, for Leonard Maylard.

Gal, Susan. 2002. "A Semiotics of the Public/Private Distinction." *Differences: A Journal of Feminist Cultural Studies* 13 (1): 77–95. https://doi.org/10.1215/10407391-13-1-77

Gray, Bethany, and Douglas Biber. 2012. "Current Conceptions of Stance." In *Stance and Voice in Written Academic Genres*, ed. by Ken Hyland, and Carmen Sancho Guinda, 15–33. New York: Palgrave Macmillan. https://doi.org/10.1057/9781137030825_2

Hearne, Thomas. 1716. *Sylloge epistolarum, a variis Angliæ principibus scriptarum*. Oxonii e Theatro Sheldoniano.

Hubbard, Eleanor. 2015. "Reading, Writing, and Initialing: Female Literacy in Early Modern London." *Journal of British Studies* 54 (3): 553–577.

Hunston, Susan. 2007. "Using a Corpus to Investigate Stance Quantitatively and Qualitatively." In *Stancetaking in Discourse: Subjectivity, Evaluation, Interaction*, ed. by Robert Englebretson, 27–48. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.164.03hun

Jaffe, Alexandra. 2009. "Introduction: The Sociolinguistics of Stance." In *Stance: Sociolinguistic Perspectives*, ed. by Alexandra Jaffe, 3–26. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195331646.001.0001

Kiesling, Scott F. 2009. "Stance as the Explanation for Patterns of Sociolinguistic Variation." In *Stance: Sociolinguistic Perspectives*, ed. by Alexandra Jaffe, 171–194. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195331646.001.0001

Lass, Roger. 2000. "Phonology and Morphology." In *The Cambridge History of the English Language*, ed. by Roger Lass, 56–186. Cambridge: Cambridge University Press. https://doi.org/10.1017/CHOL9780521264761.004

Leaper, Campbell, and Rachael D. Robnett. 2011. "Women Are More Likely Than Men to Use Tentative Language, Aren't They? A Meta-Analysis Testing for Gender Differences and Moderators." *Psychology of Women Quarterly* 35 (1): 129–142. https://doi.org/10.1177/0361684310392728

Locher, Miriam A. 2004. *Power and Politeness in Action: Disagreements in Oral Communication*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110926552

Magnusson, Lynne. 1999. *Shakespeare and Social Dialogue: Dramatic Language and Elizabethan Letters*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511483745

Marcus, Imogen. 2017. *The Linguistics of Spoken Communication in Early Modern English Writing: Exploring Bess of Hardwick's Manuscript Letters*. London: Palgrave Macmillan.

McConnell-Ginet, Sally. 2011. *Gender, Sexuality, and Meaning: Linguistic Practice and Politics*. Oxford: Oxford University Press.

Moore, Emma, and Robert Podesva. 2009. "Style, Indexicality, and the Social Meaning of Tag Questions." *Language in Society* 38 (4): 447–485. https://doi.org/10.1017/S0047404509990224

Mühlhäusler, Peter, and Rom Harré. 1990. *Pronouns and People*. Cambridge, MA.: Basil Blackwell.

Nevala, Minna. 2004. *Address in Early English Correspondence: Its Forms and Socio-Pragmatic Functions*. Helsinki: Société Néophilologique.

Nevalainen, Terttu. 1998. "Social Mobility and the Decline of Multiple Negation in Early Modern English." In *Advances in English Historical Linguistics (1996)*, ed. by Jacek Fisiak, and Marcin Krygier, 263–291. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110804072.263

Nevalainen, Terttu, and Helena Raumolin-Brunberg. 2003. *Historical Sociolinguistics: Language Change in Tudor and Stuart England*. London: Pearson Education.

Nurmi, Arja, Minna Nevala, and Minna Palander-Collin (eds). 2009. *The Language of Daily Life in England (1400–1800)*. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.183

Ochs, Elinor. 1992. "Indexing Gender." In *Language as an Interactive Phenomenon*, ed. by Alessandro Duranti, and Charles Goodwin, 335–358. Cambridge: Cambridge University Press.

OED Online. 2018. "Commend, V." *Oxford English Dictionary*. Oxford: Oxford University Press. http://www.oed.com

Palander-Collin, Minna. 2010. "Correspondence." In *Historical Pragmatics*, ed. by Andreas H. Jucker, and Irma Taavitsainen, 651–678. Berlin: Mouton De Gruyter.

Palander-Collin, Minna. 2011. "Variation and Change in Patterns of Self-Reference in Early English Correspondence." In *Historical Sociopragmatics*, ed. by Jonathan Culpeper, 83–108. Amsterdam: John Benjamins. https://doi.org/10.1075/bct.31.05pal

Pavlidou, Theodossia-Soula (ed). 2014. *Constructing Collectivity: 'We' across Languages and Contexts*. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.239

Richardson, Malcolm. 2001. "The Fading Influence of the Medieval Ars Dictaminis in England After 1400." *Rhetorica* 19 (2): 225–247. https://doi.org/10.1525/rh.2001.19.2.225

Scott, Mike, and Chris Tribble. 2006. *Textual Patterns: Key Words and Corpus Analysis in Language Education*. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.22

Simpson, Paul. 1993. *Language, Ideology, and Point of View*. London: Routledge. https://doi.org/10.4324/9780203312612

Snell, Julia. 2010. "From Sociolinguistic Variation to Socially Strategic Stylisation." *Journal of Sociolinguistics* 14 (5): 630–656. https://doi.org/10.1111/j.1467-9841.2010.00457.x

Stuart, Elizabeth. 2015. *The Correspondence of Elizabeth Stuart, Queen of Bohemia Volume 1: 1603–1631*, ed. by Nadine Akkerman. Oxford: Oxford University Press.

Tanskanen, Sanna-Kaisa. 2003. "Best Patterns for Your Imitation: Early Modern Letter Writing Instruction and Real Correspondence." In *Discourse Perspectives on English: Medieval to Modern*, ed. by Risto Hiltunen, and Janne Skaffari, 167–196. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.119.08tan

Wales, Katie. 1996. *Personal Pronouns in Present-Day English*. Studies in English Language. Cambridge: Cambridge University Press.

Wiggins, Alison. 2017. *Bess of Hardwick's Letters: Language, Materiality and Early Modern Epistolary Culture*. New York: Routledge.

Williams, Graham T. 2014. *Women's Epistolary Utterance: A Study of the Letters of Joan and Maria Thynne, 1575–1611*. Pragmatics & Beyond New Series 233. Amsterdam: John Benjamins.

Yáñez-Bouza, Nuria. 2015. *Grammar, Rhetoric and Usage in English: Preposition Placement 1500–1900*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511732522.005

Kinship references in the British Parliament, 1800–2005

Jukka Tyrkkö
Linnaeus University

Family and kinship are fundamental concepts of human society and of political governance. Proceeding from the notion that frequency of reference can be taken as a simple measure of visibility and representation, this study examines diachronic trends of kinship reference in British parliamentary debates from 1800 to 2005. Using the Hansard Corpus and pattern-driven corpus linguistic methods, I show that changes in the frequencies of reference to kin reflect societal attitudes to gender roles.

Keywords: kinship, representation, reference, parliamentary debate, corpus, pattern, political discourse

1. Introduction

The concept of family is fundamental to all human societies. Whether understood in the narrow sense as the *nuclear* or *immediate* family, in the current sense as the *postmodern* family, or in the traditional sense as the *extended* family, family and family relations are and always have been frequent points of reference in public and private discourses (Bryant 2016). Consequent to the central role played by family in social and civic life, there is a long tradition of societal interference in a wide variety of matters related to family and family relations through legislative, moral and religious regulations that govern marriage, sexual relations, inheritance, and many other aspects of life.

Kinship, a concept closely related to family, is defined by the *Oxford English Dictionary* as "the quality or state of being of kin" and further as "the recognized ties of relationship, by descent, marriage, or ritual, that form the basis of social organization" (s.v. "kinship", OED, sense I/b). Although often defined by means of consanguinity, kinship can also be established through tradition or ritual, and thereby extended beyond the family relations. In every case, however, kinship

https://doi.org/10.1075/pbns.306.04tyr

refers, either directly or indirectly, to a shared sense of common ancestry or history of some type. Given the importance of family and kin in society, it may be expected that references to kin should be ubiquitous in political discourse. This not only means legislative discourses that directly concern the regulation of family relations, but also the use of kinship references for rhetorical purposes in examples, arguments, witnesses and indices of the speaker's social circumstances. Proceeding for this assumption, it may be argued that diachronic trends in the frequency and nature of kinship references can provide a useful window into how society, or a specific segment thereof, understood and related to family and gender. By the same token, the close association between the latter two concepts and the concept of *identity* means that frequency changes should allow new perspectives into the ways in which references to family have been used over time for indexing specific aspects of the speaker's identity. For example, an increase or a decrease in the frequency of references to spouses should reveal both changing societal attitudes to, and importance attached to, marriage, but also the speakers' willingness to make discursive use of their marital status.

To test this hypothesis, I will explore the frequencies and uses of kinship references in British parliamentary debates using the 1.6-billion-word Hansard Corpus, which represents the extant record of parliamentary debates, and trace the trends of kinship references in the record of the parliamentary debates over a timespan of two-hundred years. Approaching the issue from the sociolinguistic perspective and focusing on representation and gender, I will pay particular attention to changes in proportional frequencies and use statistical grouping methods to identify patterns that signal cultural shifts. Along the way, I will also address some philological and methodological issues to do with the Hansard, arguing that the correct interpretation of Hansard data requires awareness of the editorial history of the record.

2. Theoretical background

Although the primary objective of historical sociolinguistics may be defined as analysing how social factors have shaped language across the centuries, it is fair to say that much of the scholarship in the field has also been concerned with studying linguistic practices as reflections of and influences on contemporary societies. One of the first scholars to argue in favour of analysing developments in language and culture as a dynamic relationship was Hoijer (1948: 338), who argued that "[…] changes in language, since language is an important part of the cultural pattern, must take place, in part at least, in response to cultural changes in general". The present study falls in line with the emergent research paradigm of third-wave sociolinguistics that combines large textual datasets, macro-level and primarily corpus-based methodological perspectives with sociohistorical research questions (see Eckert 2018).

Sociolinguistic and discourse analytical studies of the dynamics between language and culture have traditionally focused on relatively small datasets, careful contextual analyses and detailed socio-biographical metadata (see e.g. Lakoff 2008: 167). Such studies, valuable as they are, often suffer from the somewhat anecdotal nature of their findings. While intense focus on the language of an individual, a speech event or a discursive dynamic will almost invariably reveal intriguing and informative facts about that object of inquiry, questions may be raised concerning the extent to which such findings can provide reliable new insights into broader diachronic trends.

In recent years, the introduction of digital corpora has opened up new windows into these longer tendencies by allowing us to identify broader patterns in language use and to use data-driven methods to identify turning points and particularly interesting examples that merit closer inspection (see e.g. Michel et al. 2010; Moretti 2013; Partington 2013). Importantly, the methods offered by the nascent field of Corpus-Assisted Discourse Studies (Baker 2006; Partington et al. 2013) should not be seen as replacing discourse analytical methods focusing on close reading and qualitative analysis in context, but rather as an opportunity to develop a common methodological toolbox. The work of the VARIENG school of researchers in Helsinki, and in particularly members of the CEEC team,[1] stands out in this respect. Another recent benchmark study in this field is Baker and McEnery (2016), in which the authors demonstrate the benefits of interdisciplinary research combining the methods of sociohistorical corpus linguistics and traditional research into social history (see e.g. Baker and McEnery 2016: 11–16).[2]

Most studies of gender in politics have tended to focus on gender as a factor affecting the way a person or group speaks or is allowed to speak, or else on how a specific person or group is spoken to or portrayed in a socio-temporal instance of political discourse (see e.g. Ferrari 2010). The present study addresses these

1. The Corpus of Early English Correspondence (CEEC) and its various later versions were compiled at the VARIENG research unit of the University of Helsinki. The members of the CEEC team – Terttu Nevalainen, Minna Palander-Collin, Minna Nevala, Anni Sairio, Mikko Laitinen, Tanja Säily, Samuli Kaislaniemi – have published a considerable body of work on historical sociolinguistics using CEEC.

2. In addition to purpose-built linguistic corpora, large textual archives have also proven to be a useful source, even if the interpretation of the data needs particular care. For example, Google Books data has been used to intriguing effects by various researchers despite the fact that the data is admittedly fuzzy (Diller 2012; Mäkinen 2013; Mäkinen and Tyrkkö 2014; Laitinen and Säily 2018; Tichy 2018).

concerns as well, but views them through the prism of corpus-based methods and "Big Data" in an effort to reveal how kinship references changed over a period of two-hundred years of British parliamentary debates.[3] Although parliamentary debates were not 'public' discourses in quite the same way we might understand that concept today,[4] I will argue that the overall trends observed in the Hansard Corpus nonetheless reflect contemporary societal values and changes thereof, to a reasonable degree. As Bayley (2004: 9) argues, "[t]he activity of parliaments is, or can be seen as, linguistic activity, and while we may question the centrality of the role of parliamentary interventions in the decision-making process […] parliament is the site where government and opposition go 'on the record', where justification and criticism of legislative initiatives is made, where interests are articulated and represented […]". By observing quantitative trends in kinship references in parliamentary language we can thus establish a baseline of evidence about how familial relations were understood and represented 'on the record'.

At the most fundamental level, the present study is concerned with the representation of kin in legislative discourse and in debates associated with the legislative process. In parliamentary debates such representation is predominantly by proxy, that is, the interests of a person, group or section of society are represented by and argued for by members of parliament, not the individuals or groups themselves. At the heart of such representation is the act of reference or, to put it another way, the act of explicitly mentioning those that are represented. While a person speaking for themself may be considered to self-represent simply by being present, the person being represented *in absentia* must be mentioned. These mentions can be quantified, and in a large comprehensive dataset, changes in frequency of reference can be interpreted as reflections of how important a particular person or group is. As McConnell-Ginet (2003: 72) notes, "referring is basic to conveying information: we refer to the people we talk about". While this may seem self-evident, the seemingly simple matter becomes arguably more urgent when we involve frequency of reference in the mix. By observing and measuring the frequencies at which members of societal segments are mentioned at various points in time, we can identify trends and turning points with an added level of accuracy.

A closely related question concerns the question of labels, or the manner in which reference is made (e.g. Pahta et al. 2010; Nevala 2011). As Alexander (2011 and 2017) has argued, sudden expansions or reductions in available lexis referring to a topic or concept can signal periods of cultural shift or so-called *linguistic trauma*. By the same token, significant changes in lexical frequency can indicate

3. As always, whether any linguistic data amounts to Big Data is debatable.

4. For discussion of how public parliamentary debates really were, see Section 3.

that a particular concept, topic or, as in the case here, a segment of society, have become more or less visible. Now, it would be naïve to suggest that any increase in reference is always and necessarily a positive development and a reduction in frequency a negative one – for example, legislative discourse can refer to groups of people for the express purpose of oppressing them and denying them rights (Nevala and Tyrkkö 2018) – but it seems reasonable to argue that the general direction of the political dynamic proceeds from visibility toward representation and, consequently, that measuring visibility through frequency of reference can be a useful way of studying attitudinal changes in society.

It goes without saying that human referents can and do have multiple roles and identities; in the case of kinship terms, a male person might be a son, a cousin, a brother, a husband, a father and a grandfather all at the same time (see Watts 1991). Likewise, not all husbands or mothers are created equal, and a detailed discursive study of occurrences in context could reveal substantial differences in, say, referential polarity when it comes to social contexts (see Eckert and McConnell-Ginet 2003). As Hamlett (2009: 577–578) points out, even seemingly innocuous settings (from our modern perspective) such as the home can be charged with presuppositions and cultural baggage, and thus a reference to a male of the Victorian era spending time at home could be interpreted as something emasculating and negative (see Tosh 1999; Harvey 2009). However, when we examine a comprehensive cultural resource, such as the parliamentary record, from a quantitative perspective, it may be argued that diachronic developments in the comparative frequencies of reference types, even if they may be somewhat superficial and wanting deeper investigation, may reveal important facts about how members of a particular superordinate group, such as of a gender, were viewed by contemporary society (see e.g. Wodak 2003; Sveen 2010). Were men mostly mentioned in their role as father or brother, were women primarily seen as mothers or as wives? As will be shown, these so-called *central terms*,[5] or terms that define a person or a group, have indeed undergone dramatic changes over the timeline under investigation (see e.g. Davis and King 2018).

It goes without saying that there are many reasons and occasions for using kinship terms, and only some of them are relevant to political debates in the Houses of Parliament. For example, as Tannen (2003: 181) points out, kinship terms such as *brother* and *sister* are frequently used in modern society as address terms indicating close social relations regardless of actual familial relationship. By contrast, while legislative discourses concerning family relations, the institution of marriage, inheritance, and a variety of other topics also give rise to the need to refer to male and female individuals according to their kinship roles, the discussion is

5. See e.g. Persson (1990: 36–37); Eckert and McConnell-Ginet (2003: 242–253).

typically more distant and instrumental. However, this is not to say that references of specific individuals could not at times invoke their kin also in the parliamentary context. This was particularly common in earlier times, when a person's standing in society was to a large extent determined by their family. Likewise, the use of narratives describing events that happened to or otherwise touched the relatives of the speaker or someone else can be used to strengthen an argument. Examples of these different contexts of kinship reference will be given below.

Finally, the fact that the present study focuses on a very specific context of discourse, namely British parliamentary debates, makes it particularly relevant also to be reminded of the *community of practice* perspective to language use (see e.g. Eckert 2000; Kopaczyk and Jucker 2013). As defined by Eckert and McConnell-Ginet (1992: 8),

> A community of practice is an aggregate of people who come together around mutual engagement in some common endeavor. Ways of doing things, ways of talking, beliefs, values, power relations – in short, practices – emerge in the course of their joint activity around that endeavor. A community of practice is different as a social construct from the traditional notion of community, primarily because it is defined simultaneously by its membership and by the practice in which that membership engages.

Considering how the Parliament was for a very long time a male-only body comprising men of more or less the same age, same ethnicity and similar background, we may assume that the attitudes expressed about gender roles as well as styles of referring to kin were for the most part closely regulated by the implicit rules of the community. Furthermore, as will be discussed in detail in Section 3, in the case of the Hansard the community of the Parliament was also regulated by two other successive communities of practice, namely journalists and parliamentary transcribers. In the present context, the examination of the timeline of parliamentary debates as an evolving community of practice allows us to identify turning points and emerging practices as more than idiosyncratic expressions of individual opinions and instead as expression of attitude about gender in a culturally and societally significant public discourse.

3. The Hansard Corpus as a linguistic resource

The Hansard Corpus is a linguistic corpus, compiled by the SAMUELS project at the University of Glasgow, which comprises the entire record of British parliamentary debates from 1800 to 2000. At 1.6 billion words, more than 7.5 million speeches and 40,000 individual speakers, the Hansard Corpus is a rare example of a complete dataset, that is, rather than providing a sample drawn from a population of texts too large to study directly, it provides a complete record of all parliamentary debates during the timeline. This means that the Hansard provides us with an account of what was or was not discussed in parliament at any given time, and how such discussions were carried out. Furthermore, because the corpus represents a population rather than a sample, if observations are made with the intention of analysing cultural, political or discursive topics, there is no need to carry out inferential analysis, because any frequency differences observed are true differences in the population. Naturally, if data from the Hansard Corpus were used as a sample of wider contemporary political language use, conventional measures of inferential analysis apply. However, while the Hansard Corpus is indeed an invaluable resource for studying political, cultural and linguistic phenomena, there are also several historical reasons for why we ought to be particularly careful when using it.

First, let us consider the various methods by which the parliamentary debates were recorded for posterity and what the historiography of the Hansard tells us about the reliability of the record. The Hansard, or more formally the Parliamentary Debates in Britain, is named after printer and publisher Thomas Curson Hansard (1776–1833), the first official printer to the Parliament and the founder of the record. In the beginning, publishing information about parliamentary debates was against the law. As the newspaper industry developed, unofficial reports were circulated at increasing rate during the eighteenth century and public interest in the debates increased all the time. Reports of parliamentary debates were published with changed names and in various pseudo-fictional settings, until newsman William Cobbett began to publish reports of the debates under the title "Parliamentary Debates" as a regular section of the newspaper *Political Register* in 1802. Before long, Cobbett was brought up on charges and sentenced to a prison term and a heavy fine for sedition. Thomas Curson Hansard, the printer of Cobbett's publications, was also imprisoned for a short period. In 1812, Hansard took ownership of Cobbett's *Parliamentary Debates*, publishing it thereafter under the title *Hansard Parliamentary Debates*. During the first decades of the nineteenth century, the political climate was changing and there was a growing sense that there was a legitimate public interest to publish parliamentary debates (Rix 2014: 454). The need for a transcribed record of the debates was confirmed in the House of Commons in 1834, and the rules were relaxed which essentially made it legal for reports of the debates to be published. Hansard continued to publish the record until 1909, at which point the record was fully taken over by the Parliament.

This brief history of the Hansard is a necessary backdrop to understanding how reliable the Hansard is as a record of parliamentary debates. Up until 1878, the Hansard was compiled from various journalists' reports on the debates. Crucially for both historical and linguistic accuracy, these were rarely transcripts of actual debates, and instead reports that highlighted the major topics discussed, sometimes quoted parts of debates, gave background information, included parts of bills, and so on. In the early decades, the Hansard record was compiled not by professional parliamentary transcribers, but by journalists and others who took notes in parliament. The entries were compiled from published newspaper reports and journalists' notes, which could be sent to the politicians for approval prior to publishing. Although this in part made the publishing of the records more acceptable to the members of parliament, it also means that the record is by no means a reliable source of the actual language of the debates (Mollin 2007). From 1878, Hansard was able to employ journalists to attend parliamentary sessions and consequently the quality of the record improved. The Hansard's website quotes the decision of the House of Commons Select Committee in 1893 that the parliamentary record is to be "[…] substantially the verbatim report with repetitions and redundancies omitted and with obvious mistakes corrected but which, on the other hand, leaves out nothing that adds to the meaning of the speech or illustrates the arguments". However, according to MacDonagh (1913: 20), parliamentary language was often difficult for reporters to transcribe accurately, "not because of the speed, but its 'colloquial mode and subdued intonation'". Other contemporary sources reveal that the parliamentary reporters were also likely to ignore unimportant or uninteresting discussions and to edit their reports to improve the members' language, sometimes at their request. Furthermore, according to Mollin (2007: 192–193), even the more recent Hansard records are far from full transcriptions, with most so-called performance features of spoken language having been edited out by the reporters; Mollin notes a reduction of 18% of the authentic word count due to such measures (see also Slembrouck 1992).

As noted earlier, the year 1909 marked an important milestone in the democratic practices of the British Parliament. That is, when the Parliament commenced the policy of producing in-house transcriptions by parliamentary reporters, recording for the first time the actual spoken dialogue from the two Houses. Apart from the impact that this change in practice had on transparency and accessibility, the linguistic effects of this on the Hansard record are considerably more profound than what has generally been noted in previous scholarship. As discussed by Hiltunen et al. (2018), the switch from reported speech and descriptive narratives to transcribed spoken language marks a clear text typological shift, which is evident on every level of language use.[6] In simple terms, the early Hansard is

6. Using a pattern-driven approach to analysing typological change, Hiltunen et al. (2018) show that if the frequency trends of the 1,000 most frequent 3-grams are analysed over the period 1870 to 1930, the 3-grams are split neatly into two distinct groups around the year 1910. On pattern-driven methods in general, see Tyrkkö and Kopaczyk (2018).

essentially written newspaper language written in appropriately written register, while the post-1909 Hansard features spoken register, albeit in the formal setting of the British parliament.

In the present study, the author used a full copy of the Hansard Corpus generously made available by the SAMUELS team. The corpus was provided as a nested hierarchy of folders divided into the House of Commons and House of Lords, and then by decade, year, month and day. Within the day folders were XML-annotated files each representing an individual entry in the Hansard. Like most corpora these days, the Hansard Corpus is part-of-speech annotated, but in addition to that it is also semantically tagged using a tagset based on the Historical Thesaurus of the Oxford English Dictionary (HTOED) and a purpose-built semantic tagger developed at the University of Lancaster.[7]

The decade by decade word counts of the Hansard are given in Figure 1. As the figure shows, the word counts are very low during the first three decades of the nineteenth century; when indicated, these were left out of the analysis to avoid skewed results.

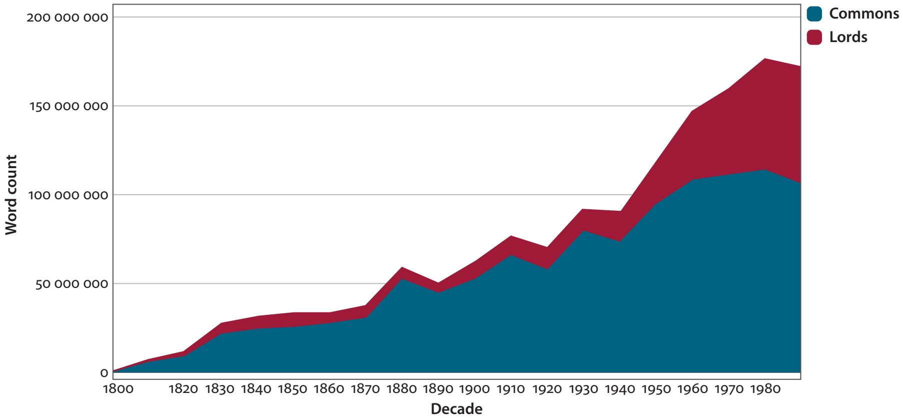

Figure 1. Word counts by decade, both Houses combined

It is also important to note that although the per-decade word counts may generally seem to be great enough to be considered intuitively representative, when it comes to fairly rare items like kinship terms, the frequencies of individual items are not that high. This can mean that the relative frequencies can be quite dramatically

7. Semantic tagging allows the user to run queries for all items that belong to a semantic field, as determined by the HTOED, alone or as part of a more complicated query. For example, in the present study there were a few occurrences of the Scots word *bairn* (=child) in the Hansard debates, which would most likely have been left out if a wordlist of kinship terms had been compiled manually.

affected by individual debates or topics. Perhaps the most striking example of this is the so-called Queen Caroline Affair, or the divorce of the newly crowned King George IV from his wife, Caroline of Brunswick in 1820. The extended debates in the House of Lords of the Bill of Pains and Penalties, which would have effectively annulled the marriage, is carefully recorded in the Hansard Corpus. The debate, which was a form of trial, includes numerous witness statements and narratives, which among other things include 77 instances of the phrase *my sister*, all of them in the interview of one speaker. This makes the frequency of *my sister* in the 1820s much higher than it is at any time in the nineteenth century, but it goes without saying that the use of the phrase by a single speaker in one interview should not be interpreted as an indication that the discussion of sisters was in general terms a significant feature of early nineteenth-century parliamentary debates. Luckily, however, this particular event appears to be something of a unique outlier in the data.

4. Operationalization of the queries and analyses

The kinship terms were identified using the tag semantic tagging of the Hansard Corpus; the HTOED category 03.01.01.03, "kinsman/relation", retrieves all items in the various subcategories, such as "parent", "child", "sibling", and so on. Because the definition of kinship in the present study is restricted to family relations, however loosely defined, some manual pruning was carried out after the initial queries to rule out references to words such as *ancestor, cadet, original, legitimate, friend, friends, allies, fostering*, etc.

In order to focus on references to kinship terms that occur in discourse and not merely in legislative texts, the focus of the present study was narrowed down to occurrences of the relevant nouns when pre-modified by a possessive pronoun.[8] Using the main kinship terms (*husband, wife, father, mother, son, daughter, brother, sister*) as an example, this narrows down the number of observations from 264,319 to 88,343. Adjectival pre-modifiers interceding between the possessive pronoun and the kinship terms are very rare in the data (ranging from 0 to 4% of raw frequency depending on the item) and they were discarded to keep the operationalization as simple as possible. It should also be noted here that references to same-sex partners are generally not included in the study unless they take the form of POSSESSIVE PRONOUNS + KINSHIP TERM and the noun in question is among the list of included items. Registered civil partnership between partners of the same sex

8. For discussion of shifts in pronoun frequencies over time in political discourse, see Tyrkkö (2016); for discussion of discursive perspectives to pronominal reference, see Mühlhäusler and Harré (1990).

became legal in Britain in December 2004, right at the end of the corpus timeline, and same-sex marriage only in 2013, which means that most earlier references to same-sex partners would have involved generic nouns such as *partner, significant other, friend*, etc., which were excluded from the study.

The calculation of standardised frequencies was carried out based on word counts calculated for Hiltunen et al. (2018). The counts included only words, not punctuation, which results in slightly lower 'word counts' than what it given on the online Hansard tool. In the visualizations, decade was deemed to be the appropriate level of granularity. The hierarchical cluster analysis used in Section 5.2 is a commonly used method of exploratory data analysis in which similarity clusters are identified from a multivariate data matrix; the agglomerative clustering algorithm used in the present study is Ward's method using Euclidean distances. Each variable (here, unique phrase type) is standardised across the timeline independent of the other variables, with the effect that all variables have equal weight in the model.

5. Findings

5.1 Overall trends

The simplest and most fundamental frequency trend is the overall frequency of references to POSSESSIVE PRONOUN + KINSHIP TERM in the Hansard data (Figure 2). For clarity, only the core kinship terms of *mother, father, wife, husband, sister, brother, daughter* and *son* are included in the figure; all the kinship terms were included in the figures below. The height of each stacked bar represents 100% of all the relevant references, and the proportion of each term is indicated by the height of the respective section of the bar.

The figure shows that references to *wife* peak in the 1930, accounting for nearly 40% of all the kinship terms. References to *husband* peak at the same time, but they account for only 23% of the total. References to *mother* are relatively few in the nineteenth century and into the twentieth century, but they start increasing around 1930. By contrast, references to *father* appear to have remained essentially unchanged throughout the entire timeline. Interestingly, both *son* and *brother* were relatively common referents during the nineteenth century, but both declined in frequency around 1920, hitting a plateau at that time. References to *daughter* were historically very rare, as were references to *sister*, but while the former has undergone a slight increase over the last few decades, the latter remains, as it always has, a very infrequent referent in parliamentary discourse. The one obvious outlier in the frequency of *sister* is the aforementioned Queen Caroline Affair in the 1820s (see Section 3).

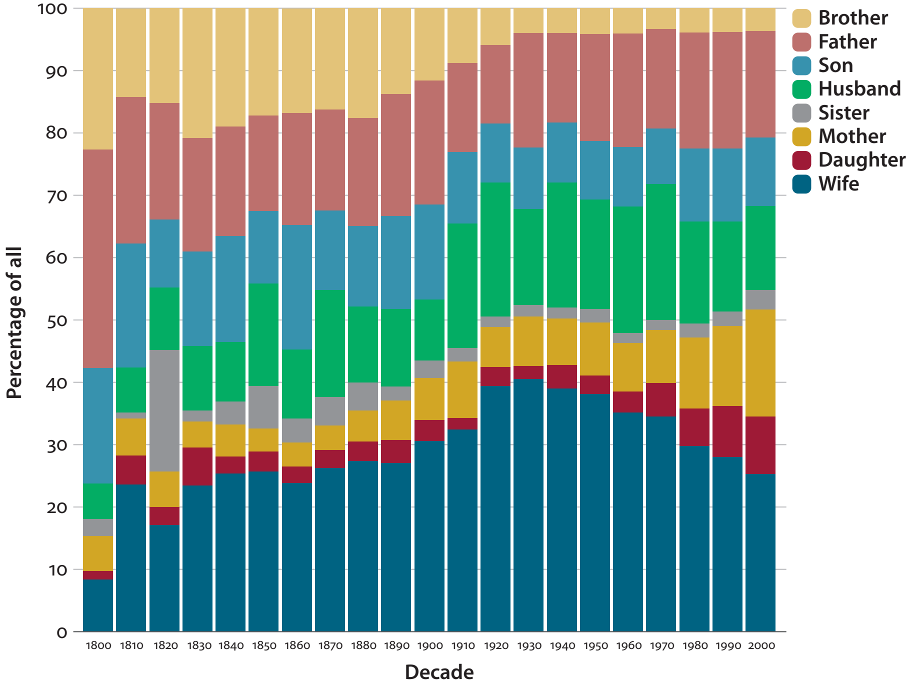

Figure 2. Stacked bar chart of the phrase *my* + KINSHIP TERM

The sociolinguistic and sociohistorical implications of these reference frequencies seem quite clear. The aggregated proportion of references to female kin begins at around 35 percent in 1810, climbs steadily and reaches 50% around 1920, and remains there for the remainder of the twentieth century. However, keeping in mind the changes in the editorial realities that underlie the Hansard record, we must consider the possibility that the apparent under-representation of women may also have been affected by the way the topics were discussed in newspapers. Since all the speakers at the time were inevitably men, and the common style was to refer to a man's male kin when describing him and his background, it is possible that some of the proportional difference may be the result of journalistic practice.

Figure 3 shows the trends in collective, female and male kinship terms. Every variable shows a clear trend here: the frequencies of collective and female nouns increase throughout the timeline and particularly after 1910, and the frequencies of male nouns decline, appearing to settle in a plateau after 1910.

The possessive pronouns used for identifying discursive usage of the kinship terms also serve as a useful indicator of the nature of the discourses in which the term occurs. Figure 4 gives the frequency trends of the six pronouns; note that the frequencies given only include instances where the possessive pronouns have a kinship term as their R1 collocate.

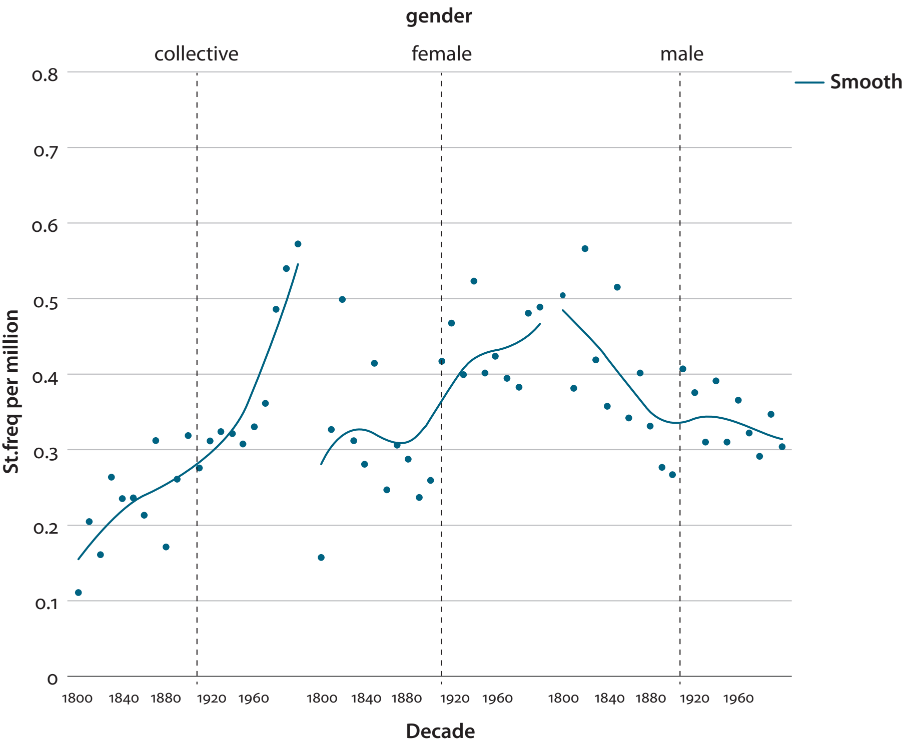

Figure 3. Trend by gender category in phrase structure POSSESSIVE PRONOUN + KINSHIP TERM

As Figure 4 shows, distinct trends can again be observed. The frequency of *his* shows a notable decline, the frequencies of *my* and *their* show strong to moderate inclines, and the frequencies of *our, her* and *your* are relatively stable. Given that the change in the nature of the parliamentary record took place in 1909, we might have expected to see a greater shift around that time, but that appears to be restricted to one pronoun. The extremely steep incline in the first person singular *my*, which begins at exactly 1910, is particularly interesting, because it relates directly to the new transcription practice. While two other pronouns, *his* and *their*, also appear responsive to the commencement of transcribing, *his* going into almost immediate decline and *their* going into an incline, neither shows as strong a trend as *my*. Although MPs may have discussed members of their family before 1909 in equal measure to thereafter, the specific possessive formulation *my* + KINSHIP TERM made it into the Hansard only rarely, typically if a particular entry happened to report on direct speech.

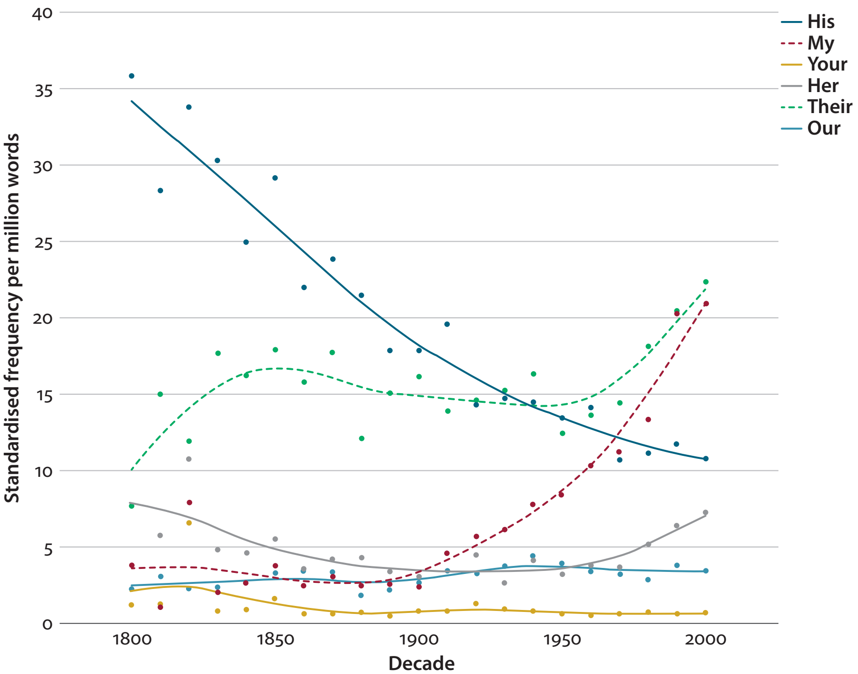

Figure 4. Trend by possessive pronoun in phrase structure POSSESSIVE PRONOUN + KINSHIP TERM

5.2 Cluster analysis and some significant trends

Hierarchical cluster analysis can be used to explore patterns in the data by grouping together decades based on similar distributions of the kinship terms (Figure 5).[9] As Figure 5 shows, the decades of the nineteenth century, from the 1830s to 1900s, form their own cluster, the decades at the turn of the century (1910s to 1950s) form a second cluster and the decades of the latter half of the twentieth century (1960s to 1990s) form a third cluster. The order of decades on the left-hand side of the visualization is not predetermined, but rather the result of the clustering method. The branching of the cluster dendrogram shows us that the most significant shift takes place around 1950. In other words, although we can detect meaningful and culturally significant changes along the way when it comes to individual phrase types, the overall cultural shifts are most notable around the turn of the century and 1960. Looking at the heatmap, which indicates the phrases that are particularly common (red) or particularly uncommon (blue) at any given time period, we can also see that there are several clusters of features that bear closer examination.

9. We discard the first three decades from the model because of the low overall word counts, which will produce somewhat skewed standardized frequencies.

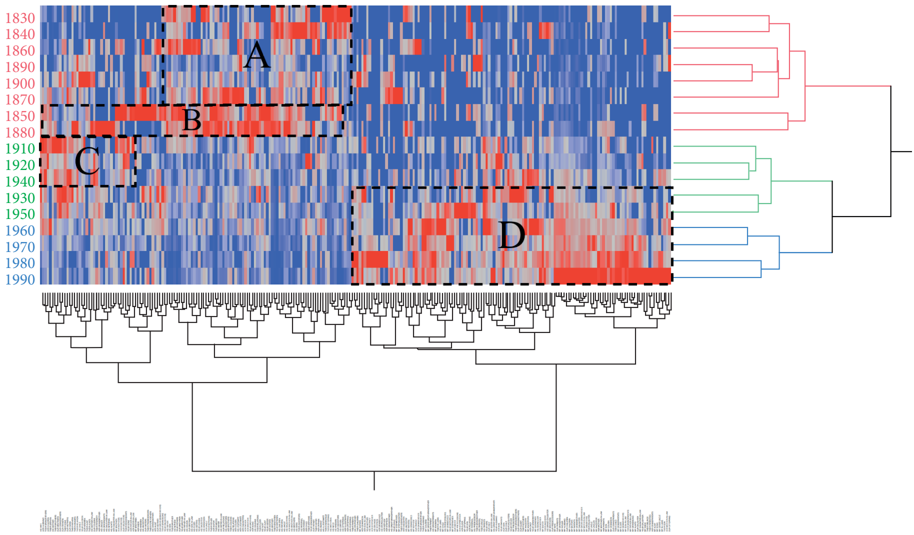

Figure 5. Hierarchical cluster analysis of all phrases with the structure POSSESSIVE PRONOUN + KINSHIP TERM

In the first cluster, (1830–1900), there are two high-frequency areas. The first, marked A, shows high frequencies, while the other, marked B, shows very low frequencies. Area A is high in phrases related to male kin – *father, son, fathers, brother, uncle, brother-in-law* – as well as younger female kin: *daughter, sister* and *niece*.[10] In nearly all of these the possessive pronoun is *his* or *their*, which is in line with the fact that the Hansard record at the time consisted of journalists' reports and reported speech. Conspicuously absent are any references to *wife* or *mother*. The frequency trends concerning *brother* and *father* mirror each other fairly closely until the 1920, showing a consistent decline in the *his* + *father/brother* pattern that extends all the way to the present day. The *my* + *father/brother* pattern remains relatively infrequent until 1910, after which references to *my father* begin a rapid increase. The phrase *my brother* remains infrequent.

The decreasing relevance of *brother* in parliamentary discourse is quite striking (Figure 6). In the nineteenth century, references to brothers (and fathers) were often offered as important background information when a person is discussed, presumably because a person's familial and social connections were an important determiner of their reliability and character. In Example (1), Mr. Isaac Butt discusses a member of Russian aristocracy, giving information on his brother and his close connections with Tsar Nicholas I.

10. For a complete picture, all the kinship terms, not merely terms referring to the nuclear family, were used in the cluster analysis.

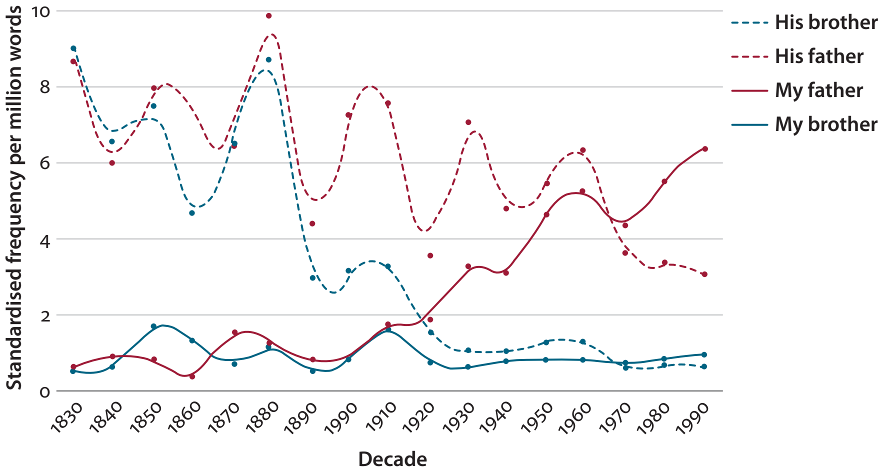

Figure 6. Most frequent phrases containing *father* and *brother*

(1) The gentleman to whom I allude is a member of a very distinguished Russian family, his name is Count Pahlen; he has been engaged in the Russian diplomatic service, and *his brother* occupies at this moment a distinguished position in the Council of the Emperor of Russia, and is said to be on the most intimate and friendly terms with him: Count Pahlen has arrived within the last ten days in this city, and since his arrival he has mixed freely in society, and it is stated that he has been introduced into some society under the special patronage and auspices of one of Her Majesty's Cabinet Ministers:

(1854, Butt, Count Pahlen – Russian subjects in London, Commons)

The horizontal bands of high frequency (area B) in 1850 and 1870 include similar kinship terms paired with the pronouns *my* and *our*, while the short band in 1880 (right-hand side of heatmap) contains references to more distant relatives in plural, such as *cousins, daughters-in-law, nephews and aunts*, typically paired with *his* or *their*. These clusters tell us that nineteenth-century kinship references were primarily used in legislative discourses that concerned family relations, such as inheritance laws and marriage.

The ongoing debate about the legality of marriage between a man and his deceased wife's sister, first sparked by the Marriage Act of 1835 and reinvigorated by the defeated Marriage to a Deceased Wife's Sister Bill of 1842, initiated a recurrent series of debates that involves high frequencies of reference to the otherwise rare item *sister*. Example (2) from 1889 illustrates the ongoing debate:

(2) Suppose *father* and *son* married two *sisters*, and suppose the *father* died and the son's *wife* died, would the *son* be entitled to marry his *mother-in-law*? Or, on the other hand, could the *father*, on the death of his *wife* and his *son*, marry his *daughter-in-law*: Further it would be quite possible for a man to marry his own *niece*: It is because of these inconsistencies, and because there is no favourable expression of opinion, that the Bill is required in Scotland, that I move the Bill be read a Second Time this day six months.
(1889, Macintosh, Marriage with a deceased wife's sister bill, Commons)

As we move down to the second cluster, 1910 to 1950, we note that there are two subclusters: the decades 1910, 1920 and 1940 do not appear to feature any phrase types at particularly high frequencies, while the decades from 1930 to 1950 show a small high-frequency area (C). Area C shows high frequencies of *mothers, brothers, sons, fathers* and *grandfathers*, as well as the first higher frequency references to *wife* and *husband*; the possessive pronouns are predominantly *his, our* and *their*. The discourse appears to have shifted away from younger female kin. Bearing in mind that the colour scheme goes from red for high frequency to grey for medium frequency to blue for low frequency, the treemap shows us how the previously high-frequency items in A are 'cooling off' during this time. Example (3) illustrates the newly emerging discourses concerning family and parenthood, which begin to question traditional values:

(3) It is the tyranny of motherhood, not only economically but psychologically, that has caused revolt on the part of countless numbers of women: They want some measure of freedom, not to give less but to give more to their children, and to develop their own minds as well as their bodies: More women are saying, "Say what you please, we are determined to have greater fulness [sic] of life than *our mothers* or *grandmothers* had," not merely for selfish reasons, but often for really unselfish reasons: When we hear about small families arising from human selfishness, does it occur to those who say that to ask how many people have children for purely unselfish reasons? How many have them because of this abstraction which we call the State? An insignificant number: They have them because they want them or because they can not stop them. (1943, Sorensen, Trend of population, Commons)

References to *her husband* and *his wife* show strongly correlated trendlines, which suggests that the two are typically discussed in the same legislative discourses (Figure 7). The strongly oscillating pattern further supports the interpretation that these two phrases occur primarily, if not exclusively, in debates about Bills that concern marriage or some other context in which spouses are discussed in relation to each other. In this rather disturbing passage from 1925, the debate concerns a husband's responsibility for his wife's debts and torts. In the course of a long speech, Lord Darling argues that a husband cannot be held responsible for his wife crimes because husbands are no longer legally permitted to administer physical punishment to their wives (Example (4)):

(4) Blackstone puts it that a man is allowed to beat *his wife* because he is responsible for her torts. How then is it reasonable that he should be still held responsible for the torts she commits, and yet is not allowed to administer modicam castigationem? The law has decided that he may. Opinion grew upon this subject and Blackstone tells us this: – "But with us, in the politer reign of Charles the Second, this power of correction began to be doubted: and a wife may now have security of the peace against *her husband*: or, in return, a husband against *his wife*. Yet the lower rank of people, who were always fond of the old Common Law, still claim, and exert their ancient privilege." This ancient privilege having been taken away it is a well known maxim of the law: cessante ratione cessat et ipsa lex. The law having taken away the power of the husband to see that *his wife* does not commit these torts by administering to her that castigation which he was accustomed to administer till the politer age of Charles the Second began, it seems to me perfectly logical, and I am sure that I could convince any woman of it, that, as the husband might beat her because he was responsible for the mischief that she did, so he ought not to be responsible for the mischief she does now he is no longer allowed to beat her.

(1925, Darling, Status of married women, Lords)

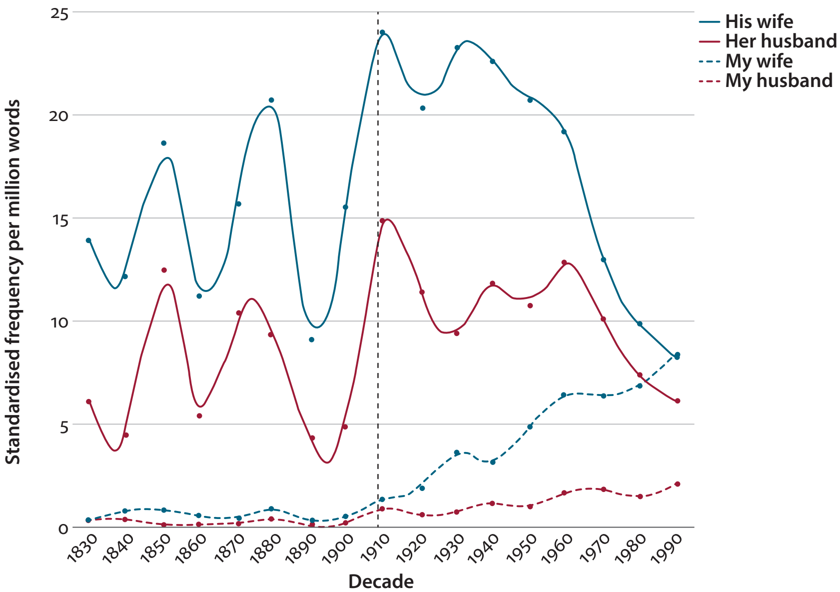

Figure 7. Most frequent phrases containing *wife* and *husband*

By contrast, references to *my wife* and *my husband* are both very infrequent during the pre-1909 period of the Hansard and begin to increase almost immediately after the Hansard record begins to reflect the actual spoken debates. While a good proportion of the nineteenth-century references to "my wife", as well as nearly all of the contemporaneous references to "my husband", occur within quoted or reported passages, there are some early instances of reported direct discourse, that is, spoken turns in which the speaker refers not to a story in which a third person mentions their spouse, but instead refers to their own. In Example (5), the assistant of an attorney has been accused of aiding his employer and is now facing criminal charges. Answering the questions of an MP, Mr. Law, the accused Mr. Pearce appeals to the House referring to his wife's background:

(5) Mr. Law: Have you a family?
Mr. Pearce: I have a wife and son, and I have no other means to support them except my labour. Besides, I have been a military man all my life, and I have learned never to break the law, or disobey my orders.
Mr. Law: And your family have no means of support if you are imprisoned on this charge?
Mr. Pearce: No. Imprisonment would be my ruin, and theirs also. I beg to tell the House, that *my wife* is allied to some of the oldest nobility, some of the first nobility in this country – I may say the first and oldest in this country – a family the patent of whose nobility is as early as 1616. She is the daughter of the hon. Philip Roper, and is first cousin to Lord Teynham. And as she is an outcast from her family for marrying a soldier, she would be totally destitute if she had not me to provide a subsistence for her. (1840, Law, Mr. Howard, Commons)

Finally, as we move to the third and final cluster, 1950 to 2000, we see that the area D appears red, indicating that the phrase types in this group are now getting relatively much use while the other areas have turned blue indicating less frequent use. It is also clear to see that nearly all the items in area D featured low frequencies in the nineteenth century. The main commonality in these items is the use of the first person singular possessive *my* and the first person plural possessive *our*, which are markers of direct speech in this context. The items toward the left-hand side of area D include kinship terms related to the parent-child relationship – *parents, parent, children, son, grandmother* – as well as, for the first time, references to *stepson, stepfather, stepmother* and *stepdaughter*. The long band of red down the bottom and middle of the treemap during the 1980s and 1990 is strongly associated with the first person possessive pronouns *my* and *our*, especially *my*.

Toward the end of the twentieth century, first person kinship references are increasingly about the MPs' own family members. This is a significant development, which goes hand in hand with a much less guarded attitude toward private topics. In Example (6) from a discussion of infertility treatments, the MP for Canterbury responds to Anne Widdecombe by giving a personal anecdote, and in Example (7), William Wallace discussed his daughter's experience of supporting a gay friend:

(6) Mr. Julian Brazier (Canterbury) *My wife* and I had a serious fertility problem, as a result of which we went through a successful IVF process. There is considerable evidence that the IVF success rate does not depend on destructive experimentation. Statistics in this country and evidence from IVF clinics in South Australia, where experimentation is illegal, point to the fact that it is the number of times that the procedure has been carried out that is the key factor determining the success rate, not destructive experimentation.
(1990, Brazier, Human Fertilisation and Embryology Bill, Commons)

(7) In the past four or five years, a friend of mine and a relative of mine have been through this experience – one with a young man and the other with a young woman – and I know something of the slow and personally painful discovery of people in their teens that they are different from the majority of their fellows. In neither case was there any 751 seduction and certainly not any promiscuity; rather a gradual realisation that that was who they are. One of them grew up in South Africa and *my daughter* went with him on the first Gay Pride march in Cape Town, at which he came out. It was for her a very emotional experience. I am glad she did it.
(1999, Wallace, Sexual Offences Amendment, Lords)

Another common function of modern references to wives in Parliament is what could be described by the colloquial term a "humble brag" (Example (8)).

(8) Much more recently, I have had experience in my Oxford capacity of the great benefit of having the Queen Mother as patron of our development campaign. It so happens that *my wife* is chairman of the National Trust and she also has had the privilege of the Queen Mother as her president. I therefore have fairly direct experience of the infusion of enthusiasm and commitment which the Queen Mother brings to two disparate organisations. We express our deep affection and respect: We hope that she may continue to enjoy life and that we may continue to enjoy the immense benefit of her service and inspiration to the nation for many years to come.
(1990, Lord Jenkins of Hillhead, Her Majesty The Queen Mother: Humble Address, Lords)

References to wives' occupations and positions of trust started becoming common in the 1970s, and when mentioned they are almost exclusively positive. The most common R1 collocates of *my wife is* [*a/an*] are: *chairman, president, patron, chair, governor, founder, director, member* [of a society, trust or club], *magistrate, doctor, teacher, trustee, director, councillor, GP, justice*, etc.

However, it is also important to note here that even in the latter half of the twentieth century, many of these references, though occurring in debates, are not references to the speaker's own kin, but rather they are indirect discourse reported by the speaker. A very common type is references to letters received from the MP's constituent or to a discussion the MP has had (Example (9)).

(9) I should like to read one from a lady, who, I believe, speaks for many in a similar position: Her husband was made redundant. She states: "From speaking to many people in this area, I feel we could have a revolt unless something is done to help the unemployed. *My husband* was made redundant after 22 years". She then gives me details of his redundancy and his work: "However, the despair and heartache is written over the faces of the people in the North, because to take away a man's right to work and support his family takes away his pride. There are also our youngsters, leaving school without hope for the future: No wonder a lot of them turn to crime: […]
(1981, Sheldon, budget_resolutions_and_economic_situation, Commons)

There are also specific kinship phrases that are employed particularly frequently for rhetorical purposes and, due to the nature of the Hansard record, such devices will generally become more visible in the data after 1909. For example, the noun phrase *our children* shows a steady positive cline all the way from the beginning of the timeline, but then begins a much steeper cline from 1910 onward (Figure 8, Example (10)).

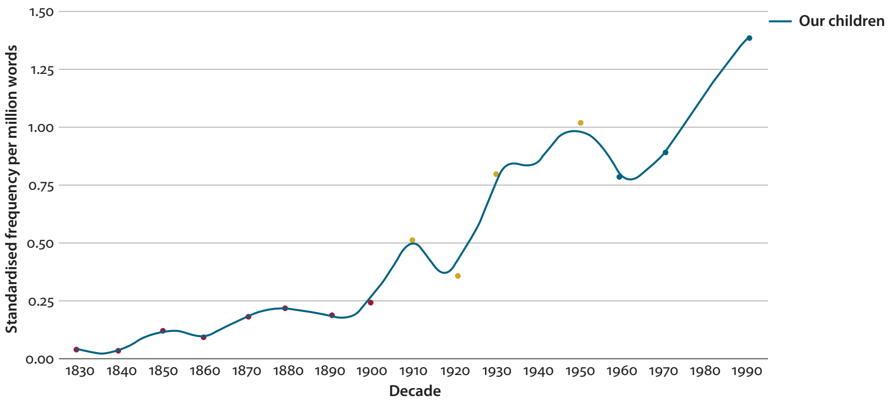

Figure 8. Trendline for *our children*

(10) If we are to have industrial development – and, heaven knows, we need it in order to provide the jobs – let it be accompanied by a little wisdom and foresight. Let us ensure that our descendants in the 21st century do not rebuke us in our graves, as we have so often rebuked those who went before us. Let us do our utmost to see that what we leave for *our children* and our children's children is a land at least as fit and clean as it was when we came into it. Indeed, let us do more and make sure that it is a better one: […]
(1970, Walker, Environmental pollution, Commons)

Another noteworthy and thematically related collective kinship term is *their parents*, which undergoes a very similar increase over the timeline (Figure 9). The timeline suggests that something was going on in British society during the 1960s that gave rise to the increasing need to discuss the parents of adolescents. The phrase is typically found as a collocate of *pupils, young people, teenagers* or *children*, and most often concerns issues to do with parents' responsibilities either for their children's welfare or behaviour (Example (11)).

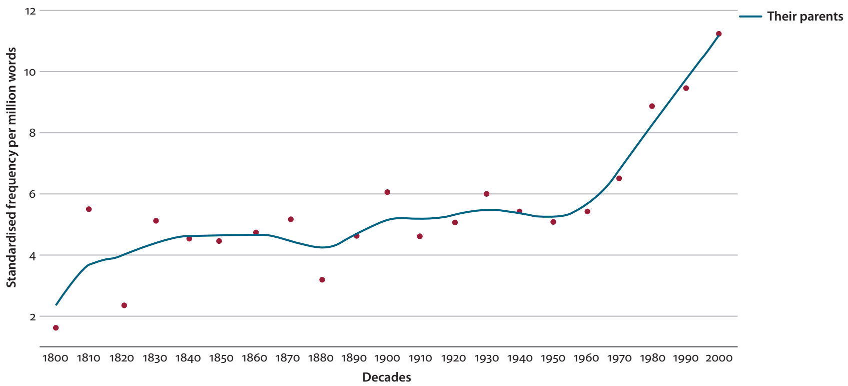

Figure 9. Trendline for *their parents*

(11) All sorts of groups of people are excluded: I entirely agree with the hon. Lady that family life is very important. It is good, proper, right and desirable that young people talk to *their parents*, discuss matters with them and receive advice from them. Unhappily, it is not true of all of them. The Bill will damage the fabric of many family relationships, because the confidentiality which the young person wishes to preserve will be destroyed and the knowledge which he does not wish his parents to have will automatically come to them. This could result in breakdowns of family life:
(1971, Lestor, Family Law Reform Act, Commons)

6. Tentative conclusions

This chapter was an exploratory pilot study into kinship and gender references in British parliamentary debates. Using the Hansard Corpus, the definitive linguistic resource on British parliamentary language, I examined the diachronic trends of kinship references over a period of two-hundred years demonstrating that both the frequency and functions of such references underwent almost constant but, significantly, systematic and unidirectional change. Over the timeline, references to female family members increased both proportionately and in absolute terms, reaching equilibrium with references to male kin by the early twentieth century. At the time, a number of significant changes had taken place in British society, such as shifts in gender relations, women's increased participation in the workforce, and the general organization and role of the nuclear family. According to the theoretical framework known as the individualization thesis (Charles et al. 2008: 2–15), the decreasing societal importance of family, a process that took place from the late Victorian era onwards, led among other things to a loss in the social capital associated with familial and social networks. In the Hansard data, this effect can arguably be seen in the proportional decline in references to male kin, whose discursive function was mainly to situate the speaker in relation to his most immediate network, namely brothers and fathers. Cluster analysis was employed to demonstrate further that there were two significant turning points in the overall use of kinship terms, the first occurring in the early twentieth century and the second around 1960. The first turning point is characterized by the final stages of the relatively rapid growth of the middle class, a process that raised public and political awareness of many issues to do with the family and afforded them more urgency. For example, as marriages across traditional class-lines became more common, the insular cultures of the upper and upper-middle classes, from which MPs had traditionally been drawn, gave way to a more heterogenous societal structure (see Miles 1999: 163–165). The second turning point likewise co-occurred with important societal developments, culminating in the Social Revolution of the 1960s as the post-war generation came of age and modern youth culture was born (Head 2002: 124–135). Here, the increase in the first person singular and plural possessives collocating with kin terms speaks volumes, as they highlight the MPs' growing willingness to establish their identities as individuals and specifically as men and women with families and personal experiences instead of as distant legislators.

Individual kinship terms also demonstrated trajectories of change, perhaps most notable being the increase in references to *wife* and the decrease in references to *brother*. In general, the findings show that the representation of female family members has grown more frequent and nuanced. The trends concerning pronouns and kinship terms suggest that members of parliament increasingly discuss their own family members, a development that goes beyond being a mere artefact of the record-keeping practices. What these shifts indicate is a transition from a detached and distant discursive style to a style that presents the MP as a person involved with and attached to their nuclear family and perhaps especially their spouse, and from a historical perspective they serve to indicate the relatively recent willingness of MPs to do away with, or at least lessen, class differences.

It goes without saying that the observations reported in this study are merely a starting point to more comprehensive analysis. For one thing, little attention could be paid to possible differences between the two Houses of Parliament or to political affiliations, let alone the habits and tendencies of individual speakers, or the topics of the specific debates. Another area for future studies is a detailed functional analysis of the syntactic roles of the kinship and gender referents, of their semantic roles, and of the verb phrases of which they function as subjects or objects. More detailed methods and closer contextual reading will be needed at the next stages of the study, but the results presented here are encouraging, suggesting that interesting and sometimes surprising findings about kinship and gender can be discovered using pattern-driven approaches.

Acknowledgements

This paper relates directly to the ongoing work of the DEMLANG research team at the Universities of Tampere and Helsinki, and indirectly Linnaeus University. The research presented here would not have been possible without access to the standalone version of the Hansard Corpus, generously provided by the SAMUELS team at the University of Glasgow; I am particular indebted to Marc Alexander and Fraser Dallachy. I also wish to acknowledge the constructive suggestions of two anonymous referees and the discussions I have had with Jenni Räikkönen, Turo Hiltunen and Minna Nevala of the DEMLANG team.

References

Alexander, Marc. 2011. "The Various Forms of Civilization Arranged in Chronological Strata: Manipulating the HTOED." In *Cunning Passages, Contrived Corridors: Unexpected Essays in the History of Lexicography*, ed. by Michael Adams, and Giovanni Imartino, 309–322. Monza: Polimetrica Press.

Alexander, Marc. 2017. "Profiles and Peaks: Lexicalisation, Focus, and Trauma in the History of English." Plenary presentation at HEL-LEX 5, Zurich, 15–19 February 2017.

Baker, Paul. 2006. *Using Corpora in Discourse Analysis*. London and New York, NY: Continuum.

Baker, Helen, and Anthony McEnery. 2016. *Corpus Linguistics and 17th-Century Prostitution*. New York: Bloomsbury.

Bayley, Paul. 2004. “Introduction: The Whys and Wherefores of Analysing Parliamentary Discourse.” In *Cross-cultural Perspectives on Parliamentary Discourse*, ed. by Paul Bayley, 1–44. Amsterdam: John Benjamins. https://doi.org/10.1075/dapsac.10.01bay

Bryant, Leah E. 2016. *The Wiley-Blackwell Encyclopedia of Family Studies*. *Wiley Online Library*. https://doi.org/10.1002/9781119085621.wbefs490

Charles, Nickie, Charlotte Aull Davies, C. C. Harris, and Colin Rosser. 2008. *Families in Transition: Social Change, Family Formation and Kin Relationships*. Bristol: Policy Press. https://doi.org/10.1332/policypress/9781861347893.001.0001

Davis, Angela, and Laura King. 2018. “Gendered Perspectives on Men’s Changing Familial Roles in Postwar England, c.1950–1990.” *Gender & History* 30 (1): 70–92. https://doi.org/10.1111/1468-0424.12333

Diller, Hans-Jürgen. 2012. “Mood in Renaissance Poetry – and in ‘Google Books’.” In *Poetry and its Language: Papers in Honour of Teresa Bela*, ed. by Marta Gibinska, and Wladyslaw Witalisz, 31–51. Bern: Peter Lang.

Eckert, Penelope. 2000. *Linguistic Variation and Social Practice*. Malden, MA: Blackwell.

Eckert, Penelope. 2018. *Meaning and Linguistic Variation: The Third Wave in Sociolinguistics*. Cambridge: Cambridge University Press.

Eckert, Penelope, and Sally McConnell-Ginet. 2003. *Language and Gender*. Cambridge: Cambridge University Press.

Eckert, Penelope, and Sally McConnell-Ginet. 1992. “Communities of Practice: Where Language, Gender, and Power All Live.” In *Locating Power, Proceedings of the 1992 Berkeley Women and Language Conference*, ed. by Kira Hall, Mary Bucholtz, and Birch Moonwomon, 89–99. Berkeley: Berkeley Women and Language Group.

Ferrari, Federica. 2010. “Gender and Identification in Political Discourse: The (American) Woman: As and Beyond a Category for Analysis.” *Political Linguistics*, 1 (39): 39–54.

Hamlett, Jane. 2009. “‘The Dining Room Should be the Man’s Paradise, as the Drawing Room is the Woman’s’: Gender and Middle-Class Domestic Space in England, 1850–1910.” *Gender & History* 21 (3): 576–591. https://doi.org/10.1111/j.1468-0424.2009.01561.x

Harvey, Karen. 2009. “Men Making Home: Masculinity and Domesticity in Eighteenth-Century Britain.” *Gender & History* 21 (3): 520–540. https://doi.org/10.1111/j.1468-0424.2009.01569.x

Head, Dominic. 2002. *The Cambridge Introduction to Modern British Fiction, 1950–2000*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511606199

Hiltunen, Turo, Jenni Riihimäki, and Jukka Tyrkkö. 2018. “Tracing the Use of Colloquial Language in the British Parliamentary Record in Late 19th and Early 20th Century.” Conference presentation at ICAME 39, Tampere, 30 May to 3 June 2018.

Hoijer, Harry. 1948. “Linguistic and Cultural Change.” *Language* 24: 335–345. https://doi.org/10.2307/410353

Kopazcyk, Joanna, and Andreas H. Jucker (eds). 2013. *Communities of Practice in the History of English*. Amsterdam: John Benjamins.

Laitinen, Mikko, and Tanja Säily. 2018. “Google Books: A Shortcut to Studying Language Variability?” In *Patterns of Change in 18th-century English: A Sociolinguistic Approach*, ed. by Terttu Nevalainen, Minna Palander-Collin, and Tanja Säily. 223–233. Amsterdam: John Benjamins.

Lakoff, Robin. 2008. “Language, Gender, and Politics: Putting ‘Women’ and ‘Power’ in the Same Sentence.” In *Handbook of Language and Gender* Volume 1, ed. by Janet Holmes, and Miriam Meyerhoff, 161–178. London: Wiley Blackwell.

MacDonagh, Michael. 1913. *The Reporter's Gallery*. London and New York: Hodder and Stoughton.

McConnell-Ginet, Sally. 2003. "'What's in a Name?' Social Labelling and Gender Practices." In *Handbook of Language and Gender* Volume 1, ed. by Janet Holmes, and Miriam Meyerhoff, 69–97. London: Wiley Blackwell. https://doi.org/10.1002/9780470756942.ch3

Mäkinen, Ilkka. 2013. "*Leselust, Goût de la Lecture, Love of Reading*: Patterns in the Discourse on Reading in Europe from the 17th until the 19th Century." In *Good Book, Good Library, Good Reading. Studies in the History of the Book, Libraries and Reading from the Network HIBOLIRE and Its Friends*, ed. by Aušra Navickienė, Ilkka Mäkinen, Magnus Torstensson, Martin Dyrbye, and Tiiu Reimo, 263–285. Tampere: Tampere University Press.

Mäkinen, Ilkka, and Jukka Tyrkkö. 2014. "Collocates and Phraseologies Associated with the Concept of Reading in Early and Late Modern Printed Texts in English." Conference presentation at the SHARP 2014, Religions of the Book, Antwerp, 17–20 September 2014.

Michel, Jean-Baptiste, Yuan Kui Shen, Aviva Presser Aiden, Adrian Veres, Matthew K. Gray, The Google Books Team, Joseph P. Pickett, Dale Hoiberg, Dan Clancy, Peter Norvig, Jon Orwant, Steven Pinker, Martin A. Nowak, and Erez Lieberman Aiden. 2010. "Quantitative Analysis of Culture Using Millions of Digitized Books." *Science* 331 (6014): 176–182. https://doi.org/10.1126/science.1199644

Miles, Andrew. 1999. *Social Mobility in Nineteenth- and Early Twentieth-Century England*. London: Springer. https://doi.org/10.1057/9780230373211

Mollin, Sandra. 2007. "The Hansard Hazard: Gauging the Accuracy of British Parliamentary Transcripts." *Corpora* 2 (2): 187–210. https://doi.org/10.3366/cor.2007.2.2.187

Moretti, Franco. 2013. *Distant Reading*. London and New York: Verso.

Mühlhäusler, Peter, and Rom Harré. 1990. *Pronouns and People*. Cambridge, MA: Basil Blackwell.

Nevala, Minna. 2011. "Altering Distance and Defining Authority: Person Reference in Late Modern English." In *Historical Sociopragmatics*, ed. by Jonathan Culpeper, 61–82. Amsterdam: John Benjamins. https://doi.org/10.1075/bct.31.04nev

Nevala, Minna, and Jukka Tyrkkö. 2018. "Lunatics, Crackpots and Maniacs: Mental Illness and the Democratization of British Public Discourses." Conference presentation at ICEHL 20, Edinburgh, 30 August 2018.

Oxford English Dictionary. Second edition, online. Oxford: Oxford University Press. http://www.oed.com

Pahta, Päivi, Minna Palander-Collin, Minna Nevala, and Arja Nurmi. 2010. "Language Practices in the Construction of Social Roles in Late Modern English." In *Social Roles and Language Practices in Late Modern English*, ed. by Päivi Pahta, Minna Nevala, Arja Nurmi, and Minna Palander-Collin, 1–27. Amsterdam and Philadelphia: John Benjamins. https://doi.org/10.1075/pbns.195.02pah

Partington, Alan. 2013. "Corpus Analysis of Political Language." In *The Encyclopedia of Applied Linguistics*, ed. by Carol Chappelle. Oxford: Blackwell. https://doi.org/10.1002/9781405198431.wbeal0250

Partington, Alan, Alison Duguid, and Charlotte Taylor. 2013. *Patterns and Meanings in Discourse: Theory and Practice in Corpus-Assisted Discourse Studies (CADS)*. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.55

Persson, Gunilla. 1990. *Meanings, Models and Metaphors. A Study in Lexical Semantics in English*. Umeå: Acta Universitatis Umensis.

Rix, Kathryn. 2014. "Whatever Passed in Parliament Ought to be Communicated to the Public. Reporting the Proceedings of the Reformed Commons, 1833–50." *Parliamentary History* 33 (3): 453–474. https://doi.org/10.1111/1750-0206.12106

Slembrouck, Stef. 1992. "The Parliamentary Hansard 'Verbatim' Report: The Written Construction of Spoken Discourse." *Language and Literature* 1 (2): 101–119. https://doi.org/10.1177/096394709200100202

Sveen, Hanna. 2010. "Good-Natured Fellows and Poor Mothers: Defining Social Roles in British Nineteenth-Century Children's Literature." In *Social Roles and Language Practices in Late Modern English*, ed. by Päivi Pahta, Minna Nevala, Arja Nurmi, and Minna Palander-Collin, 211–227. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.195.10sve

Tannen, Deborah. 2003. "Gender and Family Interaction." In *Handbook of Language and Gender* Volume 1, ed. by Janet Holmes, and Miriam Meyerhoff, 179–202. London: Wiley Blackwell. https://doi.org/10.1002/9780470756942.ch8

Tichy, Ondrej. 2018. "Lexical Obsolescence and Loss in English: 1700–2000." In *Applications of Pattern-driven Methods in Corpus Linguistics*, ed. by Joanna Kopaczyk, and Jukka Tyrkkö, 81–107. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.82.04tic

Tosh, John. 1999. *A Man's Place*. New Haven, CT: Yale University Press.

Tyrkkö, Jukka. 2016. "Looking for Rhetorical Thresholds: Pronoun Frequencies in Political Speeches." In *The Pragmatics and Stylistics of Identity Construction and Characterisation* (Studies in Variation, Contacts and Change in English 17), ed. by Minna Nevala, Ursula Lutzky, Gabriella Mazzon, and Carla Suhr. Helsinki: VARIENG. Available online at http://www.helsinki.fi/varieng/series/volumes/17/tyrkko/ (accessed 15 April 2019).

Tyrkkö, Jukka, and Joanna Kopaczyk. 2018. "Present Applications and Future Direction in Pattern-Driven Approaches to Corpus Linguistics." In *Applications of Pattern-driven Methods in Corpus Linguistics*, ed. by Joanna Kopaczyk, and Jukka Tyrkkö, 1–14. Amsterdam: John Benjamins. https://doi.org/10.1075/scl.82.01tyr

Watts, Richard. 1991. *Power in the Family Discourse*. Berlin: Mouton de Gruyter. https://doi.org/10.1515/9783110854787

Wodak, Ruth. 2003. "Multiple Identities: The Roles of Female Parlamentarians in the EU Parliament." In *Handbook of Language and Gender* Volume 1, ed. by Janet Holmes, and Miriam Meyerhoff, 671–698. London: Wiley Blackwell. https://doi.org/10.1002/9780470756942.ch29

Appendix

Kinship terms (all forms) retrieved (HTOED category 03.01.01.03); forms marked with * not included in counts

alliance*, alliances*, allies*, ancestor, ancestors, ancestral*, ancestress, ancestry, aunt, aunts, avuncular*, bairn, bairns, bastard, bastardy*, birth, blood-relations, brethren, brother, brother-in-law, brotherhood*, brothers, cadet*, cadets*, child, childhood, childlessness, children, collateral*, cousin, cousins, dad, daughter, daughter-in-law, daughters, daughters-in-law, descendants, elder, eldest, exact, father, father-in-law, fatherless, fatherly, fathers, filial*, first-born, fore-fathers, foremothers, foster-child, foster-mother, foster-parents, fostering*, fraternal*, friend*, friends*, generation*, god-daughter, godfather, godfathers, godmother, godson, grand-daughter, grand-nephews, grandchild, grandfather, grandfathers, grandma, grandmother, grandmotherly, grandmothers, grandparent, grandparents, grandson, grandsons, great-aunt, great-grandchild, great-grandfather, great-grandfathers, great-grandmother, great-grandmothers, great-grandson, great-grandsons, great-great-grandfather, great-great-grandfathers, great-uncle, guard, half-brother, half-brothers, half-nephews, half-sister, heirs*, husband, illegitimacy*, illegitimate*, in-laws, incestuous*, increase*, kindly*, kinsman, kinsmen, kinswoman, last-born, lawful, legitimacy, legitimate, lineal*, ma, majority, maternal, maternity, middle*, mother, mother-in-law, motherhood, motherless, motherly, mothers, mum, native*, nephew, nephews, next, niece, nieces, offspring, offsprings, only*, original, orphan, orphaned, orphans, pa, parent, parentage, parental, parenthood, parenting, parents, paternal, paternity, paw*, posterior*, posteriors*, posterity, posthumous, predecessor, predecessors, progenitors, progeny, relation, relations, roots, siblings, sister, sister-in-law, sisterly, sisters, son, son-in-law, sonhood, sons, source, sources, stepchild, stepdaughter, stepfather, stepmother, stepson, stepsons, succession, twin, uncle, uncles, wife

PART II

Public discourses in Present-Day English

"Thanks for the donds"

A corpus linguistic analysis of topic-based communities in the comment section of *The Guardian*

Andrew Kehoe and Matt Gee
Birmingham City University

In this chapter we use corpus linguistic techniques to analyse reader comments on the website of the UK newspaper *The Guardian*. Our research is based on a corpus containing over 500,000 articles published on the website from 2007 to 2010, along with all 6.2 million comments made on those articles. We examine the distribution of comments across articles and topics before going on to explore the commenting behaviour of individual people. The *Guardian* website is a public forum with over 20 million unique visitors per month yet, as we demonstrate, it is an online space where people are able to build distinct sub-communities around particular topics and interact regularly on an individual basis without necessarily knowing one another's true identity.

Keywords: blogs, comments, journalism, corpus linguistics, collocation, community, addressivity, reference, identity, topic

1. Introduction

Newspaper articles have been a significant component of all major reference corpora of English compiled over the past fifty years, beginning with the Brown corpus of American English (Kucera and Francis 1967) where 18% of text samples came from newspapers. Subsequent corpora in the 'Brown family' deliberately mirrored this proportion, with the LOB corpus (Johansson et al. 1978), and the Frown and FLOB corpora (Hundt and Mair 1998) being the most widely used examples. Meanwhile, the compilers of the 100 million word British National Corpus (Aston and Burnard 1997) designed its written component to include 10% newspaper articles. More recent resources such as the Corpus of Contemporary

https://doi.org/10.1075/pbns.306.05keh

American English (COCA, Davies 2008–) continue to rely heavily on newspapers. In fact, COCA contains a higher proportion of newspaper texts than any of the previous reference corpora mentioned above, with 20% of the corpus made up of this textual genre.

Although our reliance on newspaper texts in corpora continues, the last two decades have been a period of unprecedented change in the newspaper industry with a move away from traditional print media towards online publication. Accompanying this, there has been a growth in so-called 'user-generated' content: contributions to newspaper websites from people other than trained journalists, including blog posts and articles written by non-journalists as well as other Web 2.0 features such as reader comments. Herring (2013: 4) refers to this phenomenon as "convergent media computer-mediated communication" (CMCMC).

Crowston and Williams (2000) were amongst the first researchers to examine in detail the textual genres found on the web and to describe the genesis of CMCMC. In their terms, online news articles were initially "reproduced" from traditional print media but then began to be "adapted" to take advantage of web-specific features such as hyperlinking (see also Eriksen and Ihlström 2000). Our focus in this chapter is on reader comments on news articles, a feature which did not exist when Crowston and Williams carried out their research. We have previously investigated reader comments in a large corpus of blog posts, looking in particular at their potential role in improving document indexing (Kehoe and Gee 2012). Although our primary focus was on the Blogger and WordPress blog hosting sites, we also noted that many newspaper websites worldwide had launched their own blog sections with associated reader comment features. In this new research we focus on the UK newspaper *The Guardian*: the first mainstream news organisation to embrace user-generated content on a large scale.[1]

The blog section of the *Guardian* website was launched in March 2006 as "Comment is Free", a title that comes from a 1921 quote by former *Guardian* editor and owner C. P. Scott: "Comment is free, but facts are sacred".[2] In this section, authors can, by invitation, write a blog post, with readers encouraged to comment on such posts and take part in discussions. A fortnight after the launch of Comment is Free (CiF), *The Guardian* began to allow reader comments on conventional news articles across all sections of its website. Such comments appear beneath articles, or 'below the line' as it is sometimes called. Hermida and Thurman (2008: 6) report that five other UK newspaper websites were allowing reader comments on news articles by the end of 2006. Meanwhile, Santana (2011: 69) reports that the

1. http://www.theguardian.com/media/2014/feb/07/georgina-henry (All links to *Guardian* articles in this chapter were last accessed on 15 April 2019).

2. http://www.theguardian.com/commentisfree/2002/nov/29/1

Washington Post began to allow such comments in June 2006, followed by the *New York Times* in November 2007, with 75% of the top 100 US newspapers allowing comments by 2008.

Editors at *The Guardian* clearly saw the introduction of CiF as a major change in their working practices. In the week the section was launched, its founder Georgina Henry made the following observation:

> Many of the conventions ingrained by 16 years as an editor on the print version of the Guardian have been turned on their head. […] [W]e've invited several hundred people to blog as and when they want on any subject they choose and at any length. Instead of tight copy-editing – back and forth to writers, asking them to elaborate arguments, change introductions, and cut copy to fit – we're checking mainly just for libel. […] [T]hrowing open the doors to hundreds of contributors has raised a host of interesting new issues for us as editors. Balance is one. […] Tone is another. Did it matter that the language used by some of those bloggers was harsh and uncompromising – a tone you might suggest they modified if they were writing for the paper?[3]

In this study we analyse such language issues, looking at tone but focussing in particular on topic. As Henry states, CiF contributors are permitted to write "on any subject they choose". One of the key aims of our research was to determine, through corpus linguistic analyses, whether there are particular article topics that tend to generate a larger number of comments. The initial focus of the chapter is therefore on the language of *Guardian* articles rather than on the language of the comments associated with them. In the second part of the chapter, we go on to explore the communities of readers that appear to have formed around specific topics on the *Guardian* website. In doing so, we examine the behaviour of individual commenters, the various ways in which they refer to one another, and the nature of the relationships that form between them over time.

There have been previous studies of reader comments on online news articles, including Boczkowski and Mitchelstein (2011), Ruiz et al. (2011), Langlotz and Locher (2012), and Eller (2018). The research by Ruiz et al. (2011) and Eller (2018) is of particular relevance as both studies include an analysis of comments from the *Guardian* website. The key difference is that, while these previous studies tended to involve relatively small-scale manual analyses, ours is a large-scale corpus pragmatic approach which considers millions of comments posted over a four-year period. Throughout the chapter, we compare our findings with those of previous studies, in some cases providing additional evidence for the claims made therein and in others disproving them.

3. http://www.theguardian.com/commentisfree/2006/mar/18/editorsweek

2. The relationship between comments and article topic

We took our existing newspaper corpus as the starting point for the current research. This corpus consists of an unbroken flow of UK broadsheet newspapers, segmented into months and covering the period since 1984 (see Kehoe and Gee 2009 for details). All texts from 2000 onwards come from the *Guardian* website, and in this study we focus on a sub-corpus covering the period 2007 (the first full year of comments) to 2010 inclusive. This gives us 503,599 articles,[4] totalling just under 280 million words. We downloaded all reader comments associated with these articles, giving us a total of 6.2 million comments. These were stored in a parallel corpus which, coincidentally, also includes just under 280 million words. Both corpora were searched using our WebCorp Linguist's Search Engine (WebCorpLSE) software.[5] When revisiting all articles published between 2007 and 2010, we found that 22% of them had at least one comment, with the distribution shown in Figure 1.

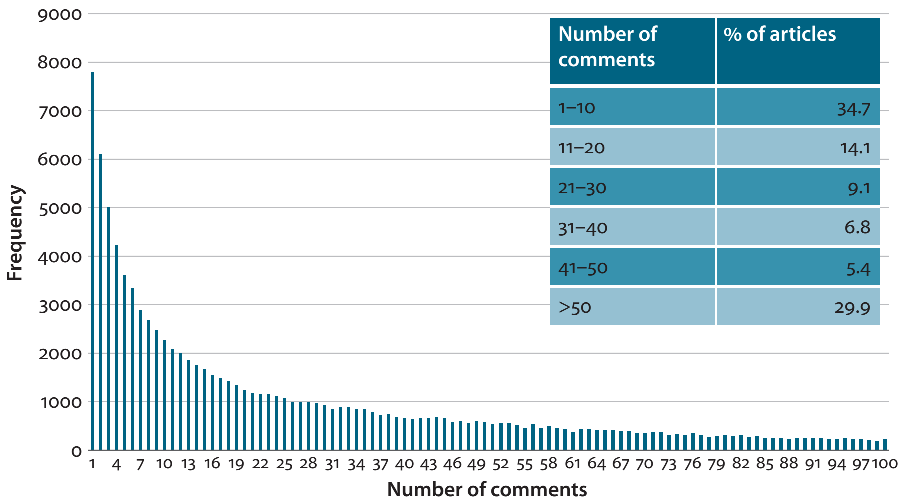

Figure 1. Distribution of *Guardian* articles with at least one comment

As we see in Figure 1, the majority of articles with at least one comment (34.7%) have between 1 and 10. However, there is a considerable proportion (29.9%) with more than 50 comments. The average number of comments per article in *The Guardian* is 25.73. These figures highlight a limitation of the research on *The*

4. We give a higher article count of 528,729 in Kehoe and Gee (2012) as we include articles from March 2006 onwards there whereas we include only January 2007 onwards here.

5. http://wse1.webcorp.org.uk/

Guardian conducted by Eller (2018), who included in her corpus only the first 20 comments from a small number of comment threads on articles published in a six week period during the same wider sampling window used in our study.

If we compare our statistics for *The Guardian* with the average number of comments per post in the *Birmingham Blog Corpus* (Kehoe and Gee 2012) – shown in Figure 2 – we see that the average *Guardian* article receives almost twice as many comments as a typical blog post elsewhere. Figure 2 also shows that the average *Guardian* comment is more than twice as long as the average comment in our wider blog corpus. This fact is even more significant given that we previously found long spam comments to inflate the average in the blog corpus (one particular comment being 71,253 words long), whereas *The Guardian* requires all users to register and has strict moderation procedures in place to eliminate spam.

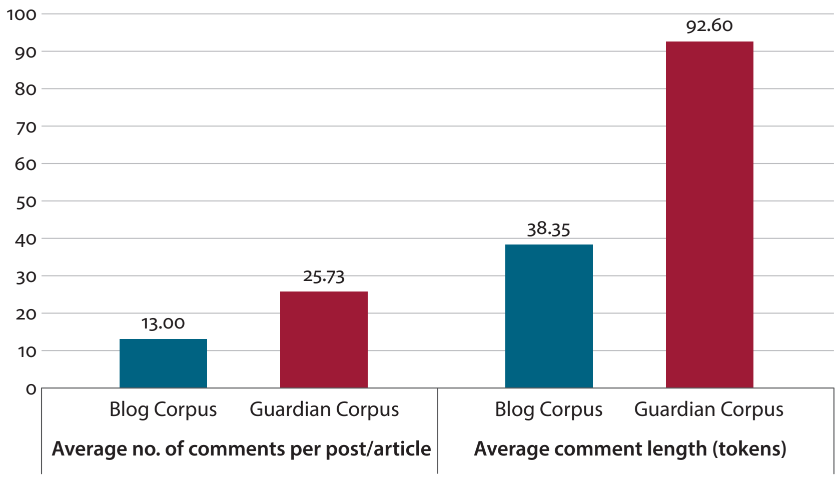

Figure 2. Comparing comments in the *Guardian* corpus and the *Birmingham Blog Corpus*

To summarise, we find that *Guardian* readers comment on only 22% of articles but, when there are comments, they tend to be long and there tend to be a lot of them. This is not surprising given that when the number of comments reaches a certain threshold there is likely to be a snowball effect, with ongoing debate encouraging additional comments. In addition, articles with lots of comments tend to be featured on the *Guardian* homepage and are thus more visible, leading to further comment.

Of course, this finding does mean that 78% of *Guardian* articles in our corpus have no comments at all. If the findings discussed above are indicative of *Guardian* readers' appetite for commenting, it seems unlikely that no reader wished to comment on any of these articles. Rather, it is more likely that, for whatever reason, they were not permitted to comment on these articles. *The Guardian's* policy on this[6] is rather vague, stating that

> In general we want to open comments up on our material wherever possible, but time and attention is finite (particularly in moderation resource) and we've learned from experience that some subjects and types of article attract less constructive or engaging debate than others. […] Comments will generally not be open on content which is sensitive for legal reasons (e.g. where there's a high risk of libel or contempt), or editorial reasons (could include: announcements of deaths, breaking news, stories about particularly divisive or emotional issues).

The key issue mentioned here that we go on to explore in the following section is the "subjects and types of article" that tend to attract most comments and those where commenting is not permitted at all.

2.1 Distribution of comments across Guardian sections

In order to explore the distribution of comments across the different sections of the *Guardian* website, we began by examining the URL of every article in our corpus. All *Guardian* URLs indicate the section of the site to which the article belongs, as illustrated by the following examples:

> http://www.guardian.co.uk/**commentisfree**/2010/may/19/india-poverty-line
> http://www.guardian.co.uk/**football**/2009/feb/17/jose-mourinho-liverpool

By analysing this information we were able to determine what proportion of articles in our corpus appeared in each section of the *Guardian* site. Then, by developing a script to re-access each URL and examine the HTML code, we were able to determine which articles allowed comments, which had received at least one, and which had received more than 40 comments. Our findings are summarised by section in Table 1.[7] Before we look at comments, the first thing to note here is that a considerable proportion of articles in our corpus (7.5%) come from the CiF section. This makes CiF the third largest section of *The Guardian*, behind only Sport (8.4%) and World (7.7%).

Turning our attention to comments, as Table 1 shows, overall 42.3% of *Guardian* articles published from 2007–2010 allowed comments, with 84.7% of these receiving at least one comment and 64.8% receiving over 40 comments. Unsurprisingly, CiF is the individual section where readers are given the most

6. http://www.guardian.co.uk/community-faqs

7. This table includes the top 30 sections of the *Guardian* website in terms of number of articles, as extracted from our data.

Table 1. Distribution of reader comments across sections of *The Guardian*

Guardian section	Proportion of whole corpus %	Comments allowed %	Of these, with 1+ comments %	With >40 comments %
commentisfree	7.5	98.4	99.5	58.7
sport	8.4	22.1	95.0	39.7
news	0.9	31.0	93.9	16.7
tv-and-radio	0.5	68.3	92.7	25.4
world	7.7	7.3	89.0	21.6
law	0.1	59.2	81.1	7.4
global	0.2	50.5	79.4	29.5
media	6.7	39.9	70.9	7.0
technology	2.5	76.4	64.9	12.3
football	7.1	24.9	59.8	39.8
leeds	0.1	83.4	58.5	0.0
politics	4.1	36.0	58.3	16.2
cardiff	0.1	77.8	56.0	0.4
artanddesign	1.0	39.9	53.6	11.1
environment	2.1	58.1	52.2	17.4
culture	2.0	49.7	50.9	6.0
edinburgh	0.1	85.0	46.4	0.0
uk	4.3	18.8	44.1	5.9
lifeandstyle	5.3	37.2	41.7	11.7
money	2.4	57.1	33.5	3.5
education	1.7	52.8	31.6	5.7
film	2.6	69.5	30.1	6.3
stage	1.5	74.6	30.0	1.5
science	0.6	75.2	29.3	7.5
society	2.4	44.4	29.1	3.5
theobserver	0.4	34.5	27.0	3.4
books	3.2	69.0	25.8	5.9
music	4.6	70.6	22.5	5.1
business	6.4	49.8	21.7	4.2
travel	1.4	15.8	20.3	4.4
unknown/other	12.1	–	–	–
OVERALL	**100.0**	**42.3**	**84.7**	**64.8**

freedom to comment although, contrary to the name, 1.6% of articles in this section do not allow comment! Almost all (99.5%) of the CiF articles that allow comments have at least one, and 58.7% have over 40.[8]

Our approach thus far is not without its limitations. The biggest issue is that CiF contains many different topics but it is not possible to determine what these are by URL alone. There are also several other rather general categories in Table 1 such as News, which clearly covers a wide range of different topics on a daily basis. In the next section we carry out a more fine-grained linguistic analysis to allow us to discover the most contentious topics in our corpus.

2.2 Distribution of comments by article topic

In our previous study (Kehoe and Gee 2012) we used the key words approach to explore textual 'aboutness', or topic, in blog posts. We adopt a similar method here but we now compare all articles with more than 40 comments against all articles without any comments. We are comparing the articles themselves rather than the comments on those articles. Our aim is to reveal the words which appear significantly more frequently in articles with lots of comments, as an indicator of the topic of those articles. The results appear in Table 2, sorted by log-likelihood (L-L) score (Dunning 1993).

In Table 2, we have classified all words manually into one of three broad categories: politics (red), religion/world affairs (green), or other (blue). There is of course some overlap between the categories, and the classification is approximate. Nevertheless, it is clear that political words in particular dominate the list in Table 2. What we see here is that if an article in our corpus contains one or more of the words *Labour, political, Cameron, Obama*, etc. it is more likely to have attracted lots of comments than an article which does not contain such words.[9]

It may not be particularly surprising to find political and religious words dominating this list. Mishne and Glance (2006) found a similar focus on politics and religion in their wider study of disputed topics in blogs, while Boczkowski and Mitchelstein (2011: 1) found the most commented upon stories in US newspapers to be on "political, economic, and international topics". We do, however, find

8. The 'Leeds', 'Cardiff' and 'Edinburgh' sections were an experiment run by *The Guardian* in 2010 to allow people in these areas to write (and comment on) blog posts about local issues. It was later discontinued: https://www.theguardian.com/help/insideguardian/2011/apr/27/guardian-local-update

9. At the other end of the spectrum, not shown in Table 2, are the key words in articles without any comments. Top of this list is the word *said*, perhaps reflecting the fact that conventional news articles with large amounts of reported speech are less likely to have comments enabled.

Table 2. Key words in articles with > 40 comments, classified manually (sorted by log-likelihood)

Type	L-L	Type	L-L	Type	L-L
Labour	22,313.97	Conservatives	4,192.41	Cameron's	2,626.23
political	13,301.61	war	4,182.18	Israel's	2,623.64
Cameron	12,196.36	yet	4,136.60	economy	2,617.73
Obama	8,831.98	question	4,128.31	Palestinians	2,568.12
even	8,797.73	change	3,994.69	country	2,557.45
Tories	7,970.83	Muslim	3,884.04	governments	2,536.76
Lib	7,790.93	economic	3,852.29	crisis	2,534.60
election	7,295.39	coalition	3,823.94	blog	2,511.51
politics	7,214.12	vote	3,708.41	policies	2,496.66
Brown	7,205.11	power	3,681.64	law	2,489.84
government	6,765.16	liberal	3,498.59	Tehran	2,403.93
public	6,631.67	human	3,445.87	Conservative	2,394.84
Clegg	6,311.56	seems	3,433.13	wrong	2,345.27
Israel	6,150.00	moral	3,296.49	Britain	2,337.55
party	6,052.20	simply	3,191.63	Afghanistan	2,315.50
religious	5,989.99	Iraq	3,188.61	Jewish	2,277.20
debate	5,822.55	democratic	3,158.59	leaders	2,262.79
Dems	5,822.20	Blair	3,098.41	Palestinian	2,244.21
state	5,799.60	parties	3,098.12	reason	2,202.37
Tory	5,333.55	course	3,092.64	football	2,187.30
society	5,293.63	faith	3,040.48	truth	2,147.81
voters	5,291.18	rights	3,002.14	far	2,110.51
policy	5,235.27	Jews	2,871.24	God	2,089.48
climate	5,173.91	Israeli	2,859.96	perhaps	2,069.40
reform	4,573.67	surely	2,844.34	freedom	2,045.20
Labour's	4,399.87	parliament	2,820.19	cuts	2,007.11
politicians	4,353.21	progressive	2,819.48	EU	1,982.44
Iran	4,343.78	Democrats	2,792.68	MPs	1,941.88
religion	4,313.92	speech	2,785.65	BNP	1,896.09
democracy	4,308.29	electoral	2,722.48	women	1,876.59
Muslims	4,306.48	right	2,708.55	Islam	1,854.89
fact	4,288.38	citizens	2,681.86	seem	1,843.61
argument	4,265.27	tax	2,636.62	gay	1,830.02

several words outside these categories in our 'other' group (blue), some of which relate to particular topics with a tendency to attract a higher number of comments: *human, rights, freedom, football, climate, women, gay*. We also see *change* and *reform*, concepts which tend to provoke discussion regardless of the exact form they take, as well as the debated concepts of *fact, truth, right* and *wrong*.

3. Commenter behaviour

In the preceding sections we have explored the relationship between the topic of *Guardian* articles and the number of comments they receive. In this section we turn our attention to the commenters themselves. Our corpus contains 6.2 million comments made by 286,878 users on 112,241 articles from the *Guardian* website. It may sound like a large number of people are involved in commenting, but we must bear in mind that the *Guardian* website attracts approximately 25 million unique visitors per month.[10] The fact that we observe 286,878 unique commenters in a four-year period indicates that only a small proportion of those 25 million visitors engages with the commenting feature.

This is consistent with the findings of previous studies. For example, Ruiz et al. (2011: 476) also studied *The Guardian* and found that in a corpus of 16,493 comments a large majority of users (88.6%) commented just once, with only 3% of the community contributing more than 6 comments and 3 users contributing more than 21 comments. From this it could be concluded that regular commenting is unusual. However, expanding the scale of the study, as we have done, reveals that this core group of regular commenters is actually substantial in size.

Table 3 shows the usernames of the 20 users who comment most frequently in our dataset, with the most frequent commenter being *MartynInEurope*, who made 15,233 comments on 4,874 articles in the four-year period. In fact, overall there are 108 users responsible for 10% of all 6.2 million comments in our corpus, 913 users responsible for a third of all comments and 2,529 responsible for half.

Relating this to our previous discussion of topic, we also observe that these commenters are active in different parts of the *Guardian* website. Taking three of the most frequent commenters and plotting the sections in which their comments appear (Figure 3), we find that *MartynInEurope* comments mostly in CiF, *CunningStunt* comments mostly in Technology, and *miroljub* comments mostly in the Sport and Football sections. Figure 3 also includes the key words in the set of articles commented upon by each of these three users, calculated using the same technique as in Table 2. This reveals that *MartynInEurope* appears to be a

10. https://advertising.theguardian.com/advertising/why-us

Table 3. The 20 most frequent commenters in our corpus

Username	Number of comments	Number of articles
MartynInEurope	15,233	4,874
CunningStunt	11,500	412
lefthalfback	11,447	2,674
MoveAnyMountain	10,718	4,288
Berchmans	10,033	3,272
miroljub	10,001	2,476
chubster2010	9,733	916
RogerINtheUSA	9,075	3,959
stevehill	8,489	4,417
stevejones123	8,119	3,470
kizbot	7,894	1,872
petrifiedprozac	7,827	1,780
WheatFromChaff	7,697	2,240
LordSummerisle	7,490	3,369
ElliottCB	7,125	762
edwardrice	6,752	2,192
MouthoftheMersey	6,720	2,353
Limni	6,475	781
BeautifulBurnout	6,363	1,866
lovingu	6,313	1,787

prototypical commenter, commenting on the same religion and world affairs topics that our analysis in Table 2 showed to be those attracting the most comments more generally. The key words in Figure 3 also offer a more fine-grained view of *CunningStunt's* commenting behaviour, revealing that the Technology articles this user comments on are actually articles on the specific topic of video games.[11]

11. "Chatterbox" (one of the key words for *CunningStunt* in Figure 3) was until 2017 the section of *The Guardian* website dedicated to the discussion of video games. The fact that *CunningStunt* managed to make 11,500 comments on only 412 articles in this section is related to the nature of the section, where many of the 'articles' are actually single sentences designed only to encourage contributions in the comment section below.

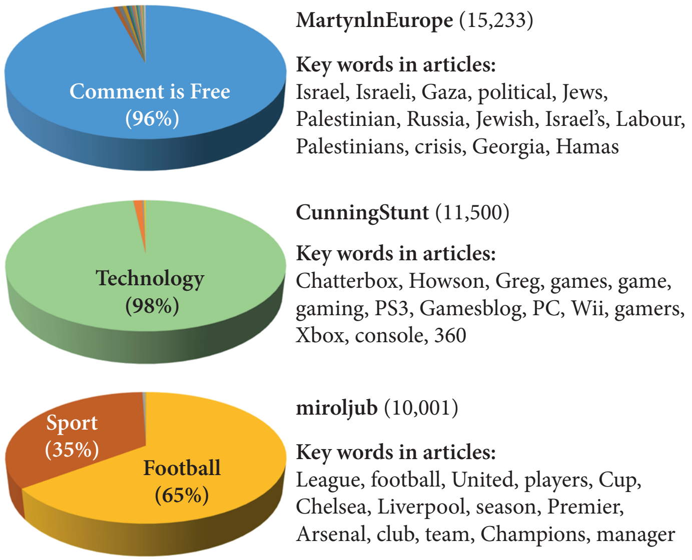

Figure 3. The sections in which three of the top commenters comment

A few of the users in Table 3 also signal their interests or reveal something of their identity through their chosen usernames. For example, *MartynInEurope, RogerINtheUSA* and *MouthoftheMersey* appear to make reference to where they are based or where they are from (Europe, the USA, and Merseyside in the UK respectively). The username *lefthalfback* suggests an association with football, which correlates with the sections in which this user tends to comment. Most of the usernames in Table 3 are pseudonyms but in some cases (e.g. *edwardrice*) users appear to have combined their actual first and surnames, possibly choosing to reveal their real-life identities.

The commenters on the *Guardian* website can be seen as a relatively small group of very active users. Given this, we wanted to discover how these users interact with each other, in particular:

1. Do the users interact with the same users regularly and thus form communities, possibly based around specific topics?
2. What is the nature of the relationships between users, as can be observed from the language used?

In order to answer these questions, we chose to employ large-scale analysis methods as described in the following sections. These have the benefit of allowing us to include the majority of our dataset in our study. This use of a large-scale, data-driven approach to study pragmatic and social phenomena is, however, experimental, and thus we also wished to investigate to what extent these methods are suitable for answering such questions.

3.1 Interactions between commenters

To discover the extent to which users interact with each other, we needed a way to identify the specific comments in which those interactions happen. Some commenting systems include a reply feature, where users can write a comment clearly labelled as a reply to another comment. At the time of data collection, however, *The Guardian* did not include a reply feature in its commenting system. Herring (1999) describes the ways in which users circumvent the restrictions imposed by CMC technology to bring coherence to the discourse. Unlike spoken conversation, where each turn is expected to be relevant to the preceding turn, online comment threads function as multi-participant asynchronous exchanges. This leads to the problem Herring (1999) refers to as "disrupted turn adjacency", where a response to a message in a CMC system may be displayed a long distance from (or on a separate webpage to) the original message. Herring found that users frequently employ an "addressivity" tactic to overcome this limitation. Specifically, they include the username of the author of the message to which they are responding at the start of their own message. We have found this tactic to be employed by *Guardian* commenters, as we describe in more detail below.

The first step in finding user interactions was to find usernames in our collection of comments. For each comment in our corpus we collected both the text of the comment and its metadata, such as author and date. We started by extracting all usernames from the metadata. Users with few comments overall would not have had the opportunity to interact repeatedly, so we filtered this list to exclude users with fewer than 10 comments, which gave us 42,347 usernames. However, some of these usernames were common English words (such as *teacup* or *conifer*) which needed to be excluded from our search to avoid skewing the results. Thus, we compared the list of usernames against a list of words with frequency greater than 100 in the *Guardian* articles, removing any usernames which were found in this wordlist. After the filtering process, 2,690 usernames (6.4%) were removed from our list leaving 39,657 usernames.

The filtered username list was then used to search our comment corpus. Where a match was found, we extracted the author of the comment from the metadata, allowing us to construct a frequency table showing how often commenters include another user's name in their comments (Table 4).

Table 4. The top 40 references to a user by a commenter. C > R shows the frequency with which the commenter includes the referenced username in their comments. R > C shows the reciprocal.

Commenter	Referent	C > R	R > C
everwonthetreble	aceal	367	75
parisa	deadgod	311	126
freewoolly	beor	306	118
sonofwebcore	tincanman	288	26
kevinnevada	ngavc	272	105
parisa	pinkroom	270	64
woollymindedliberal	peitha	263	20
styxdweller	petrifiedprozac	232	176
steenbeck	nilpferd	226	123
freewoolly	savvymum	223	22
aceal	donwendyagain	222	157
lefthalfback	aleatico	206	34
iamnothere	parisa	204	144
exarmy	permare	201	131
zephirine	pepp	195	0
bitethehand	allyf	192	51
freewoolly	donella	192	56
sonofwebcore	steenbeck	187	54
parisa	anytimefrances	185	96
zdenekv	olching	185	1
nefastus	roverdc	182	3
tincanman	tonnl	181	2
sonofwebcore	rockingmitch	179	169
parisa	smpugh	179	50
petrifiedprozac	styxdweller	176	232
goody84	limni	175	1
sonofwebcore	magicman	174	47
rockingmitch	sonofwebcore	169	179
cunningstunt	limni	166	17
petrifiedprozac	twoswords	165	17
lefthalfback	ngavc	164	142
heverale	styxdweller	163	77

Table 4. (*continued*)

Commenter	Referent	C > R	R > C
berchmans	akus	162	94
monctonian	marsman	158	60
donwendyagain	aceal	157	222
woollymindedliberal	mrpikebishop	155	0
tincanman	daddypig	153	28
parisa	hic8ubique	150	56
deadgod	pinkroom	149	30
shachtman	berchmans	147	82

As can be seen in Table 4, some users make reference to other users with high frequency (shown in the C > R column). Eller (2018), after studying the first 20 comments from each thread included in her study, claims that user-user interactions are rare in *Guardian* comment threads. She suggests the lack of a built-in reply feature as the reason for this. In contrast, our findings show that the larger picture is one involving a great deal of interactivity between the most active users. Furthermore, the high frequency of references users make to one another suggests that they may be well acquainted (within the bounds of the *Guardian* website). However, the reciprocal frequencies (in the R > C column) show that the same number of references are not always returned. Overall the frequencies in Table 4 point towards a mixture of one-and two-way interactions.

Taking a closer look at references to users by other users, we see interactions that span multiple comment threads over several months or even years. Examples (1) to (3) show usernames being used as terms of reference or address (the first name is the author of the comment and the second name in italics is the referent's username).

(1) aceal: *Donwendyagain* I think your memory of the recent past is a little wayward.

donwendyagain: *Aceal*, in relation to the officials decisions and possible points dropped I refer you to my response to riisenpeas. […] But dont let the facts get in your way *aceal*.

aceal: *Donwendyagain* yeah ok whatever! It must be great to have a crystal ball that you know exactly what is going to happen for the rest of this season […] (Football, 26/10/07)[12]

12. http://www.guardian.co.uk/football/2007/oct/26/liverpool.premierleague

(2) aceal: Everwon where's *Donwendyagain* what did I do to offend that blogger? Eh by gum it's *Donwendyagain*! Where are you lad? Come back all is forgiven lad!!!

(Football, 21/10/09)[13]

(3) aceal: *Donwendyagain Donwendyagain* you're back!!! The legend returns! How are you lad?
donwendyagain: I'm grand *Aceal* lad, thanks for asking. I dont need to ask how you are […] (Football, 21/10/09)[14]

In Examples (1) to (3) we see the "addressivity" tactic in operation. It is worth noting that it does not only function as described by Herring (1999), that is, by prepending the username to the start of the comment. In our data we find that users also address each other within the content of their comments. In the majority of comments this enables users to specify an intended recipient, functioning to help organise and structure the discourse. We also note that the reference made to *donwendyagain* in Example (2) is a mention rather than a direct address – in fact a different intended recipient (*Everwon,* an abbreviation of *everwonthetreble* who appears in Table 4) is specified at the beginning of the comment.

Taking the two users in Examples (1) to (3), *aceal* and *donwendyagain*, we see that they regularly debate matters relating to the topic of football (as in Example (1)). Examples (2) and (3) occur two years later, but it should be noted that during the intervening two years the two users make reference to each other several times in their comments. Example (2) is taken from a comment thread where *donwendyagain* did not comment, but *aceal* was expecting to receive a response from them, something which *aceal* goes on to note. Example (3) comes from a new comment thread started later the same day about the same sporting event. In this example, *donwendyagain* does comment and this time *aceal* makes reference to their return. It seems that these users identify each other as 'jousting partners' within the discussions, clearly familiar with each other over years of commenting on similar topics.

A different type of relationship is shown in Examples (4) to (6).

(4) rockingmitch: for my news, I've just found out that Mrs. RockingMitch and I are to become grandparents for the first time.

13. http://www.guardian.co.uk/football/blog/2009/oct/21/champions-league-liverpool-lyon

14. http://www.guardian.co.uk/football/blog/2009/oct/21/rafael-benitez-liverpool-debate

sonofwebcore: *ROCKINGMITCH* – Conbloodygratulations old chap, and to Mrs Mitch too. Know exactly how you feel.
(Music, 24/10/08)[15]

(5) sonofwebcore: *ROCKINGMITCH* – It is done!! I'll call you tomorrow.
sonofwebcore: EVERYBODY – *ROCKINGMITCH* phoned me today.
(Music, 31/10/08)[16]

(6) rockingmitch: @ *Sonofwebcore* I do hope everything's ok with you and your other half. All good wishes.
sonofwebcore: MAKINAVAJA + *ROCKINGMITCH* – Thanks lads, she's doing fine, especially since she came off the morphine.
(Music, 11/02/10)[17]

In these examples we see conversations between *rockingmitch* and *sonofwebcore*. The examples follow the "addressivity" tactic of prepending the username to the comment, but also include an instance of a username being mentioned rather than addressed (Example (5)).

Examples (4) to (6) also show some variation in the form of the references. Using all capitals for a username clearly signals this as being separate from the rest of the comment. Alternatively an @ character is used to mark the word as a username, as popularised by Twitter. Example (6) shows that multiple recipients can be addressed by joining the usernames with the + character. By contrast, a reference to "EVERYBODY" signals a comment aimed at a much larger group (see Example (5)). Different extents of reference are possible here: the users already active in the thread, the users who comment in the music section of the website, all active commenters on the *Guardian* website or conceivably all readers of *The Guardian* (as the comments are publicly accessible). We return to the concept of audience in Section 3.4.

Rather than being based around debate, Examples (4) to (6) reveal a more familiar and personal relationship. It seems that the relationship started online in the *Guardian* comment threads (Example (4)) and an offline element was added as time progressed (Example (5)). We also see these users share quite personal information with one another, wishing each other happy birthday in some threads

15. http://www.guardian.co.uk/music/musicblog/2008/oct/24/readers-recommend-afternoon-songs

16. http://www.guardian.co.uk/music/musicblog/2008/oct/31/readers-recommend-travelling-songs

17. http://www.guardian.co.uk/music/musicblog/2010/feb/11/readers-recommend-whispering-songs and talking about their partners' health (Example (6)). Through congratulating, thanking, showing concern and general openness these users present amiable and affable identities for themselves and build rapport with one another. These relationships are maintained over periods of time, with the difference in time between Examples (4) and (6) being 15 months.

We noted above references to users which are not intended to address them (Examples (2) and (5)). This illustrates a limitation of a fully automated approach. Although we are able to extract all references to users from the comment corpus, without manual analysis we are unable to differentiate mentions from direct address. A solution would be to restrict the analysis to instances where the username occurs at the beginning of the comment. However, as will be shown in Sections 3.4 to 3.6, a number of frequently used constructions require the addressee's username to be part of the comment body (similar to Examples (2) and (3) above). Thus, we chose not to filter our data in this way, but we were sure to bear this limitation in mind throughout our analysis.

Thus far, we have presented an analysis of interactions between only a few users. In the following sections, we employ methods that help us understand to what extent these findings can be generalised across the corpus.

3.2 Topic-based commenter communities

We have seen in the previous section that there are certain pairs of commenters who interact regularly in comment threads on the *Guardian* website and are able to build sustained relationships of various kinds. Our next question was to what extent these individual interactions intersect with each other to form a wider community of users. In order to investigate this we turned to Network Graphs. Network Graphs allow links between nodes to be visualised and plotted according to the relative strength of those links. In our case, the nodes are users and the links are the number of comments in which one user refers to another (i.e. the data in Table 4).

We built a Network Graph using the Gephi software package (Bastian et al. 2009). We used the Force Atlas algorithm, which is a 'force-directed' graph drawing algorithm designed to pull nodes close together based on the strength of the link between them (in our case, the number of references) and simultaneously to push nodes apart by a constant value to prevent the graph collapsing into a single point. We took the 200 commenters who had made the most comments, calculated the number of times they referred to each other and plotted the graph (Figure 4).

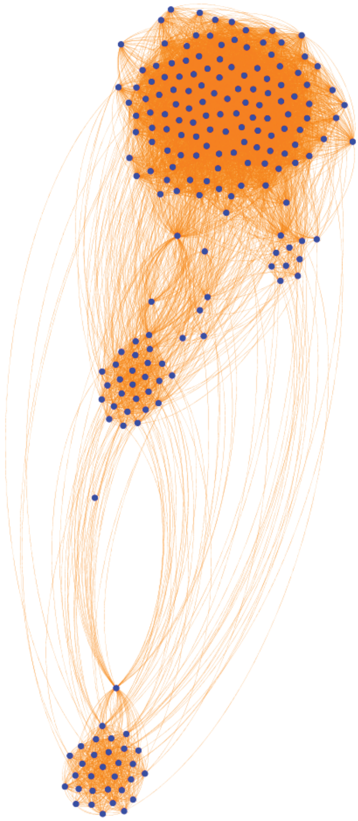

Figure 4. A force-directed graph where the nodes (dots) are users and the edges (lines) are references made by one user to the other

In Figure 4 we see three clear clusters of nodes. By investigating the users in each cluster, we have been able to determine that the large cluster at the top consists of users who comment mostly in CiF, the centre cluster is users who comment mostly in the Football section, and the bottom cluster is users who comment mostly in the Technology section. There is a small cluster at the bottom right corner of the top-most one, where we see users who comment mostly in the Music section. Where users appear between clusters, they have links to users within the clusters near to them and can be observed to comment in multiple sections. Given that these groups of users make frequent reference to one another within specific parts of the *Guardian* website (namely CiF, Football, Technology, and Music), this seems to indicate that these sections of the newspaper have developed their own communities centred around shared interests which reflect the topics of those newspaper sections. We explore the Music section in more depth in Section 3.6.

3.3 The nature of user interactions

We have observed a large number of interactions between users and shown that they tend to cluster around sections of the *Guardian* website. Now we turn our attention to the nature of those interactions, once again taking a large-scale data-driven approach, namely collocation.

In a related study (Lutzky and Gee 2018), reactions to blog posts were studied by investigating words and phrases at the beginning of comments in the *Birmingham Blog Corpus*. By limiting the position under investigation to the start of comments and analysing the most frequent words and clusters appearing in that position, the pragmatic ties between blog posts and comments could be uncovered. Greetings, interjections, positively evaluative words and expressive speech acts such as thanking and congratulating were found to occur frequently in this position. These all enable the bloggers and commenters to establish interpersonal relations and build rapport. In the present study, we similarly limit our focus to a specific position in the newspaper comments by investigating the text surrounding usernames. Whereas the comment-initial position revealed reactions to the blog post or a preceding comment, our hypothesis here was that the immediate context of usernames would reveal reactions to that specific user's previous comment and possibly yield insights into the relationship between users or the nature of their interaction.

In Section 3.1, we noted that users addressed each other both by prepending the intended addressee's username to the comment and by referencing the username in the body of the comment. Therefore, we opted to use a method which would allow us some flexibility when summarising the language used near usernames. Specifically, we used span-based collocation, where we define collocates as words which occur within a span of four words to the left and right of the node word (the node word being the username). Again, we used our filtered list of usernames and constructed a list of all collocates. Stopwords (i.e. closed-class grammatical words) and usernames were removed from the collocate list. Each collocate was scored for its strength of association with the usernames using the z-score measure. The top 90 collocates are presented in Table 5.

The first thing we note is that few words relating to the topics of comments are present in Table 5. This suggests that targeting the span of words around usernames has had the desired effect; that is, the words we find mostly relate to the nature of the interaction rather than the specific topic of discussion.

Table 5 also shows that user interactions in *Guardian* comment threads have an informal, spoken flavour, as indicated by exclamations (*oh, wow, gosh*), interjections (*er, hmm*) and laughter (*lol, haha*). We also note a number of colloquialisms that have a regional aspect to them. Of chiefly British use or origin are, for instance, *cheers, mate, aye* and *nah*.

Table 5. The top 90 collocates of usernames in the comments (span 4, sorted by z-score, stopwords and usernames removed)

Collocate	Score	Collocate	Score	Collocate	Score
thanks	670.04	absolutely	142.41	quote	94.97
agree	577.24	hey	142.13	yup	94.15
hi	362.44	reply	141.60	see	93.45
sorry	362.13	indeed	141.41	hear	92.86
thank	358.04	ok	141.16	sounds	88.71
re	306.59	interesting	139.49	say	87.54
think	294.40	fair	138.85	forgot	86.61
post	282.99	lol	134.85	totally	86.11
said	282.58	know	134.53	firstly	86.09
point	232.15	read	134.23	missed	85.91
dear	228.71	dond	131.15	perhaps	84.85
oh	228.16	actually	130.07	asks	84.71
ah	210.73	posted	125.55	ha	81.15
right	208.37	nice	124.93	funny	78.77
says	198.46	disagree	122.48	mean	78.76
spot	196.54	correct	121.43	seconded	78.49
cheers	186.35	really	114.94	others	78.06
donds	178.08	true	114.80	course	77.96
excellent	166.95	quite	114.70	maybe	76.19
posts	165.19	mate	111.12	thought	75.45
apologies	163.44	link	110.48	btw	74.73
please	156.97	wow	110.06	aye	74.58
hello	155.17	nope	107.59	hmm	72.41
comment	154.71	pointed	107.09	brilliant	71.36
agreed	152.93	glad	105.71	answer	71.15
good	152.41	thankyou	105.36	congratulations	69.98
yep	152.38	response	105.15	nah	69.18
points	148.63	just	99.90	welcome	69.07
exactly	144.18	comments	96.97	got	68.79
sure	143.61	er	96.20	afraid	68.59

In order to explore the collocational information in more detail, we extracted the most frequent three word clusters within which the collocates of usernames occurred. The three word clusters consist of sequences of three words that include one or more of the significant collocates and that occur near a username (within a span of 4 words). For example, the most frequent clusters containing *thanks* and occurring near a username were *thanks for the* (frequency 5,153), *thanks for your* (2,123) and *thanks for that* (1,574). The most frequent clusters containing *agree* and occurring near a username were *i agree with* (8,551), *agree with you* (2,711) and *i agree that* (1,502). Table 6 in the Appendix includes an expanded list of these clusters. In the next sections, we discuss three features of the interactions between users which became salient to us through this analysis.

3.4 Argumentation and discussion

In our data we found a considerable number of examples relating to criticism, discussion, agreement and disagreement. This is not surprising in the context of newspaper discussion threads, which are largely intended as spaces in which readers can engage in debate about the articles and their topics. This aspect of newspaper comments has previously been studied in some detail, for example by Eller (2018). Eller found that comment authors would often be the target of criticism from other commenters, with aspects such as the reasoning, abilities, knowledge, attitude or character of the comment author being most frequently targeted. However, it should be noted that Eller focussed exclusively on instances of criticism without considering supportive comments.

Within the context of a newspaper discussion thread, user interactions are naturally stance-saturated (for research into stance and its effect on identity see in particular Jaffe 2009). Much of the interaction between commenters is characterised by evaluation as users assess each other's arguments and put forward their own. Acts of evaluation are simultaneously acts of positioning, i.e. acts of alignment or disalignment with or against others' positions. In our analysis of collocates of usernames and their clusters, we were able to extract the core constructions used by commenters to display agreement and disagreement, and thus show how they present their stance and position themselves in line with or in opposition to their interactants. These techniques are illustrated in Examples (7) to (9), which include matter-of-fact statements of agreement or disagreement.

(7) *I agree with* ngavc: It was moving and sad speech. (CiF, 26/04/10)[18]

18. http://www.guardian.co.uk/commentisfree/michaeltomasky/2010/apr/26/usa-barack-obama

(8) Mawaltrees you make *a good point* about the top four finish.
(Football, 19/12/09)[19]

(9) So Salmonberry, *I disagree with* you. People who go along with a weak system are weak people, morally. (CiF, 31/03/09)[20]

It is worth noting that in these examples the commenters make reference to the authors of previous comments either by username (7) or by username and the pronoun "you" ((8) and (9)). Example (7) is a mention of a username whilst (8) and (9) are examples of address, but in all three cases the focus is placed on the authors of the previous comments. Alternatively, commenters may refer to the comment itself, making the comment rather than the user the focus of attention, as in (10). A third option is for users to state their own opinion in a manner which forefronts themselves as an individual, as in (11). In this example, "actually" signals that the commenter disagrees and "I think" allows them to place the focus on themselves. Through these means commenters can signal varying degrees of association or dissociation, depending on whether the addressee, the comment made by the addressee, or the addresser are foregrounded.

(10) exiledlondoner, that's *an excellent post*. (CiF, 26/08/08)[21]

(11) Ilikedthe80s *Actually I think* it's more that his job was incidental, not his sexuality. (CiF, 02/11/09)[22]

Commenters also express their degree of agreement or disagreement. The adverbs *totally, quite* and *absolutely* collocate strongly with usernames (see Table 5) and these occur most frequently in clusters with terms used to express agreement, as in Examples (12) to (14).

(12) @chesney9 *I totally agree* – the benefits system is a basketcase.
(CiF, 18/01/10)[23]

19. http://www.guardian.co.uk/football/blog/2009/dec/19/liverpool-rafael-benitez-portsmouth

20. http://www.guardian.co.uk/commentisfree/2009/mar/31/jacqui-smith-mps-expenses-media

21. http://www.guardian.co.uk/commentisfree/2008/aug/26/israelandthepalestinians.middleeast

22. http://www.guardian.co.uk/commentisfree/2009/nov/02/baynham-homophobia-hate-crime-vigil

23. http://www.guardian.co.uk/commentisfree/2010/jan/18/family-marriage-tax-cameron-labour

(13) smpugh: *You're absolutely right* about the regularity of scansion;
(Books, 16/02/09)[24]

(14) But *you're quite right*, zephirine, the game in the WI has long been in need.
(CiF, 11/06/08)[25]

The fact that we find such examples of agreement is interesting in itself. While the prevailing stereotype of newspaper comment threads is that they consist predominantly of criticism and conflict (see e.g. Eller 2018), the examples discussed so far show that users are clearly willing to support each other.

Nevertheless, we do find examples of strong disagreement in our data too. When studying the collocates of usernames (see Table 5), we also came across the forms *really, oh* and *sorry*. Taken as individual words these may not seem to relate to disagreement, but the most frequent clusters in which they occur are *you really think, oh come on/off, dear oh dear, I'm sorry but* and *[I] am sorry but*. These are illustrated in Examples (15) to (17).

(15) IAMJGW do *you really think* they would have told the tabloids.
(Football, 19/12/09)[26]

(16) *oh come on* Seagulljavea, you can't be one of those people who think Beckham was a success ON THE PITCH in Spain. (Football, 24/07/07)[27]

(17) Marisd->*I'm sorry but* the government does not own my body.
(CiF, 15/01/08)[28]

We interpret Examples (15) to (17) as instances of incredulity. The commenter is expressing their disbelief at a previous comment and, thus, they are distancing themselves from that point of view. Simultaneously, they are projecting properties such as 'naïve' or 'foolish' (Examples (15) and (16)) onto the addressee, influencing how their interlocutors may be perceived by other readers. In this way, they attempt to weaken their interlocutor's strength of argument and future credibility.

24. http://www.guardian.co.uk/books/booksblog/2009/feb/16/lyke-wake-dirge-poem-week

25. http://www.guardian.co.uk/commentisfree/2008/jun/11/israelandthepalestinians.middleeast

26. http://www.guardian.co.uk/football/blog/2009/dec/19/liverpool-rafael-benitez-portsmouth

27. http://www.guardian.co.uk/football/2007/jul/24/sport.cricket

28. http://www.guardian.co.uk/commentisfree/2008/jan/15/politics.publicservices

3.5 Terms of address

A number of terms of address can be found amongst the collocates of usernames. The word *mate* appears in Table 5, and *sir, dude, pal* and *buddy* are also significant collocates of usernames, illustrated in Examples (18) and (19).

(18) pondwatching, No worries *mate*! I like the wind ups! Thats why I am always here! (Football, 08/05/07)[29]

(19) Bluedaddy, I salute you *sir*! :) (Sport, 25/10/07)[30]

In both examples the terms of address are used to express positivity towards the interlocutor. The word *mate* is an informal term, and in Example (18) it helps the commenter to appear relaxed and easy-going, and thus downplay any previous offence that may have been taken from the preceding "wind up". On the other hand, *sir* is an honorific, used in a congratulatory manner and to show respect to another person, as in (19). Even though *sir* is traditionally a formal term, in this case it is used in an informal manner, as indicated by the emoticon.

We also see the terms of address overlapping with the instances of agreement and disagreement, in particular with "spot on", as in Examples (20) and (21).

(20) Marketsarem […] Spot on *mate*, a very cogent analysis. (CiF, 18/01/07)[31]

(21) EscapeVelocity – you, *sir*, are spot on. (CiF, 16/03/08)[32]

In these examples, terms of address once again enable the commenter to adopt an informal and relaxed style, which can be observed throughout the collocates (as discussed above). We suggest that this helps with building rapport and maintaining civility in a forum environment which might otherwise feel hostile due to the amount of argumentation.

On the other hand, we noted in Section 3.4 that users frequently show incredulity in response to previous points in a debate. *I'm sorry but* was the most frequent three word cluster containing *sorry*, and the most frequent three word cluster containing *mate* is a variation on this: *sorry mate but*, as in Example (22).

(22) SteveL – *Sorry mate, but* what total drivel. (CiF, 16/12/08)[33]

---

29. http://www.guardian.co.uk/football/2007/may/08/newsstory.sport16

30. http://www.guardian.co.uk/sport/2007/oct/25/cricket.comment1

31. http://www.guardian.co.uk/commentisfree/2007/jan/18/comment.china

32. http://www.guardian.co.uk/commentisfree/2008/mar/16/europeneedssolidarityoverc

33. https://www.theguardian.com/commentisfree/2008/dec/16/musicindustry-popandrock

Here the commenter is disagreeing with their interlocutor and so there is the risk that the interlocutor will take offence or react in a hostile manner. We see two possible reasons for using the cluster in Example (22). One, the commenter's use of "mate" is an attempt to lessen the potential impact or likelihood of a hostile response. The OED (s.v. mate, *n. 2.*, Sense I.1.c) defines *mate* as "a form of address to a person […] regarded as an equal", and the commenter may therefore be attempting to place both users on an equal footing to avoid further hostility. Two, the commenter is seeking to establish an identity for themselves as being 'one of the good guys' or as being reasonable and personable. At the same time, the commenter is distancing themselves from their interlocutor, in a similar manner to that discussed in Section 3.4.

3.6 Dond(s): An example of in-group lexis

Perhaps the most striking words found in the collocates of usernames (Table 5) are the neologisms *dond* and *donds*. There is no entry for *dond(s)* in the OED but it does have an entry in Urban Dictionary, a website which allows users to provide their own definitions of words. Here *donds* is defined as "[s]upporting or agreeing with a previous statement. Seconding a motion".[34] The entry goes on to explain how *donds* originated in the "Readers Recommend" music blog on the *Guardian* website, starting as *secdonds*, a typo for *seconds*, before being truncated to *donds*. Use of the term started before our data collection period, but we see examples of *donds* in the Music section of our corpus, which corroborate the *Urban Dictionary* description.

(23) Nilpferd, I'll *dond* your He Loved him Madly. (Music, 29/08/08)[35]

(24) @craigoh – massive *dond* for Muttonbirds. (Music, 19/02/09)[36]

(25) Ooh zombus, *donds* for Penny Lane! (Music, 15/02/08)[37]

(26) @Steenbeck thanks for the *donds*. (Music, 4/12/08)[38]

34. The definition was provided by *durf47* on 23/09/09: http://www.urbandictionary.com/define.php?term=donds

35. https://www.theguardian.com/music/musicblog/2008/aug/29/readersrecommend29808

36. http://www.guardian.co.uk/music/musicblog/2009/feb/19/readers-recommend-nature-songs

37. http://www.guardian.co.uk/music/musicblog/2008/feb/15/thesongsthatsoundlikehome

38. https://www.theguardian.com/music/musicblog/2008/dec/04/readers-recommend-songs-life-lessons

In Examples (23) to (26) we see that since its creation the term *dond(s)* has been used both as a noun and verb. Examples (23) to (25) illustrate the most typical use of the term, that of seconding a previous suggestion (in these cases a recommendation of a song or artist). In Example (26), "donds" is used outside of the typical construction of seconding something, where the commenter thanks their interlocutor for the "donds" received previously. Therefore, a term that started out as a typo has become firmly ingrained within the lexicon of *Guardian* commenters.

As we noted above, the term *donds* was coined by commenters in the Music section of the *Guardian* website. In our corpus, there is little evidence of the term having spread to the comment threads of other sections. The combined frequency of *dond* and *donds* in comments on the Music section is 17,765. There are 386 instances in comments on the Film section, with the next highest usage being in Books (just 17 instances). This is the result of a particular practice within the comment threads of the Music section of *The Guardian*: that of users suggesting songs to which others should listen. The word *dond(s)* then fulfils a specific role for the Music community by allowing the users to respond positively to those song suggestions. They could have continued to use *second*, but instead have taken on and claimed collective ownership of a term that was created within their community. Thus, *dond(s)* is an example of an in-group term which forms part of the language and identity of commenters in the *Guardian* Music community.

4. Conclusion

In this chapter we have examined the growth of reader comments on online news articles since the mid-2000s, focussing in particular on the UK newspaper *The Guardian*. By carrying out a large-scale corpus linguistic analysis we have been able to identify sub-sections of the newspaper's website where commenting is most prevalent. Taking the analysis further, through the extraction of key words we have identified the specific topics which are most likely to generate debate, often relating to politics, religion and social issues.

Building on our initial analysis of the *Guardian* sections and topics attracting the most comment, we have also explored the commenting behaviour of individual users. Our most surprising findings here were that six users made over 10,000 comments each between 2007 and 2010, and that there were 2,529 readers who were together responsible for half of the 6.2 million comments made during this period. That fact that *Guardian* comment threads attract a relatively small but highly active group of users encouraged us to explore the interactions between these users and the relationships they formed over time.

By searching for usernames within comments we were able to discover that, contrary to claims made by Eller (2018), there are particular users who refer to other users by name repeatedly, in some cases over 300 times within the four-year period studied. When we explored these relationships in more detail, we found instances of interactions between pairs of users spanning multiple comment threads over several months or even years. Furthermore, we were able to give specific examples of interactions where users come to recognise one another's usual commenting behaviour – a specific aspect of their online identity – and expect particular users to comment on particular topics.

We were then able to use network analysis to reveal that distinct sub-communities of commenters have formed around particular sections of the *Guardian* website, specifically Comment is Free, Football, Technology, and Music. In the case of the Music section, we also found a clear example of in-group lexis, with the word *dond(s)* used over 17,000 times by commenters on articles in this section but used very little in comments on articles in other sections. Previous research has tended to focus on criticism and conflict in newspaper comment threads. For example, Eller (2018: 177) notes that criticising other people "seems to be one of the core functions of comments posted on newspaper websites". Our findings corroborate this to some extent, but there still exists a core of non-critical dialogue that deserves further attention, especially in relation to community building. Topic again plays a role here. We have noted that articles about politics, religion and world affairs are more likely to generate the longer comment threads indicative of heated debate, whereas readers active in communities centred around less controversial topics appear to approach their interactions in a more supportive way. This is especially true of users engaged in threads where they can offer positive suggestions to each other, such as Reader Recommends in the Music section.

Another important contribution of our work is in the use of collocational analysis to explore the words and clusters appearing close to usernames in comments. Through this analysis, we discovered that usernames play a variety of roles dependent on context of use. On the one hand, they allow commenters to address one another personally, offering thanks, congratulations and apologies. On the other hand, the inclusion of usernames allows commenters to develop their own online identities by positioning their views in relation to those of other commenters. Through stancetaking commenters can index and lay claim to particular identities for themselves. In addition, they can project undesirable properties or statuses onto others. In our data we see the tools (i.e. the lexis) with which alignment and disalignment with others is performed in the medium of newspaper comment threads where the interactions are largely dyadic in nature. However, in a public forum like this, a wider audience of overhearers (i.e. other *Guardian* commenters and readers) is present in significant numbers. Given that some commenters are extremely active, managing the identity they present to a wider audience may well be of high importance to them. If identity can be constructed from a history of stancetaking, as Jaffe (2009) suggests, then each user's online identity will in part be constructed from a complex combination of multiple interactions of the nature demonstrated by the examples discussed throughout the chapter.

Overall, our large-scale data-driven approach has revealed key patterns and trends in comment data, spanning multiple comment threads over a four-year period, which would not be detectable through the analysis of individual interactions in isolation. Thus, we feel our work offers further evidence of the vital role corpus linguistic methods can play in the analysis of pragmatic and discursive features in a vast online environment.

References

Aston, Guy, and Lou Burnard. 1997. *The BNC Handbook: Exploring the British National Corpus with SARA*. Edinburgh: Edinburgh University Press.

Bastian, Mathieu, Sebastien Heymann, and Mathieu Jacomy. 2009. "Gephi: an Open Source Software for Exploring and Manipulating Networks." *International AAAI Conference on Weblogs and Social Media*.

Boczkowski, Pablo J., and Eugenia Mitchelstein. 2011. "How Users Take Advantage of Different Forms of Interactivity on Online News Sites: Clicking, E-Mailing, and Commenting." *Human Communication Research* 38 (1): 1–22. https://doi.org/10.1111/j.1468-2958.2011.01418.x

Crowston, Kevin, and Michelle Williams. 2000. "Reproduced and Emergent Genres of Communication on the World-Wide Web." *The Information Society* 16 (3): 201–216. https://doi.org/10.1080/01972240050133652

Davies, Mark. 2008. *The Corpus of Contemporary American English*. http://corpus.byu.edu/coca/

Dunning, Ted. 1993. "Accurate Methods for the Statistics of Surprise and Coincidence." *Computational Linguistics* 19 (1): 61–74.

Eller, Monika. 2018. "'no prizes to anybody spotting my typo, by the way': The Interplay between Criticism and Identity Management in the Comments Sections on Newspaper Websites." In *The Discursive Construction of Identities On- and Offline*, ed. by Birte Bös, Sonja Kleinke, Sandra Mollin, and Nuria Hernández, 177–202. Amsterdam: John Benjamins.

Eriksen, Lars B., and Carina Ihlström. 2000. "Evolution of the Web News Genre: The Slow Move beyond the Print Metaphor." In *Proceedings of the Thirty-Third Hawaii International Conference on System Sciences*. Los Alamitos, CA: IEEE. https://doi.org/10.1109/HICSS.2000.926697

Hermida, Alfred, and Neil Thurman. 2008. "A Clash of Cultures: The Integration of User-Generated Content within Professional Journalistic Frameworks at British Newspaper Websites." *Journalism Practice* 2 (3): 343–356. https://doi.org/10.1080/17512780802054538

Herring, Susan C. 1999. "Interactional Coherence in CMC." *Proceedings of the 32nd Hawaii International Conference on System Sciences*.

Herring, Susan C. 2013. "Discourse in Web 2.0: Familiar, Reconfigured, and Emergent." In *Georgetown University Round Table on Languages and Linguistics 2011: Discourse 2.0: Language and New Media*, ed. by Deborah Tannen, and Anna M. Trester, 1–25. Washington, DC: Georgetown University Press.

Hundt, Marianne, and Christian Mair. 1998. "'Agile' and 'Uptight' Genres. The Corpus-Based Approach to Language Change in Progress." *International Journal of Corpus Linguistics* 4 (2): 221–242. https://doi.org/10.1075/ijcl.4.2.02hun

Jaffe, Alexandra (ed). 2009. *Stance: Sociolinguistic Perspectives*. Oxford: Oxford University Press. https://doi.org/10.1093/acprof:oso/9780195331646.001.0001

Johansson, Stig, Geoffrey Leech, and Helen Goodluck. 1978. *Manual of Information to Accompany the Lancaster-Oslo/Bergen Corpus of British English, for Use with Digital Computers*. Oslo: Department of English, University of Oslo.

Kehoe, Andrew, and Matt Gee. 2009. "Weaving Web Data into a Diachronic Corpus Patchwork." In *Corpus Linguistics: Refinements and Reassessments*, ed. by Antoinette Renouf, and Andrew Kehoe, 255–279. Amsterdam: Rodopi. https://doi.org/10.1163/9789042025981_015

Kehoe, Andrew, and Matt Gee. 2012. "Reader Comments as an Aboutness Indicator in Online Texts: Introducing the Birmingham Blog Corpus." In *Aspects of Corpus Linguistics: Compilation, Annotation, Analysis*, ed. by Signe Oksefjell Ebeling, Jarle Ebeling, and Hilde Hasselgård. Helsinki: VARIENG. Available online at: http://www.helsinki.fi/varieng/journal/volumes/12/kehoe_gee/ (accessed 15 April 2019).

Kucera, Henry, and W. Nelson Francis. 1967. *Computational Analysis of Present-Day American English*. Providence: Brown University Press.

Langlotz, Andreas, and Miriam A. Locher. 2012. "Ways of Communicating Emotional Stance in Online Disagreements." *Journal of Pragmatics* 44: 1591–1606. https://doi.org/10.1016/j.pragma.2012.04.002

Lutzky, Ursula, and Matt Gee. 2018. "'I Just Found Your Blog.' The Pragmatics of Initiating Comments on Blog Posts." *Journal of Pragmatics* 129: 173–184. https://doi.org/10.1016/j.pragma.2018.03.016

Mishne, Gilad, and Natalie Glance. 2006. "Leave a Reply: An Analysis of Weblog Comments." *Third Annual Workshop on the Weblogging Ecosystem (WWW 2006)*.

OED. *Oxford English Dictionary*. Second edition, online. Oxford: Oxford University Press. http://www.oed.com

Ruiz, Carlos, David Domingo, Josep L. Micó, Javier Díaz-Noci, Koldo Meso and Pere Masip. 2011. "Public Sphere 2.0? The Democratic Qualities of Citizen Debates in Online Newspapers." *International Journal of Press/Politics* 16 (4): 463–487. https://doi.org/10.1177/1940161211415849

Santana, Arthur D. 2011. "Online Readers' Comments Represent New Opinion Pipeline." *Newspaper Research Journal* 32 (3): 366–381. https://doi.org/10.1177/073953291103200306

Appendix

Table 6. The top 50 collocates of usernames in the comments (span 4, sorted by z-score, stopwords and usernames removed) including the most frequent three word clusters in which they occur. Freq. shows the number of times each collocate occurs near a username. The numbers in brackets show the frequency of each word cluster (minimum frequency 10)

Collocate	Freq.	Score	Most frequent 3 word clusters
thanks	25,456	670.04	thanks for the (5,153), thanks for your (2,123), thanks for that (1,574)
agree	33,431	577.24	i agree with (8,551), agree with you (2,711), i agree that (1,502)
hi	5,878	362.44	hi thanks for (19), say hi to (15)
sorry	16,111	362.13	i am sorry (515), i'm sorry but (469), sorry but i (318)
thank	10,668	358.04	thank you for (3,510), thank you i (327), thank you very (295)
re	7,857	306.59	re your comment (48), re your post (47), good point re (38)
think	45,622	294.40	i don't think (4,377), i think you (3,941), i think the (2,361)
post	14,983	282.99	your post is (237), for your post (151), the post by (150)
said	22,967	282.58	as i said (747), i never said (342), well said i (227)
point	21,455	232.15	the point is (750), a good point (629), point is that (600)
dear	5,935	228.71	dear oh dear (127), oh dear i (104), oh dear you (94)
oh	11,715	228.16	oh come on (366), oh i see (133), oh i don't (128)
ah	4,234	210.73	ah yes the (148), ah i see (139), ah the old (54)
right	23,218	208.37	you are right (1,808), you're quite right (336), is right the (312)
says	8,982	198.46	says it all (173), says it is (87), says this is (79)
spot	4,615	196.54	is spot on (396), absolutely spot on (213), spot on the (173)
cheers	2,733	186.35	cheers for the (349), cheers for that (160), cheers for your (22)
donds	2,022	178.08	for the donds (81), thanks for donds (78), donds for the (67)
excellent	5,432	166.95	an excellent point (145), an excellent post (112), is an excellent (90)
posts	5,214	165.19	your posts are (104), the posts of (75), good posts from (65)
apologies	2,743	163.44	apologies for the (148), my apologies i (97), my apologies for (66)
please	8,560	156.97	please tell me (198), could you please (195), can you please (179)
hello	2,096	155.17	say hello to (15), hello to you (12), hello and welcome (11)
comment	12,201	154.71	for your comment (173), your comment is (171), the comment by (132)
agreed	3,247	152.93	i agreed with (56), agreed but i (26), agreed but the (24)
good	23,655	152.41	a good point (629), good to see (493), a very good (388)

(*continued*)

Table 6. (*continued*)

Collocate	Freq.	Score	Most frequent 3 word clusters
yep	1,721	152.38	yep i agree (22), yep i think (22), yep you are (13)
points	6,153	148.63	points out the (270), rightly points out (154), some good points (152)
exactly	6,815	144.18	what exactly is (182), that's exactly what (158), that is exactly (156)
sure	11,475	143.61	i'm not sure (1932), i am sure (588), not sure what (554)
absolutely	5,271	142.41	you are absolutely (355), you're absolutely right (297), is absolutely right (247)
hey	3,086	142.13	No clusters above frequency threshold
reply	2,457	141.60	in reply to (319), for the reply (269), for your reply (246)
indeed	6,624	141.41	it is indeed (125), is indeed a (118), indeed it is (115)
ok	5,415	141.16	ok fair enough (57), ok so you (53), so it's ok (34)
interest-ing	7,161	139.49	an interesting point (178), that's an interesting (175), interesting that you (153)
fair	6,293	138.85	to be fair (964), fair enough i (300), fair enough but (256)
lol	2,089	134.85	lol i think (19), lol are you (11)
know	19,773	134.53	i don't know (2806), do you know (897), i know what (492)
read	11,062	134.23	if you read (418), you read the (412), i have read (306)
dond	1,128	131.15	for the dond (96), a dond for (73), a dond to (43)
actually	12,356	130.07	actually i think (250), you actually read (171), do you actually (164)
posted	2,832	125.55	i posted a (45), posted a link (40), i have posted (33)
nice	6,115	124.93	nice to see (455), would be nice (124), nice try but (110)
disagree	3,370	122.48	i disagree with (363), i don't disagree (252), have to disagree (233)
correct	3,957	121.43	you are correct (499), correct me if (144), is correct in (84)
really	16,713	114.94	do you really (654), you really think (388), you really are (333)
true	7,289	114.80	it is true (247), this is true (223), it's true that (136)
quite	9,525	114.70	you are quite (428), you're quite right (336), are quite right (265)
mate	2,378	111.12	sorry mate but (88), spot on mate (55), sorry mate i (34)

From ‘country’ to ‘confederation’ – debating terms of reference for the EU on a Wikipedia talk page

Susanne Kopf
Vienna University of Economics and Business

This chapter deals with Wikipedia editors’ (‘Wikipedians’) discussions about the EU between 2001 and 2015 on the talk page (TP) accompanying the Wikipedia article “European Union” (EU). This TP provides a platform for Wikipedians to debate controversial issues regarding the article on the EU. Taking a corpus-assisted approach to these debates allows an insight into how and why Wikipedians apply and discard different terms of reference and concepts for the EU. Findings indicate that, in particular, the applicability of the terms ‘confederation’, ‘federation’ and ‘country’ was heatedly discussed throughout the TP conversations. The key factor that determined whether Wikipedians supported or rejected use of these terms was how much sovereignty the term in question ascribed to the EU and its members.

Keywords: European Union, Wikipedia, Corpus-Assisted Discourse Studies, public sphere, digital discourse, terms of reference

“We must build a kind of United States of Europe.” (W. Churchill)

1. Introduction

As the quote by Winston Churchill from 1946 indicates, ideas of European cooperation or even unification have been discussed since before the establishment of the EU and its predecessor organizations. What is even more striking is that – more than 50 years later – uttering the same sentence in the context of a Wikipedia discussion page on the EU would probably cause an outcry and trigger yet another discussion about the current nature of the EU amongst Wikipedia editors (also called Wikipedians). Indeed, a substantial part of Wikipedians’ debates about the EU between 2001 and 2015 revolved around how to refer to the present-day

https://doi.org/10.1075/pbns.306.06kop

European Union and the associated views on the EU's identity – as a country, federation, confederation or something else entirely (see Nevala and Lutzky, this volume, for more on nominal forms of reference).

These debates among Wikipedia contributors take place on sites separate from, but connected to, the widely known Wikipedia article pages, i.e. the encyclopaedic entries most of us are familiar with. Each of these article pages is accompanied by a so-called talk page (TP) which allows contributors to debate controversial (editing) issues with regard to the article in question. As the Wikipedia article on the European Union was already created in 2001, the talk page accompanying the article has served as a platform for Wikipedians to engage in debate about the institution since then, i.e. editors have had ample time to visit and revisit certain controversies regarding the representation of the EU. This chapter homes in on one aspect of talk page debate in particular, namely the Wikipedia community's negotiations regarding terms of reference used to capture the nature of the EU. Examining the terms of reference used, and their connected concepts, allows a glimpse into the Wikipedia community's perspective on the EU – a glimpse into the EU's identity as constructed by the Wikipedia community.

The next section presents the theoretical background informing this study. This is followed by an overview of the data and associated methods of analysis. Finally, the main findings are presented.

2. The European Union and Wikipedia as a public sphere

Unsurprisingly, the EU has received substantial research attention in linguistics. Among the topics that have been studied are questions about language policy and discourses surrounding topics relevant or connected to the EU, such as EU discourses on un/employment (e.g. Muntigl et al. 2000; Krzyżanowski and Wodak 2011; Gazzola 2016). In addition, research has focused on individuals – EU citizens – to explore the potential emergence of a European identity (e.g. Millar and Wilson 2007). With respect to the internet and the EU, the EU's online presence and representation(s) have been subject to investigation (e.g. Wodak and Wright 2006; Van Os et al. 2007). Furthermore, a considerable body of research – both EU internal and EU external – has addressed and lamented the lack of a public sphere spanning the EU (e.g. Commission of the European Communities 2001; Eriksen 2005; Wright 2007; Fossum and Schlesinger 2008; Bärenreuter et al. 2009; Triandafyllidou et al. 2009). While some research has addressed online platforms and the internet as potentially giving rise to such a European public sphere, it has not yet progressed to one of the consistently most visited websites globally – Wikipedia (Alexa 2016). In fact, Wikipedia suffers from a general lack of linguistic research as past studies are mostly limited to bias detection, examinations of collective authorship and tracing of article development (e.g. Emigh and Herring 2005; Callahan and Herring 2011; Page 2014).

This lack of research persists despite the fact that the collaboratively created online encyclopaedia constitutes a fascinating venue of research. One particularly notable aspect of Wikipedia in the context of this study is its division into outwardly directed encyclopaedic entries and discussion pages. Wikipedia articles and their associated talk pages can be conceived of in Goffman's terms (1959): the articles can be understood as the front stage as they are globally accessible sites intended for public consumption. By comparison, the talk pages are the back stage. While also publicly accessible, these platforms consist of threaded discussions amongst editors that are not intended for an audience beyond individuals involved in the collaborative article-creation process – the purpose of TPs is "to provide space for editors to discuss changes to its associated article" (Wikipedia:Talk Page guidelines, Wikipedia 2016). To elaborate briefly on this point, Wikipedia depends on collaboration and cooperation amongst Wikipedians not least because consensus is "the primary way decisions are made on Wikipedia" (Wikipedia:Consensus, Wikipedia 2016). This means that Wikipedians, who wish to have particular (controversial) content represented in the Wikipedia article, have to negotiate with their peers and convince collaborators why this content ought to be included. By comparison, editors who make changes to the article without others' consent run the risk of incurring disciplinary action, including being permanently blocked from Wikipedia editing (Wikipedia:Blocking policy, Wikipedia 2017).

Another aspect worth noting is that as the TPs are publicly accessible and individuals from potentially all over the world can contribute and engage in discussion, Wikipedia TPs are part of *transnational* public discourse.[1] What is more and particularly notable in light of past research that has highlighted the importance but also criticised the EU's lack of public sphere(s) (see above), the Wikipedia talk page on the European Union can function as a, not European, but general transnational public sphere. That is, an arena that allows people to communicate information and exchange opinions (Habermas 1992: 436). Here, a postmodernist conception of the public sphere is assumed, i.e. the notion that multiple publics exist in parallel, which redresses the effects of inequality of access stemming from Habermas' previous understanding that there is one overarching but highly exclusive public sphere (Fraser 1995: 291). Eriksen's definition of "general public spheres" adds a central component to the classification of Wikipedia talk pages – or at least this

1. Internet access and a degree of digital literacy are prerequisites. Moreover, individuals' rights to access and contribute to Wikipedia without fear of prosecution depend on their countries of residence.

particular talk page – as public sphere: *general* public spheres are defined as arenas for deliberation and opinion formation "not aimed at achieving particular results" (Eriksen 2005: 345; see also Wright 2007: 1170). Indeed, Wikipedia talk pages might serve as general transnational public spheres, i.e. as metaphoric spaces that allow opinion formation amongst a theoretically unlimited number of interlocutors. Still, different public spheres are not equally influential. While Wikipedia has a sizable audience, especially in Europe (Wikimedia Foundation 2011b), the talk pages are not as widely received as the article pages and might not impact public opinion formation directly. However, the *results* of the Wikipedia-internal opinion formation are received by readers accessing Wikipedia articles.

On the TP associated with the European Union, the Wikipedia community negotiates many controversial issues pertaining to the institution. Yet, one aspect stands out among other discussions – the community's struggle with defining the nature of the EU. That is, on this particular TP private individuals negotiate the EU's collective identity (cf. Kleinke et al. 2018: 2) by applying and discarding a number of expressions of reference.[2] Throughout the TP discussions, the Wikipedia community debates whether, and why, certain concepts are applicable to the EU and if particular terms of reference should or should not be used in connection with the organization. In this context, it is important to note that this chapter conceives of the relationship between 'term of reference' and 'concept' as corresponding with signifier and signified. Wikipedians' choosing different terms first to describe and then to refer to the EU gives an insight into how the community makes sense of the institution and how it aims to represent the institution in certain ways on the outwardly-directed front stage. The talk page debates also permit an insight into the key conceptual differences between the particular terms of reference, and how these different conceptions lead to acceptance or rejection of descriptors and then terms of reference applied to the EU. Therefore, this chapter, apart from merely identifying terms of reference applied to the union, also delves into how and why Wikipedians choose to accept or discard certain terms.

3. Data and methodology

Wikipedia provides free access to talk page data. Hence, all TP data, that is, each conversation thread – from the creation of this particular TP in 2001 to December 2015 – were sampled and divided into sub-corpora in accordance with topic focus.

2. It is worth noting that while the Wikipedians involved in the debates may or may not be part of this collective identity, their position in relation to the EU is not subject to examination in this chapter (Davies and Harré 1990: 46; see also Fetzer 2018).

This categorisation by topic was possible due to three factors: first, Wikipedia policy mandates that each TP thread ought to introduce and focus on one topic only (Help:Using Talk Pages, Wikipedia 2016). Second, the Wikipedia community enforces this rule, for instance by stating that "we're starting to go a bit off topic" (Talk:European Union/Archive 16, Wikipedia 2007). Finally, the community repeatedly discusses similar issues, which further facilitated the grouping of conversation threads into sub-corpora.

The sub-corpus examined for this study deals with Wikipedians' repeated debates about the nature of the EU. It encompasses 124 threads and consists of 118,175 tokens, that is, 19 per cent of all TP debate on the EU between 2001 and 2015 focus on this issue. Using the programme AntConc (Anthony 2015), this chapter takes a qualitatively-oriented corpus-assisted approach to this data set. This means that apart from considering frequency rankings and collocations (i.e. the statistically significant co-occurrence of tokens) the focus of examination is on concordance lines. That is, items are always examined as embedded in co-text to avoid mere identification of which items are used in the corpus. Rather, the aim is to understand how, precisely, the Wikipedia community makes sense of the EU in its discussions (McEnery and Wilson 2001: 76).

With regard to the collocation calculation, I use t-score to glean an insight into certainty of co-occurrence of particular items with a span of five tokens to the left and five to the right of the node. In order to shed light on precisely how certain co-occurrences are realised, I examine the concordance lines of collocates. In connection with this, it is also worth noting that Tables 1 to 3 do not merely give an overview of the number of concordance lines of co-occurrences. Rather, since several co-occurrences are part of one and the same posting, I give information on both the number of concordance lines and the number of postings these lines are a part of. This is done in order to avoid overstating the number of rejections or acceptances of using, for instance, 'country' in reference to the EU.

Moreover, to permit a more in-depth understanding of the debates and Wikipedians' reasoning, Toulmin's model of argumentation is used. Among other elements, Toulmin identifies three components necessary for a well-formed argument – claim, data and warrant – but also emphasises that there is no clear form-function relationship, that is, the linguistic realisation of different parts of arguments may vary (Toulmin 2003: 87–88). The claim (C) is the assertion a Wikipedian makes, that is, it is the part of an utterance of which an editor tries to convince the other Wikipedians (2003: 88–90). In contrast, data (D) are defined as the "foundation upon which [a] claim is based" (Toulmin 2003: 90–91). Put in other words, data are the pieces of information – the fact or facts – the claim builds on and follows from. Complementing these two elements, the warrant (W) establishes a connection between data and claim and is thus crucial for any argument.

Toulmin also highlights that, although essential to ensure interlocutors can relate data to claim, warrants are frequently left implicit – omitting the warrant might serve to disguise that the connection between data and claim is dubious (Toulmin 2003: 91). In addition, the representation of social actors is taken into account, i.e. the idea whether an actor is sociologically activated or passivated and whether the actor(s) is / are specified or genericised (Van Leeuwen 1996).

In contrast to the TP, the Wikipedia front stage – the article on the European Union – is not examined in depth. However, Wikipedia provides access to historic versions of articles, which allows me to trace article development. This is done to show how the back stage discourse is reflected and recontextualised on the front stage.

4. Data discussion

4.1 The EU as country

The term ‘country’ collocates with ‘eu’ at a t-score of 7.6. The item’s concordance lines shed light on the central defining feature associated with the term ‘country’, namely sovereignty, for example, country is “usually defined by sovereignty” and “[c]ountries are defined by a set of powers”. That is, in addition to allowing an insight into the central signified when using the term ‘country’ – sovereignty – the community’s rejection or acceptance of this term indicates particular views of the EU as a sovereign or not.

The concordance lines of the co-occurrence ‘country’ and ‘eu’ show that 35 concordance lines, as part of 28 postings, reject the idea that ‘country’ is an adequate term and concept to capture the nature of the EU (see Table 1). By comparison, nine postings in nine lines support the idea that ‘country’ is a concept applicable to the EU.

Table 1. Concordance lines and postings supporting / rejecting ‘EU as country’

28 rejection postings/35 lines*	– unsupported rejections: 14 postings
	– lack of sovereignty: 7 postings
	– lack of unity: 3 postings
	– future speculation: 2 postings
	– miscellaneous: 2 postings
9 support postings/9 lines	

*concordance lines

4.1.1 *The EU is not a country*

The sheer number of postings that reject the term 'country' illustrate that this is the dominant perspective – more than 70 per cent (i.e. 28 postings) reject this term of reference (see row one in Table 1).

The examination of argumentation in the data supports this finding – Wikipedians do not back their view in 14 of the counter-country postings, which might indicate that they do not see the need to defend this view or make a persuasive case for it, e.g. "The EU is not a country. Full stop". This example is particularly illustrative since the poster even emphasises that they do not support their claim by adding a spelled out version of sentence final punctuation "[f]ull stop".

The remaining counter-country postings provide an understanding of what the concept 'country' entails. In fact, seven postings reinforce that the most prominent element that defines the term 'country' appears to be sovereignty, e.g. "[anonymised] doesn't seem to understand the difference between the EU and a sovereign country". Here the editor highlights the "difference between" EU and a country by referring to sovereignty – according to this editor the difference between country and EU is that the former concept describes a sovereign entity. This posting also incorporates:

> note the words independent, sovereign countries. Im not sure how to explain it any simpler than that. But i will leave you with a final message. THE EU IS NOT A COUNTRY[3]

Here, the Wikipedian again stresses "independent, sovereign" country and then proceeds to negate 'EU as country' in capitals letters, i.e. rather forcefully.

Another example of how sovereignty is used to reject the term 'country' was posted in 2004:

> The EU is not a country in that sovereignty ultimatly lies with the member states, who can leave the union at any time. […] The idea that the EU is a 'country' would horrify most Europeans!

The structure of the argument is as follows: data: member countries can leave → warrants: (a) the ability to leave "any time" indicates that sovereignty "lies with" the member states and (b) an entity whose constituents are sovereign cannot be described as a country → claim: the EU is not a country. Thus, rejecting 'country', i.e. a term defined by its status as sovereign (see above), this Wikipedian represents the EU as a non-sovereign entity. Moreover, this posting sheds light on how EU members are envisioned in relation to the EU: The members are described

3. To facilitate readability of my data discussion, I do not include [sic] to signal deviation from standard spelling or grammar each time such an instance occurs in the examples discussed.

as sovereign and this retention of sovereignty eliminates all possibility that the EU is a sovereign / a country. That is, an either / or situation is established where 'members = sovereign' means 'EU = not country' but the inversion of this argument is also thinkable as 'EU = country' would mean that 'members = not sovereign'. Another interesting point is that the data of the argument are actually incorrect – at the time of posting there was no formalised way for an EU member to leave the EU since the Lisbon Treaty, which stipulates exit procedures, came into force only in 2009.

Connected to this, the posting also exemplifies that the Wikipedia community does not wish the EU to attain a status of sovereignty – the editor states that the notion that the EU is a country would horrify 'Europeans' and emphasises this by use of an exclamation mark. The reference to "Europeans" serves to hide that it might be the poster's personal reaction that is given here since the poster definitely cannot speak for "most Europeans". However, the vagueness of the genericised "most Europeans" makes the notion that this group might have such an emotional reaction almost impossible to contest since the group cannot be clearly identified and questioned. The reference to "Europeans" also serves to strengthen the claim that the EU is not a country since allegedly even the EU in-group – Europeans – does not condone the idea. Finally, the alleged negative response of "most Europeans" to the idea of 'EU as country' follows the claim that the EU is not a country. In turn, this idea builds on the notion that EU members are sovereign (and not the EU) with the implied reversal of the argument: 'the EU as country would mean loss of members' sovereignty'. Thus, the poster implies that a loss of sovereignty on the level of the EU members would "horrify most Europeans".

Furthermore, three postings that reject the idea of referring to the EU as 'country' address the degree of unification and integration in the European Union, such as the following example:

> The US is an ethnically and culturally diverse place, but it undoubtedly has a common legal and political culture, which wouldn't be possible without a cohesive wider culture. We can't say that of the EU […] as things are the EU isn't a country in the sense that its members and the US are. […] The EU doesn't have a culture in the sense that a country has.

This excerpts presents a comparison argumentation (Garssen 2001: 92): the poster presents their view of the US as factual and then proceeds to negate a similar state of being for the EU. They argue that the EU is not a country based on the diversity of cultures / lack of united culture within the Union. A similar negated comparison is used in the following example: European citizens "do not associate themselves with the country and nation of the EU anywhere near as strongly as your Chinese American would to the US". This indicates that Wikipedians view the EU as a highly diverse entity without cohesion in the sense of one overarching 'culture' – it is considered more diverse than the US and too diverse to be a country. In turn, the term 'country' is apparently conceived of as describing a relatively homogenous entity with regard to the legal system and politics but also with respect to culture.

Interestingly, four postings that reject 'EU as country' contain speculations about the EU's future and / or about the EU's motivations.[4] For instance, "the CIA […] they think the EU will become a country. We do not crystal ball gaze in this way". This editor ascribes the belief that the EU is moving towards becoming a country to a respected agency outside of Wikipedia and, thereby, makes this view more credible (argumentum ad verecundiam). However, they also separate the Wikipedia community from this source by use of the personal pronoun 'we' as an in-group marker in reference to Wikipedians – the poster posits that "[w]e" do not engage in such speculation by drawing on the metaphor of a fortune teller's crystal ball.

The following is a posting that goes beyond future speculation but ascribes a particular goal-orientation to the European Union:

> decision by some or all member states to cooperate in a number of ways doesn't make the EU a country […] the member states are the 'masters of the treaty'. […] while the EU isn't a country, it does have a tendency to attempt to acquire more country-like attributes, and that might or might not eventually lead to its transformation into a country[comment: 'treaty' is a reference to the treaties that formed the EU]

The poster emphasises member states' position of power and authority by activating them as decision-makers in terms of increasing cooperation. The subsequent use of "masters"[5] further evokes the idea of the EU member states as having dominance, power and authority over the treaty(ies) that form the EU and, by proxy, the EU itself. This is followed by depicting the EU as an actor who 'attempts to acquire X' – the use of the verb 'to attempt' ascribes the EU with goal-orientation and the goal is the accumulation of country-like characteristics. As the poster first rejects the term 'country' on grounds of the member states' master status, and then claims that the EU aims to acquire country-like characteristics, this Wikipedian implies that the EU aims to reduce member states' authority and power. Moreover, while the reader does not learn who is / are the human actor(s) hidden in the

4. Two postings that reject 'EU as country' because of concerns about sovereignty also speculate about the EU's future.

5. It is worth noting that the phrase "masters of the treaty" was previously used in several academic papers, EU-related documents and court decisions.

metonymic reference to the EU, it represents the EU as already able to have goals it can "attempt" to reach.

4.1.2 *The EU is a country*

It is clear that, predominantly, Wikipedians do not consider the term *country* applicable to the EU as is. Out of the overall low number of pro-country postings (nine postings – see Table 1), only one presents a definitive view to that effect: "So how can you define a country? if you look at all things [the EU] is a COUNTRY". Here, a Wikipedian poses the question of how to identify a country. This is followed immediately by an unhedged claim without presenting any data to support this claim. That is, the Wikipedian does not answer their own question or specify what features the EU has that make it a country.

The remaining pro-country postings present very weak forms of the pro-country perspective. One means Wikipedians use to weaken their pro-country postings is avoiding formulations that claim that the EU *is* a country – instead of intensive attributive relational processes they use, for example, material processes. That is, these editors focus on how the EU *acts* (material processes) or on possessive attributive relational processes – what attributes it *has*. For instance, the "EU operate almost as a country" instead of e.g. the "EU is a country". This example already points to another strategy, namely how 'country' is modified and graded. Other examples refer to the EU as having certain "country-like elements", the EU as "semi-country" and as "'somehow almost' a country". The use of the quotation marks in the last example is indicative of the unease with which Wikipedians have started to establish a continuum between country and not-country rather than capturing this relationship as a binary opposition.

4.2 The EU as federation

The term 'federation' collocates with 'eu' at a t-score of 4.6 and, again, the Wikipedia community repeatedly defines federation in terms of sovereignty, e.g.

> A Federation is a group of states that maintain local control of affairs but have a stong central governemnt which also represents the group internationally as a single sovereign power.

The community also draws on 'confederation' to capture different degrees of sovereignty, for instance, "the difference between a federation and a confederation is sovereignty" and the "difference between a confederation and a federation is that constituent parts of the former are legally sovereign". These definitions emphasise the fact that federations are seen as having constituents – "parts" and "group of states" – that are not fully sovereign.

Table 2. Concordance lines and postings supporting / rejecting 'EU as federation'

50 rejection postings/64 lines	– sovereignty: 13 postings
	– the EU as a mix of concepts: 10 postings
	– unsupported rejections: 11 postings
	– future speculation: 9 postings
	– miscellaneous: 7 postings
5 support postings/9 lines	

4.2.1 *The EU is not a federation*

Altogether, there are 73 concordance lines in 55 postings where Wikipedians discuss if the term *federation* is applicable to the EU (see Table 2).[6] 64 concordance lines in 50 postings, i.e. approximately 90 per cent, take a counter-federation stance (see row one in Table 2). The high number of rejections points towards the dominant view that the EU is not a federation.

Argumentation analysis confirms this – eleven postings present claims to that effect without any data to back the claim, e.g. "'federation' is used to denote federal states, which EU is not", "[t]he European Union is not a federation" or "I cannot believe this comment about the EU as a federation… the EU is simply not a federation. Not in any sense. At all. […] absolutely not a federation". The last example is also notable in terms of intensity of rejection ("Not in any sense. At all", "absolutely not"). Not unlike the unbacked rejections of 'EU as country', the fact that Wikipedians make such unsupported claims suggests that they are confident that they do not need to support or defend this position but that it is accepted as factual.

The examination of argumentative strategies also sheds light on which aspects of the concept 'federation' the Wikipedia community takes issue with, i.e. why the EU should not be described as a federation. Editors draw on issues of sovereignty, lack thereof and its general distribution between the EU and its member states as reason(s) to reject 'federation' in more than 20 per cent of rejection postings (13 postings). For instance, one Wikipedian rejects 'EU as federation' and adds "[a] federation is a type of sovereign state". Here the claim 'EU = not federation' is backed by data in the form of a short definition: 'federation = sovereign state'. The warrant is only implied: 'the EU is not a sovereign state', that is, the EU is represented as a non-sovereign entity. Another Wikipedian also claims that the EU is not a federation, then provides the data to their claim ("A federation is […] a state in which the federated units lack total sovereignty"), and thus implies that the EU members are fully sovereign. Another example involves a poster rejecting 'EU as

6. Due to the relatively low number of co-occurrences of 'eu' and 'federation' all concordance lines of 'federation' were examined.

federation' and adding: "[t]oo much associations with a sovereign state, which the EU is not". Here, the data are that 'the EU is not a sovereign state' and since using the term 'federation' would elicit association with sovereignty (warrant), the EU should not be described as such.

The examination of the remaining concordance lines that reject 'federation' shows an additional pattern, namely that Wikipedians conceive of the EU as a mix of several concepts or something novel / different entirely without detailing the characteristics of the EU that allegedly make it such a mix / novel entity. Ten postings reject 'EU as federation' based on the view that the nature of the EU cannot be captured by use of 'federation' alone. Two examples that illustrate this point are "[t]he EU has elements of federation, confederation, and international organization" or "EU is neither a federation nor a confederation in the classical sense". A common structure in the context of explaining the EU as a 'mix' is a spatial metaphor – the EU as an entity is placed in space, i.e. it is or lies in-between concepts, for example "[t]he EU is currently somewhere between a country / federation and an international treaty organisation", "the EU lies between a federation and confederation […]" and "the EU […] nowhere near a federation". Alternatively, the Wikipedia community debates the EU in terms of an ontological metaphor – as a substance – that is 'more than' or 'less than' certain terms can capture: "Although the EU is not a federation in the strict sense, it is far more than a free-trade association" and "it is more than a free trade association, but less than a federation".

As with 'country', the rejections of 'federation' also show that the Wikipedia community draws on the term to speculate about the EU's future, e.g. "there are movements to develop the EU into a federation" and "[s]ome want it to become [a federation], but that certainly would not be accurate at present". Interestingly, the social actors driving this movement or having the desire ("want") are entirely suppressed, which might serve to make the claim that such movements exist incontestable, since nobody can question if a group actually pursues the cited agenda, if no particular group is ascribed this agenda.

Among the remaining rejections, seven postings point to the lack of legal basis for calling the EU a federation ("The EU is not de jure a federation") or find that the concept referred to by the term 'federation' is unclear, too vague or confusing, e.g. "Federation in itself is vaguely defined, so that it becomes too arbitrary to say the EU [is a federation]".

4.2.2 *The EU is a federation*

The low number of postings expressing support for 'EU as federation' indicates that the Wikipedia community opposes the application of this term. In addition, there is only one posting that expresses unmitigated support for the application of 'federation' to the EU:

> very strongly support. Very good term for EU. "Federation is a political entity characterized by a union of partially self-governing states or regions united by a central (federal) government"

Apart from the repeated use of 'very', which functions as an intensifier (Vukovic 2014: 46), the posting is notable as it hinges on an argumentum ad verecundiam – the poster quotes an outside authority (as indicated by the quotation marks). The poster – similar to Wikipedians who reject 'federation' – makes reference to the idea of sovereignty ("self-governing"), but here it is used to support 'EU as federation'. Moreover, the argument structure shows that the warrant remains implicit but is actually a vital aspect of how this Wikipedian conceives of the EU (see Figure 1).

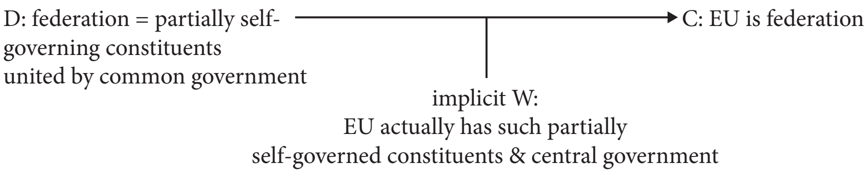

Figure 1. EU as federation

The data 'federation = partially self-governing constituents' back the claim 'EU = federation' connected by the warrant 'since the EU has partially self-governing members'. Thus, this Wikipedian implies that the EU member states are (only) partially self-governing – a view that is markedly different from the above idea of member states as 'masters of the treaties'. It is also worth noting that this poster's understanding of 'federation' contains the notion of discrete and identifiable constituents that form a centrally governed entity.

The remaining postings supporting 'EU as federation' do not claim that the EU is, in fact, a federation as is. Not unlike the Wikipedians who reject 'federation' on the basis of the EU being a mix of concepts, the pro-federation postings acknowledge that 'federation' does not capture the nature of the EU entirely. However, in contrast to the rejections, these postings argue that the EU can be best described by applying a number of concepts *including* 'federation', e.g. "I would add the term 'federation' to the mix", i.e. to the mix of terms used to describe and then refer to the EU. Another Wikipedian also suggests using a hybrid term and draws on an argument from authority to do so: "CIA fact book describes it as a hybrid Confederation-Federation".

4.3 The EU as confederation

The term 'confederation' collocates with 'eu' at a t-score of 7.1. Interestingly, the signifié of this term is different from 'country' and 'federation' in terms of sovereignty. While 'country' is understood as a sovereign entity and 'federation' as a sovereign entity consisting of non-sovereign constituents, 'confederation' is defined as comprising sovereign constituents. In 2004, confederation is initially discussed as "a large state composed of many self-governing regions" but this view is immediately contested with 'confederation' defined as "a group of independent nations, states or tribes more or less permanently united by treaty or alliance for joint action". This latter understanding is re-enforced in 2006 ("an association of sovereign states or communities, usually created by treaty but often later adopting a common constitution"). In 2008, another poster defines confederation as a "permanent union of sovereign states for common action in relation to other states". Repeatedly, varying degrees of sovereignty are used to distinguish 'confederation' from 'federation'. For example, in 2012, in response to a poster's definition of 'confederation' as "permanent union of political units for common action in relation to other units", 'sovereignty' is highlighted: the "difference between a confederation and a federation is that constituent parts of the former are legally sovereign" (see also Section 4.2).

Examining the co-occurrences of 'eu' and 'confederation' yields 28 postings that comment on the idea of the 'EU as confederation' – 16 postings reject this concept for the EU whereas six postings support the application of the term to the EU (see Table 3).

Table 3. Concordance lines and postings supporting / rejecting 'EU as confederation'

16 rejection postings/24 lines	– lack of sources: 9 postings
	– unsupported rejections: 4 postings
	– sovereignty: 3 postings*
6 support postings/10 lines	
6 neutral postings/6 lines	– no stance, based on lack of sources: 3 postings
	– indifference: 3 postings

*One of these postings also refers to a lack of source material.

4.3.1 *The EU is not a confederation*

As with 'federation' and 'country', 'confederation' is rejected, even though the rejection rate (24 lines in 16 postings, i.e. approximately 60 per cent) is lower than with regard to the former concepts (see Table 3). Still, the rejections outweigh the pro-postings and four rejections even present unsupported claims to this effect, e.g. "it is falsely describing the EU as a confederation" and "completely wrong-headed, EU is not a confederation".

Additionally, 'confederation' is rejected on the basis of different reasons compared to 'country' and 'federation'. Indeed, the dominant reason for Wikipedians to reject 'confederation' is that there are not enough or no reliable sources that describe the EU as a confederation (see row one in Table 3). Nine postings make use of this strategy; for instance, "[e]ither provide an official source describing the EU as a confederation or it needs to be removed" and "it does not matter much whether we believe the EU meets the criteria listed in the article 'confederation'. Only reliable secondary sources do matter".

The issue of sovereignty does not serve to reject 'confederation' as comprehensively as it does, for example, with regard to 'federation'. First, the rejection rate of 'confederation' on the basis of potential implications concerning sovereignty is lower than with regard to 'country' and 'federation'. Second, it is notable *how* sovereignty is debated in the context of 'EU as confederation'. The following example illustrates how a Wikipedian draws on both the issue of sources and of sovereignty to back their claim that the EU is not a confederation:

> […] no law or constitution to say that the EU is a confederation, nor does the EU describe itself as a confederation. All EU members are independent states, this must be reflected in the opening sentence. 'confederation' is speculation and clearly POV, it implies that the EU is sovereign, which it currently is not. ALL EU member states are themelves, independent sovereign states, there is NO confederation.

This posting presents one claim with two supporting data and warrant combinations. The claim of the whole argument remains implicit until the last sentence – the EU is not a confederation. One set of data presented to back this claim is given in the first sentence: the aforementioned lack of sources that declare the EU a confederation and an argumentum ad verecundiam that the EU itself does not use 'confederation' ("no law or constitution to say that the EU is a confederation, nor does the EU describe itself as a confederation"). The warrant for this is given in "is speculation and clearly POV", that is, the Wikipedian reminds interlocutors of key Wikipedia policy, namely to only present factual information and not opinion on Wikipedia – the NPOV policy (Wikipedia:Neutral Point Of View, Wikipedia 2015). The poster thus proceeds from data: 'lack of sources' ("no law or constitution to say […]") to the claim: 'the EU is not a confederation' with the warrant: 'since that would be speculation / POV and Wikipedia does not speculate'.

The Wikipedian also presents another set of data aimed at supporting the same claim. The poster states the data: 'EU members are independent and sovereign' twice and inserts additional data: 'the EU is not sovereign'. Altogether, the structure is as follows: the data 'EU members are sovereign and independent and the EU is not sovereign' supporting the claim 'the EU is not a confederation'. The warrant is left implicit 'since confederation implies that the EU is sovereign'. This posting gives a notable perspective on the relationship between the EU and its members: the statement that the EU is not sovereign is enveloped by the idea that

EU members are fully sovereign. This implies an either / or situation – either the members are fully sovereign and the EU is not sovereign or the EU is sovereign and the members are not.

The Wikipedia community challenges and ultimately rejects this editor's posting. Indeed, the poster's understanding of 'confederation' is questioned. Subsequent postings focus on correcting the poster's understanding of the term, for instance: "To [anonymised], no one is saying that the individual member states are not sovereign. That is not what being a confederation means". Thus, respondents do not focus on whether the EU is sovereign. Rather, they focus on the constituents' roles and emphasise that the term 'confederation' does not mean that its constituents' sovereignty is in any way limited.

Generally, the matter of source material is the most persistent element of Wikipedians' considerations with regard to classifying the EU as a confederation or not. Even when posters do not take a stance for or against defining the EU as a confederation explicitly, they address the issue of source material, e.g. "I don't know if there are sources out there that describe the EU as a confederation".

4.3.2 *The EU is a confederation*

Compared to 16 rejection postings, there are only six postings that support using 'confederation'. Thus, it can be concluded that 'the EU as confederation' is the non-dominant perspective on the EU in this Wikipedia talk page corpus.

The finding that 'the EU as confederation' is the non-dominant conception of the EU is supported by the fact that five of the pro-postings contain mitigating devices. An example includes the clause "it seems to me that EU is indeed a confederation", where "seems to me" conveys that this might be a subjective impression, even though this is slightly counterbalanced by "indeed", which arguably lends some emphasis to the claim that the EU is a confederation. In one posting an editor argues that "'confederation' is obviously contentious, though it can be (and sometimes is) argued that the EU is a type of confederation". This Wikipedian exercises particular caution concerning the concept: first, the poster acknowledges that there is controversy concerning the term. Second, they passivise the verb 'to argue' and thereby suppress the sayer. Therefore, they do not assume responsibility as an individual that argues this point. Third, in terms of epistemic modality, they weaken their claim with the modal verb 'can'. Although they acknowledge that the claim that the EU is a "type of confederation" is made "sometimes", the poster does so in brackets, which weakens the impact of the claim because there is usually less emphasis on a proposition given within brackets. Finally, they use the adverb "sometimes", which further mitigates the claim.

Interestingly, one pro-editor, that is, an editor who supports the use of 'confederation' to describe and to refer to the EU, refers to the issue of source material and evidence to prove or disprove the idea that the EU can be defined as a confederation:

> the evidence the EU is a confederation is somewhat lacking. […] we should adopt colloquial use [of the term confederation], which can be based on a dictionary definition. So in my view, the task for editors objecting to the EU being labelled confederation should provide a reliable source […]

After initially acknowledging that evidence supporting 'the EU as confederation' is problematic, they proceed to commit the fallacy of putting burden of proof on the counter-confederation side of the argument. Thereby they actually violate core Wikipedia policy, which mandates that any content included in the article requires sourcing (Wikipedia:Identifying reliable sources, Wikipedia 2016). This violation is also possibly the reason the poster's peers do not accept this shift in burden of proof and continue to reject 'EU as confederation' based mostly on a lack of sources.

Finally, and comparable to the pro-country postings, the example of "the EU is a type of confederation" illustrates the strategy of grading and adapting the concept of 'confederation'. Another example of this is the use of "most of" in a posting expressing weak support but at least acceptance of using the term 'confederation': the poster deems the use of the term 'confederation' "[a]cceptable The EU fits most of the definition of a confederation, but the term is never used in practice". This editor also weakens support by not giving a straightforward classification of what the EU *is* in the form of an intensive attributive relational process, but uses a metaphorical material process – the EU is considered as merely meeting ("most of") the unnamed criteria that define a confederation. The poster further weakens their pro-confederation stance by adding the adversative conjunction "but" and drawing attention to the fact that, though allegedly mostly applicable, the term is not used in practice.

4.4 The EU as community and union – the Wikipedia article

'Community' and 'union' were not as widely debated as the terms mentioned in Sections 4.1 to 4.3. 'Community' inspired a short debate in 2007 due to the term's connotations – as Musolff (2004) notes, certain terms may evoke notions of close, even familial relation. Indeed, Wikipedians are aware that 'community' might be problematic in this respect due to its potential to frame the EU as a metaphorical family / group of friends: "Have to say I didn't like the word 'community'. […] The word 'community' is not normally used to describe something as hard-edged as a complex business deal like this" and "I remain bothered by 'community', which still to me implies too much a collection of happy friends living together". Finally – on 19 November 2007 – the Wikipedians involved in the debate arrived at "community of states" as the phrase to be used in the Wikipedia article on the EU. The postmodification "of states" was added to avoid association with personal relationships, i.e. a community of *people*.

However, the TP discussion continued despite this apparent consensus to use 'community' in the article, e.g. "I prefer the terms political and economic union" and "[s]urely the European Union can be described as a Union?". Finally, on 23 June 2008, the word 'union' was introduced on the Wikipedia article page (European Union:Revision history, Wikipedia 2016). The Wikipedia community has retained 'union' on the front stage from that point onwards. On the talk page there is only minor disagreement about whether the article's use of 'union' should also contain a link to the Wikipedia article on the concept 'union' or not. This discussion stems from the fact that the article on the concept describes 'union' as a form of state, as one Wikipedian points out: "The first sentence of Wikipedia's article on political unions says: 'A political union is a type of state which is composed of or created out of smaller states'". The Wikipedians ultimately decide to leave the term 'union' in the article but without a hyperlink to the article on 'union' in order to avoid implications of the EU being associated with a sovereign state in any form.[7] In 2012, the term 'union' was once again confirmed for usage: "Strongly support. This was the long-standing consensus, arrived at after lengthy discussions. The European Union describes itself as a union". All in all then, the Wikipedia community arrives at a tautology as outcome of its discussions – describing and referring to the European *Union* as 'union' does not actually provide readers with a clearer understanding of this political entity.

5. Conclusion

This paper has shown that the use of various nominal forms as terms of reference (e.g. 'country', 'federation', 'confederation') allows the Wikipedia community to make sense of the EU's identity in versatile ways. In the discussions about particular terms of reference, the editors consistently refer to sovereignty, that is, the concepts (the signified) underlying the terms (signifiers) debated are predominantly defined and distinguished by the different degrees of sovereignty associated with them. Especially the terms 'federation', 'confederation' and 'country' are differentiated (and rejected) based on their implications concerning sovereignty. In the given corpus, these three terms and connected concepts can be conceived of as placed along a continuum that is graded with respect to the degree of sovereignty

7. State and country are conflated in the given corpus, e.g. "these two words [country and state] are often synonymous".

the overarching entity and its constituents hold. Figure 2 illustrates this scale of sovereignty reaching from country to confederation.

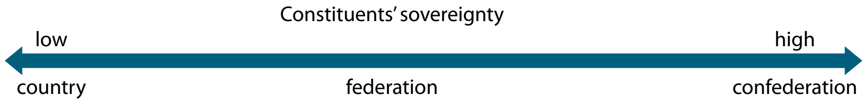

Figure 2. Scale of sovereignty

It is not possible to draw ultimate conclusions with regard to where the Wikipedia community places the EU on this continuum of sovereignty. Still, the community appears to favour a position close to the 'confederation' end of the spectrum seeing as the rejection rate is lowest across the three terms 'country', 'federation' and 'confederation'. What is more, the main reason for rejecting 'confederation' is a lack of source material stating that the EU can be identified as such. In contrast, both 'country' and 'federation' are rejected mainly based on issues connected to sovereignty. Another notable point that this investigation brought to light is that the Wikipedia community ascribes an agenda to the EU or to certain groups of people, namely the goal to further empower the EU and possibly have it progress from confederation to federation to, finally, sovereign unitary country. Generally, this hypothetical change in the EU's status is represented as not desirable.

It is also worth mentioning that the scale of sovereignty does not incorporate 'community' and 'union'. With regard to the former, the concept associated with 'community' was not debated in terms of sovereignty as the term was rejected by Wikipedians due to its implications concerning interpersonal relations (cf. Musolff 2004: 13–29). The term 'union' is particularly notable as the Wikipedia article on 'union' implies that the signified connected to the term might have implications concerning sovereignty. However, the Wikipedians intentionally reject this possible concept of 'union' by not linking to the associated Wikipedia article. That is, in this context, the community actively rejects a conceptual aspect tied to the term of reference 'union' – the community thus appropriates and determines the signified of its chosen signifier.

Unsurprisingly, the back stage TP discussions directly affect the front stage, i.e. the article. The Wikipedia community rejects the three main terms of reference discussed in this chapter and resolves the dilemma through a quasi non-solution, by referring to the European *Union* as 'union'. As already mentioned, the community ensures that this term does not contain implications concerning sovereignty by deliberately discarding the option of creating a link to the Wikipedia article on 'union' (which describes 'union' as a form of merged entity with a degree of sovereignty). Moreover, the back stage affects the front stage inasmuch as the community includes explicit reference to the difficulty of categorising the EU and acknowledges the controversial nature of the issue: “The classification […] has been much debated […] by some criteria, it could be classified as a confederation; but it also has many attributes of a federation” (European Union, Wikipedia 2015).

Overall, my corpus of talk page data from 2001 to 2015 addresses a rather profound issue with respect to the EU, namely the understanding of what type of body the institution is and what degree of power it has over its constituents. What is more, the controversy is not actually resolved conclusively and the community resorts to using the term ‘union’, which does not give an insight into the institution’s standing in terms of sovereignty. Seeing as Wikipedia TPs can function as transnational public spheres that allow discussants to engage in opinion formation, the fact that such a fundamental question concerning the EU’s identity is debated heatedly and left unanswered is notable. It suggests that a transnational although predominantly European audience and community[8] is still conflicted and unsure about key aspects of the EU, which does not bode well for the institution – see, for example Hirsh and Inzlicht (2008) for a discussion of individuals’ tendency to respond negatively to uncertainty and lack of clarity.

References

Alexa. 2016. “The Top 500 Sites on the Web.” http://www.alexa.com/topsites (accessed 15 April 2019).

Anthony, Laurence. 2015. *AntConc*. Tokyo: Waseda University. http://www.laurenceanthony.net/software.html (accessed 15 April 2019).

Bärenreuter, Christoph, Cornelia Brüll, Monika Mokre, and Karin Wahl-Jorgendsen. 2009. “An Overview of Research on the European Public Sphere.” Eurosphere Working Paper Series 3. Available online at https://www.academia.edu/9087443/An_Overview_of_Research_on_the_European_Public_Sphere (accessed 15 April 2019).

Callahan, Ewa, and Susan Herring. 2011. “Cultural Bias in Wikipedia Content on Famous Persons.” *Journal of the American Society for Information Science and Technology* 62 (10): 1899–1915. https://doi.org/10.1002/asi.21577

Commission of the European Communities. 2001. “European Governance: A White Paper.” Nr. 428. Unpublished manuscript. http://europa.eu/rapid/press-release_DOC-01-10_en.htm (accessed 15 April 2019).

Davies, Bronwyn, and Romano Harré. 1990. “Positioning: The Discursive Production of Selves.” *Journal for the Theory of Social Behaviour* 20 (1): 43–63. https://doi.org/10.1111/j.1468-5914.1990.tb00174.x

Emigh, William, and Susan Herring. 2005. “Collaborative Authoring on the Web: A Genre Analysis of Online Encyclopedias.” *Proceedings of the 38th Annual Hawaii International Conference on System Sciences*, 1–11.

8. The typical Wikipedian is from Europe or the US (See e.g. Wikimedia Foundation (2011a) for more demographic information about Wikipedians).

Eriksen, Eric. 2005. “An Emerging European Public Sphere.” *European Journal of Social Theory* 8 (3): 341–363. https://doi.org/10.1177/1368431005054798
European Union. Wikipedia, 2015. https://en.wikipedia.org/w/index.php?title=European_Union&oldid=695821747 (accessed 5 February 2016).
European Union: Revision history. Wikipedia, 2016. https://en.wikipedia.org/w/index.php?title=European_Union&action=history (accessed 19 January 2016).
Fetzer, Anita. 2018. “And You Know, Jeremy, My Father Came from a Very Poor Background Indeed: Collective Identities and the Private-Public Interface in Political Discourse.” In *The Discursive Construction of Identities On- and Offline*, ed. by Birte Bös, Sonja Kleinke, Sandra Mollin, and Nuria Hernández, 227–248. Amsterdam: John Benjamins.
Fossum, John, and Philip Schlesinger. 2008. *The European Union and the Public Sphere: A Communicative Space in the Making?* London, New York: Routledge.
Fraser, Nancy. 1995. “Politics, Culture, and the Public Sphere: Toward a Postmodern Conception.” In *Social Postmodernism: Beyond Identity Politics*, ed. by Linda Nicholson, and Steven Seidman, 287–312. Cambridge, New York: Cambridge University Press. https://doi.org/10.1017/CBO9780511520792.012
Garssen, Bart. 2001. “Argument Schemes.” In *Crucial Concepts in Argumentation Theory*, ed. by Frans van Eemeren, 81–100. Amsterdam: Amsterdam University Press.
Gazzola, Michele. 2016. “Multilingual Communication for Whom? Language Policy and Fairness in the European Union.” *European Union Politics* 17 (4): 546–569. https://doi.org/10.1177/1465116516657672
Goffman, Erving. 1959. *The Presentation of Self in Everyday Life*. Garden City, New York: Anchor Books.
Habermas, Jürgen. 1992. *Faktizität und Geltung: Beiträge zur Diskurstheorie des Rechts und des demokratischen Rechtsstaats*. Frankfurt am Main: Suhrkamp.
Help: Using Talk Pages. Wikipedia, 2016. https://en.wikipedia.org/wiki/Help:Using_talk_pages#Indentation (accessed 21 January 2016).
Hirsh, Jacob B., and Michael Inzlicht. 2008. “The Devil You Know: Neuroticism Predicts Neural Response to Uncertainty.” *Psychological Science* 19 (10): 962–967. https://doi.org/10.1111/j.1467-9280.2008.02183.x
Kleinke, Sonja, Nuria Hernández, and Birte Bös. 2018. “Introduction: Identity Construction in Complex Discourse Contexts.” In *The Discursive Construction of Identities On- and Offline*, ed. by Birte Bös, Sonja Kleinke, Sandra Mollin, and Nuria Hernández, 1–12. Amsterdam: John Benjamins.
Krzyżanowski, Michal, and Ruth Wodak. 2011. “Political Strategies and Language Policies: The European Union Lisbon Strategy and Its Implications for the EU’s Language and Multilingualism Policy.” *Language Policy* 10 (2): 115–136. https://doi.org/10.1007/s10993-011-9196-5
McEnery, Tony, and Andrew Wilson. 2001. *Corpus Linguistics: An Introduction*. Second Edition. Edinburgh: Edinburgh University Press.
Millar, Sharon, and John Wilson (eds). 2007. *The Discourse of Europe: Talk and Text in Everyday Life*. Amsterdam: John Benjamins. https://doi.org/10.1075/dapsac.26
Muntigl, Peter, Gilbert Weiss, and Ruth Wodak (eds). 2000. *European Union Discourses on Un/employment: An Interdisciplinary Approach to Employment, Policymaking and Organizational Change*. Amsterdam: John Benjamins. https://doi.org/10.1075/dowi.12
Musolff, Andreas. 2004. *Metaphor and Political Discourse*. London: Palgrave Macmillan. https://doi.org/10.1057/9780230504516

Page, Ruth. 2014. "Counter Narratives and Controversial Crimes: The Wikipedia Article for the 'Murder of Meredith Kercher'." *Language and Literature* 23 (1): 61–76. https://doi.org/10.1177/0963947013510648

Talk: European Union/Archive 16. Wikipedia, 2007. https://en.wikipedia.org/wiki/Talk:European_Union/Archive_16 (accessed 20 December 2015).

Toulmin, Stephen Edelston. 2003. *The Uses of Argument*. Cambridge, New York: Cambridge University Press. https://doi.org/10.1017/CBO9780511840005

Triandafyllidou, Anna, Ruth Wodak, and Michal Krzyżanowski (eds). 2009. *The European Public Sphere and the Media: Europe in Crisis*. Basingstoke, New York: Palgrave Macmillan. https://doi.org/10.1057/9780230271722

Van Leeuwen, Theo. 1996. "The Representation of Social Actors." In *Readings in Critical Discourse Analysis*, ed. by Carmen R. Caldas-Coulthard, and Malcom Coulthard, 32–70. London, New York: Routledge.

Van Os, Renée, Nicholas Jankowski, and Maurice Vergeer. 2007. "Political Communication about Europe on the Internet During the 2004 European Parliament Election Campaign in Nine EU Member States." *European Societies* 9 (5): 755–775. https://doi.org/10.1080/14616690701412871

Vukovic, Milica. 2014. "Strong Epistemic Modality in Parliamentary Discourse." *De Gruyter Open Linguistics* 1 (1): 37–52. https://doi.org/10.2478/opli-2014-0003

Wikimedia Foundation. 2011a. "Research:Wikipedia Editor Survey 2011/Results." https://upload.wikimedia.org/wikipedia/commons/7/76/Editor_Survey_Report_-_April_2011.pdf (accessed 21 March 2019).

Wikimedia Foundation. 2011b. "Research:Wikipedia Readership Survey 2011/Results." https://meta.wikimedia.org/wiki/Research:Wikipedia_Readership_Survey_2011/Results (accessed 18 October 2017).

Wikipedia:Blocking policy. Wikipedia, 2017. https://en.wikipedia.org/wiki/Wikipedia:Blocking_policy (accessed 26 March 2017).

Wikipedia:Consensus. Wikipedia, 2016. https://en.wikipedia.org/wiki/Wikipedia:Consensus (accessed 17 November 2016).

Wikipedia:Identifying reliable sources. Wikipedia, 2016. https://en.wikipedia.org/wiki/Wikipedia:Identifying_reliable_sources (accessed 7 October 2016).

Wikipedia:Talk Page guidelines. Wikipedia, 2016. https://en.wikipedia.org/wiki/Wikipedia:Talk_page_guidelines (accessed 19 January 2016).

Wodak, Ruth, and Scott Wright. 2006. "The European Union in Cyberspace: Multilingual Democratic Participation in a Virtual Public Sphere?" *JLP* 5 (2): 251–275. https://doi.org/10.1075/jlp.5.2.07wod

Wright, Scott. 2007. "A Virtual European Public Sphere? The Futurum Discussion Forum." *Journal of European Public Policy* 14 (8): 1167–1185. https://doi.org/10.1080/13501760701656403

After we #VoteLeave we can #TakeControl

Political campaigning and imagined collectives on Twitter before the Brexit vote

Sylvia Jaworska and Tigran Sogomonian
University of Reading

This study explores the use of personal pronouns in the context of political agitation on Twitter in the run-up to the EU referendum 2016. Using a combination of corpus linguistic and discourse analytical techniques, it shows notable differences in the way in which personal pronouns were employed by Leavers and Remainers. In particular, *we* emerged as a significant factor distinguishing the online rhetoric of the two camps with Leavers using *we* more often and in a much more versatile manner. This study contributes to the growing body of research on ambient affiliations in political communication offering insights into the ways in which personal pronouns are strategically deployed to create imagined collectives for the purpose of political bonding and agitation online.

Keywords: Twitter, Brexit, corpora, identity construction, personal pronouns, ambient affiliations, imagined collectives

1. Introduction

Brexit has been one of the major political events in modern British history. In the EU Referendum held on 23 June 2016, the Leave campaign won over the pro-European Remain campaign with a narrow majority of 51.9%. Much of the Leavers' success has been attributed to populist rhetoric fuelled by the fear of immigration, economic losses and strong nationalist sentiments. While this rhetoric undoubtedly played an important role in swaying the public opinion towards Brexit, some commentators argue that it was not just economic inequality or concerns over immigration but a wider range of issues pertaining to values and identities that motivated the Leavers (Kaufmann 2016).

Similar to other political events such as the Arab Spring or Occupy Movement, social media sites, Twitter in particular, have been considered influential and

https://doi.org/10.1075/pbns.306.07jaw

consequential in swaying public opinion during the run-up to the EU referendum (Grčar et al. 2017). Researchers in political and computational sciences see contents of tweets as mirrors of public attitudes and political sentiments, and hail Twitter as a predictor of election results (Tumasjan et al. 2010; Di Fatta et al. 2015; Grčar et al. 2017). Not just the content of messages sent on Twitter is important; its affordance to reach large and diverse audiences, the portability and replicability of messages (boyd 2011) and the possibility of establishing and strengthening socio-emotional relationships, or what Zappavigna (2014) calls a "communion of feeling", makes Twitter a particularly useful tool for political agitation, which very much depends on creating affiliations and allegiances (Fetzer and Bull 2008).

Despite the significance of Twitter in political campaigning (Jungherr 2014; Papacharissi 2015), linguistic practices of networked bonding around political issues have to date received less attention (but see Coesemans and De Cock 2017; Sadler 2018). Linguistic devices for identity construction in political communication, specifically devices such as personal pronouns and metaphors used to construct in- and out-groups, have been extensively investigated under the methodological umbrella of Critical Discourse Analysis (CDA), but most of this research has been concerned with offline texts and practices (Wodak et al. 1999; Charteris-Black 2004; Baker et al. 2008; Musolff 2015). Although the distinction between offline and online is blurred, mobile technologies do come with certain affordances and constraints that shape the ways in which political identities can be constructed and communicated. The broadcastability and accessibility of Twitter allows for communication with large audiences in a short spell of time without any delays caused by editorial boards or other authoritarian bodies. At the same time, the word limit of 140 characters (increased to 280 in 2017) requires users to use this space economically, condensing tweets into key messages (Coesemans and De Cock 2017). The shortness of messages and their mostly episodic, bullet point-like style (chronicling of events and sharing information) (Humphreys et al. 2013) means that the typical cohesive flow of written texts established through the use of anaphora and cataphora is absent. The resultant contextual absences can instil referential ambiguity, for example, in our understanding of personal pronouns and what or who they stand for. Sometimes this ambiguity could be rhetorically *useful* in that it could appeal to diverse audiences (Fetzer and Bull 2008) and multiple audiences characterise communication on social media sites such as Twitter.[1] Other times, tweeters might offer specific contextualisation cues to make sure that a particular identity is meant and not another.

1. In fact, the presence of multiple and specified audiences makes it impossible to know precisely who receives a message posted on Twitter – something which Marwick and boyd (2010) call "context collapse".

The focus of this study is on the use of personal pronouns in the context of political agitation on Twitter. Studying the use of personal pronouns in political communication is significant for at least three reasons: Personal pronouns are the paramount linguistic devices of indexicality used to refer to the self and other(s); they act as important tools of signalling and maintaining individual or group membership; and greater use of personal pronouns, especially first and second person pronouns, is considered indicative of an involved style, whereas lesser use of these devices points to a nominal and impersonal style (Biber and Conrad 2009). Because political campaigns build strongly on involvement to create affiliations and allegiances, the extent to which personal pronouns are deployed could indicate on the one hand the degree of involvement, and on the other hand the strategies of in- and out-grouping. Some personal pronouns that index identities, specifically *we* and *they*, are ambiguous and fuzzy in that they can simultaneously refer to multiple entities or identities hidden in the nouns or noun phrases for which they stand (Fetzer and Bull 2008; Kleinke and Bös 2018). This ambiguity makes them *useful* strategic devices in political communication allowing politicians to divert attention from themselves, to shift responsibilities, or to strengthen allegiances when necessary.

This study, therefore, explores the use of personal pronouns in the context of political agitation on Twitter in the run-up to the EU referendum. It is particularly interested in demonstrating the extent and types of personal pronouns used by Leavers and Remainers on Twitter to gauge the degree of involvement with ambient audiences. Given that some pronouns have a dual function, for example, *we* can include or exclude depending on the context, this study explores in more detail the specific kind of affiliations and memberships that influential tweets produced by the two camps evoke through the use of *we*.

Influential tweets in our study are tweets that were produced by politicians, public figures and political organisations affiliated with or representing either the Leave or Remain campaign and that were retweeted at least 100 times. This study contributes to the growing body of research on networked identities and imagined collectives (boyd 2011) showing the ways in which personal pronouns are strategically deployed to create imagined collectives for the purpose of political bonding and agitation online.

2. The role of pronouns in political persuasion

Personal pronouns are the paramount linguistic devices used to refer to the self or the other, often in replacement of a noun or noun phrase to which they can be anaphorically or cataphorically traced (Quirk et al. 1985). Although the relationship between pronouns and the noun or noun phrase for which they stand is grammatically straightforward, the actual use of personal pronouns in discourse is not always so clear cut (Fetzer and Bull 2008). With the exception of *I* which is unambiguous, other first and second person pronouns have a wide referential range and flexibility (Pavlidou 2014). For example, the first person plural *we* can refer to just a speaker and listener (*I* and *you*) but also extend to larger groups or identity memberships such as a political party, a government, a country or the whole of humanity (Pavlidou 2014). Quirk et al. (1985) identify eight special uses of *we* including the self-referential *we*, the generic *we*, the inclusive authorial *we*, the editorial *we*, the rhetorical *we, we* in reference to the hearer, in reference to a third party, and the rarely used royal *we*. Mühlhäusler and Harré (1990) classify the referential meanings of *we* into two groups: inclusive *we* referring to speaker, hearer and possibly some other people, and exclusive *we* including references to a group of people including the speaker but excluding the hearer (the authorial *we*, the editorial *we* and the royal *we*).

The polyvalent character of some personal pronouns makes them *convenient* rhetorical devices in political spheres, which depend on affiliations and are inherently driven by in- and out-grouping (Fetzer 2014). Pronouns can be useful for asserting allegiances and drawing boundaries between groups (*us* vs. *them*). Equally, they can signal inclusiveness and be strategically used to appeal to larger audiences. *We* in particular has been shown to be a strategic tool of constructing, reconstructing and deconstructing collectivity (Fetzer 2014).

A great deal of attention has been dedicated to pronouns in studies investigating discursive representations of marginalised groups, specifically immigrants and ethnic minorities. In the field of CDA, there is now a large body of research documenting diverse linguistic strategies of 'othering', often underpinned by specific uses of pronouns. For example, Wodak et al. (1999) show how in right-wing agitation in Austria, the plural *we* was strategically used to refer to one particular national identity (Austrians), while ethnic minorities were referred to with a vague impersonal *they* without making clear to which group *they* actually belonged. Similar trends have been established in nationalist propaganda found in other national and political contexts including the UK, Germany, Russia, and Poland.

The differentiated use of personal pronouns is not only restricted to right-wing or extremist propaganda; politicians across the board use pronouns manipulatively for political effects. For example, Wilson (1990) has shown how pronominal references can be strategically used to divert attention from a political persona, while De Fina (1995) demonstrates how in a self-promotional manner they can help present a speaker as an active agent. In more recent research, Fetzer and Bull (2008) investigate the use of pronouns in political interviews and show how pronominal shifts (for example, from *you* to *we* and vice versa) can be used as a means of both over- and under-inclusion and as devices of evasion to discursively *wriggle out* of difficult questions. Analysing a panel discussion about Europe's future in the context of the 2008 annual meeting of the World Economic Forum, Cramer (2010) explores pronouns as tools of indexing Europeanness. The study shows how, for example, the inclusive *we* is used to construct European identity and alignment with European values and beliefs.

Given the rhetorical usefulness of pronouns, it is not surprising that they are a salient feature of political communication. Tyrkkö (2016) investigates quantitatively the use of personal pronouns in 875 political speeches delivered since the early nineteenth century. The diachronic analysis shows a significant rise in the use of personal pronouns over the last two centuries. Specifically, the self-referential plural pronouns *we, us*, and *our* increased quite dramatically, whereas references to the self (*I, me* and *mine*) experienced a substantial reduction. The author argues that this change reflects a shift from a person-centred rhetoric to a more group-centred style brought about by wider changes in democratic societies, dislike for political self-promotion, and the development of broadcast and technology. Interestingly, a rise of references to *the other* has been observed at least since the early twentieth century, suggesting a stronger tendency for exclusiveness through in- and out-grouping.

The vast majority of studies concerned with the use of pronouns in political contexts examined a variety of offline contexts, whereas increasingly important online communication has received little attention. A recent study by Coesemans and De Cock (2017) presents the first important contribution to our understanding of the use of pronominal references in the context of political microblogging. Studying tweets produced by Belgian and Spanish politicians during the 2014 European elections campaign, the authors identify several communicative practices that politicians adopt and adapt in order to increase their online visibility and influence. First, the authors notice a prominent use of self-references through first person pronouns. Second, through retweets, hashtags, and Twitter handles, politicians also use third person means to refer to themselves. The authors conclude that in the studied contexts politicians have shown a heightened awareness of the affordances and constraints of Twitter and adapted their communicative strategies accordingly. They utilised the platform not just for political messaging but as a tool for personal and professional branding, reflecting a more person-centred style, rather than the group-centred style observed in political speeches in the study by Tyrkkö (2016). Thus, it seems that communication on social media sites such as Twitter 'encourages' politicians to be more self-centred. Yet, it must be noted that the politicians whose tweets were analysed by Coesemans and De Cock (2017) were individuals competing for seats in the European parliament and thus self-presentation played an important part in this contest. This might not be the same in the context of collective actions such as the Leave and Remain campaigns. Against this background, this paper endeavours to address the following research questions:

1. Which personal pronouns were utilised in influential Leave and Remain tweets and to what extent?
2. Does the use of personal pronouns in influential Leave and Remain tweets pertain to a more person- or group-centred style?
3. What kind of identities and memberships did influential Leave and Remain tweets draw on to mobilise voters?

Whereas the first two questions are answered using a quantitative corpus-based methodology, specifically adopting the framework proposed by Tyrkkö (2016), the third question draws on qualitative discourse-analytical techniques used in previous research on the use of pronouns in political communication (e.g. Fetzer and Bull 2008). The next section outlines in detail the data collection procedures and methods used to conduct this study.

3. Data and methods

The data under investigation comprises two corpora consisting of 1,456 Leave tweets (30,302 words) and 1,458 Remain tweets (35,691 words) respectively, which were retweeted in the run-up to the EU referendum. They were compiled from two sources of tweets. The first set of tweets was collected from the Twitter API platform between 20 February and 22 July 2016. All tweets posted during this period which include one of the keywords *Brexit, EU, referendum, voteleave*, or *votestay* were collected in real time. In total, 55,551,810 tweets were collected. The data was initially used for a larger project investigating the influence of Twitter on the Brexit vote. As the focus of this study is on identity politics with a wider reach, we filtered the large dataset for only those tweets that we deemed influential for the current study.

There are three main ways in which influence on Twitter can be measured: (1) by the number of followers indicating the size of a user's audience, (2) by the number of mentions of the user indicating her or his ability to engage in conversation with other users, and (3) by the number of retweets indicating the ability of the user to write content of interest that is widely disseminated (Grčar et al. 2017). Since we were not interested in the influence of a particular politician but in the creation of networked political affiliations and identities by the Leave and Remain campaign, we decided to use the number of retweets as a marker of influence. Beyond information diffusion to large audiences, retweeting is also an important 'conversational tool' and a means of participation and engagement with others (boyd et al. 2010). What is retweeted can be assumed to be an ongoing relevant topic of conversation with which audiences engage. Only tweets that were retweeted at least 100 times were marked as *influential*, whereas the rest were removed. Each remaining tweet was then manually classified as either Remain or Leave depending on the known stance of the handle that posted it.

It is not surprising that most of the influential tweets were tweeted by politicians and media institutions such as the BBC or the *Daily Mail*, with the majority coming from politicians' accounts. This is already an interesting result, indicating that views of single people *from the street* are less likely to be widely retweeted on Twitter in the context of political campaigns.

The set of influential tweets retrieved from the big data pool was small and dominated by singular voices (e.g. Boris Johnson, Louise Mensch, Chuka Umunna), which we felt did not reflect the whole campaign. Therefore, we decided to complement the data set by collecting tweets from MPs on Twitter, by utilising a historical archive of all tweets posted by past and present Members of Parliament. This constituted the second data set.

Tweets created between 20 February and 22 July 2016 containing the same keywords as above (*Brexit, EU, referendum, voteleave, votestay*) were downloaded from mpsontwitter.co.uk. Tweets which had been retweeted 100 times were included in the data set. Any duplicates were removed. As above, each tweet was then classified as either Remain or Leave depending on the known stance of the MP during the timeframe. Tweets from four MPs who had either remained neutral during the campaign or had changed their stance on the referendum within our timeframe (as in the case of Sarah Wollaston, @sarahwollaston) were discarded. Using regular expressions and Notepad++, both sets of tweets were subsequently cleaned by removing URL links and converting unwanted html entities (e.g. *&amp*) to their applicable characters (e.g. *&*). Only the actual tweet texts were analysed including metadata in hashtags (#) and mentions (@) that the tweets contained.

We began our analysis by retrieving all instances of personal pronouns from the two corpora, Leave and Remain. When identifying candidates for the retrieval, we followed the analytical framework proposed by Tyrkkö (2016). Similar to his study, we were interested in personal references as markers of identity and therefore included personal pronouns proper (in both subject and object position) as well as possessive pronouns. Reflexive pronouns were not considered because they tend to be very low frequency items. The final set of pronouns included twenty-one items.

Both corpora were uploaded onto the corpus linguistic software programme Sketch Engine (Kilgarriff et al. 2004), which was used to retrieve the selected pronouns. Instances in which double meanings were possible were subsequently checked manually to resolve ambiguities. For example, all instances of *us* in which *us* was used as the abbreviation of the United States were removed. Similarly, occurrences of non-personal use of *they* and *them* were also excluded by manually checking the tweets in which the pronouns occurred. To identify general trends in the use of personal pronouns, the twenty-one pronouns were subsequently grouped into four main semantic categories in accordance with their primary referent – an approach which we adopted from Tyrkkö (2016). These include the following categories: (1) singular self-references (*I, my, me, mine*), (2) audience-references (*you, your, yours*), (3) plural self-references (*we, us, our, ours*), and (4) other-references (*he, his, him, she, her, hers, they, them, their, theirs*). Raw frequencies (RF) were retrieved for all twenty-one pronouns and combined in the relevant semantic categories. All results were normalised per 1,000 words using the following formula: the number of instances of a personal pronoun in a given corpus divided by the size of that corpus (in words) and multiplied by 1,000. To gauge whether the difference in usage of the pronouns in the two corpora was significant, we used the log-likelihood (LL) test available through the log-likelihood wizard created at the University of Lancaster.[2] The LL relies on the normalisation procedure based on total corpus size and hence it is well suited to deal with large quantities of non-parametric data. It is a good measure for testing the significance of differences between token counts of specific linguistic features across corpora (Dunning 1993). When reporting results, the following standard notation has been used to indicate the level of statistical significance: NS not significant; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$ and **** $p < 0.0001$.[3]

As discussed in the previous section, some personal pronouns in English can have a wide referential range and the kind of identities that they refer to can only be discerned from investigating in detail the context in which they are used. For this reason, the general quantitative insights were expanded by a qualitative analysis of the influential tweets. Because the first person plural *we* emerged as a significant distinguishing feature of influential tweets produced by Leavers and Remainers, the qualitative analysis focused on the use of this pronoun.

In order to understand what kind of identities *we* encompassed in the context of the Leave and Remain campaign, all influential tweets with the pronoun were analysed to identify and code the meaning(s) and identities that it evoked. Social actors

2. http://ucrel.lancs.ac.uk/llwizard.html (accessed 15 April 2019).

3. P stands for p-value, also referred to as calculated probability. Conventionally, a result is considered significant if it reaches $p < 0.05$ (less than 1 in 20 chance of being wrong), which is usually indicated with *. If the p-value is less than 0.01 ($p < 0.01$), this is indicated with two stars **, pointing to very significant results; *** are used for highly significant results, when the p value is less than 0.001, or **** for $p < 0.0001$.

are described and identified not just through the use of nouns and modifiers but also through the actions that they perform or are subject to (Van Leeuwen 1995), as well as the objects or effects of the actions. For example, in the sentence "He beat the hell out of Tom" it is the action and the modifier of the action that construct *he* as a violent person who does harm to others. Hence, whole tweets including hashtags and mentions were scrutinised for cues that evoke *we*-identities and their qualities.

The categorisation of the instances of *we* in the two corpora proceeded by applying the categories of meanings of *we* identified by Mühlhäusler and Harré (1990). No instances of exclusive *we* were identified; only inclusive *we* was used in the data albeit many different groups and memberships were evoked, sometimes in one tweet. For example, the following tweet "After we #VoteLeave, the British people can decide how to spend the £350m we send to the EU each week" includes two instances of *we*; the first one is clearly a reference to voters in support of Leave (signalled by the use of the hashtag), whereas the second *we* extends this reference to all British people.

The next section presents first the general trends regarding the use of personal pronouns in the corpora of influential Leave and Remain tweets identified quantitatively. Subsequently, insights obtained from the qualitative analysis of the constructions of *we*-identities are discussed.

4. Results and discussion

Table 1 presents the raw (RF) and normalised frequencies (NF, per 1,000 words) of the 21 personal pronouns retrieved from the Leave and Remain corpora.

Generally speaking, we find more personal pronouns being used in the influential tweets produced by Leavers with approximately 41 pronouns per 1,000 tokens as opposed to Remainers, who used 35 pronouns in every 1,000 tokens. The difference is significant and could potentially indicate a more involved stance (Biber and Conrad 2009) of Leavers' behaviour on Twitter. When we look at the use of specific pronouns, further conspicuous tendencies can be observed. For example, both campaigns made frequent references to men as indicated by greater use of *he, him* and *his* as compared to the female references *she, her* and *hers*, which were almost absent. This suggests that regarding the use of pronominal references, male voices and perspectives dominated the influential tweets in the run-up to the EU referendum, highlighting a gendered dimension of the campaign. This is perhaps not surprising given that in the UK politics is still dominated by men with the female participation being around 30% (one third of all MPs).[4] The *we* pronoun

4. http://www.bbc.co.uk/news/election-2017-40192060 (accessed 15 April 2019).

Table 1. Raw and normalised frequencies of personal pronouns in influential Leave and Remain tweets

Pronoun	RF Leave	NF Leave	RF Remain	NF Remain	LL	P
I	131	4.3	156	4.4	0.01	NS
me	59	1.9	31	0.9	14.04	***
my	46	1.5	122	3.4	24.35	****
mine	1	0.0	1	0.0	0.01	n/a
you	87	2.9	97	2.7	0.14	NS
your	45	1.5	41	1.1	1.42	NS
yours	0	0.0	0	0.0	0.00	n/a
he	77	2.5	49	1.4	11.70	***
him	14	0.5	4	0.1	7.64	**
his	59	1.9	29	0.8	15.93	****
she	16	0.5	9	0.3	3.30	NS
her	4	0.1	11	0.3	2.35	NS
hers	0	0.0	0	0.0	0.00	n/a
we	325	10.7	288	8.1	12.39	***
us	66	2.2	98	2.7	2.14	NS
our	161	5.3	159	4.5	2.48	NS
ours	1	0.0	2	0.1	0.20	NS
they	79	2.6	68	1.9	3.61	NS
them	23	0.8	39	1.1	1.97	NS
their	40	1.3	44	1.2	0.10	NS
theirs	1	0.0	0	0.0	1.56	n/a
TOTAL	**1235**	**40.8**	**1248**	**35.0**	**14.6**	***

was used to a significantly different extent by Leavers and Remainers; it occupied 10.7 per 1,000 tokens in influential Leave tweets in comparison to only 8.1 per 1,000 tokens of Remain's. This result suggests that Leavers relied more on creating a sense of inclusiveness since all instances of *we* in the corpora were examples of inclusive *we* (see below). Conversely, Remainers tended to refer more often to themselves through *my*, which stands out as a frequent feature of the influential tweets produced by the Remain camp. This suggests that influential voices supporting the Remain campaign were more self-promotional, since *my* as a possessive refers to entities that belong to the speaker (*my* article, *my* blog). However, when the twenty-one personal pronouns were grouped into semantic categories, the difference in the use of singular self-references was less striking. As indicated in Table 2, when the frequencies of all singular self-references were aggregated, the difference was not significant. What Table 2 supports and Figure 1 visualises, though, is a greater use of plural self-references and other-references in the influential Leave tweets suggesting that the Leave campaign was more often speaking on behalf of or for wider communities (Fetzer 2014).

Table 2. Referential categories of personal pronouns in influential Leave and Remain tweets

Pronoun	RF Leave	NF Leave	RF Remain	NF Remain	LL	P
Singular Self-Reference	237	7.8	310	8.7	1.48	NS
Plural Self-Reference	553	18.2	547	15.3	8.37	**
Audience	132	4.4	138	3.9	0.96	NS
Other	313	10.3	253	7.1	19.98	****

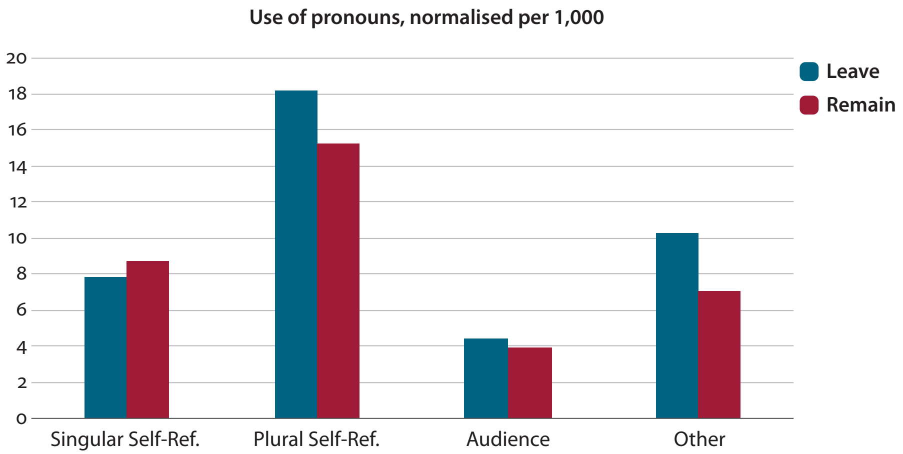

Figure 1. Referential categories of personal pronouns in influential Leave and Remain tweets, normalised per 1,000 words

In general terms, the findings fit the tendencies regarding the use of personal pronouns in political speeches observed by Tyrkkö (2016); a greater use of plural self-references and other-references and a relatively smaller proportion of audience- and singular self-references confirm the shift in political style from a more person-centred to a stronger group orientation. Since the use of the first person plural *we* emerged as a significant distinguishing feature of the influential tweets produced by Leavers and Remainers, we now turn our attention to the use of the pronoun in context.

Identifying the kind of identities evoked by *we* in the two corpora revealed some striking differences in the ways in which this pronoun was employed by the two camps. The Leavers did not only use the pronoun more frequently, they used it in a much more versatile manner. Rarely the kind of identities that were signalled by *we* involved anaphora or cataphora referring to specific names of groups (e.g. *citizens of this country*); rather, the identities were constructed through the content of the messages including devices such as hashtags or @handles that all worked together to discursively evoke a particular *we*-identity. What is more, the messages together with hashtags and @handles implied certain evaluative and affective characteristics with which the *we*-identities were then imbued.

In the corpus of the influential Leave tweets, seven clear Leavers' identities could be identified. These are: disadvantaged people of Britain, members of the EU, opponents of established political elites, World War Two heroes, advocates of democracy, global citizens and supporters of British families, farmers and consumers. These identities entail diverse memberships and can potentially speak to large numbers of audiences, uniting those who might not have shared the same political views before the campaign. We begin by analysing tweets in which *we*-identities are constructed on the basis of a negative self-presentation.

An image of disadvantaged British people dominates the self-presentations. Examples (1) to (5) are indicative illustrations of this construction in our corpus.

(1) Amount we send to Brussels each week: £350m. Cost to build a new, fully staffed NHS hospital: £350m. It's time to #VoteLeave

(2) The @EU_Commission bills the UK £350m per week – after 40 years. Wonder what level it will reach if we suffer another 40 years? #VoteLeave

(3) We do need more money in our public services… imagine what we could do with an extra £350m a week… #TakeControl #VoteLeave

(4) Another great interview @patel4witham @itvnews! £350m recognition of money we don't control! #VoteLeave #TakeControl

(5) If we #VoteLeave on 23 June we can #TakeControl of the £350m we send to the EU every week.

In these tweets, *we* has a dual indexicality; it signals the identity of Leave voters (*If we #VoteLeave*) but extends it to all British people who are portrayed as being at a disadvantage. The image of the disadvantaged British is not created through concrete naming strategies but through foregrounding causal links between policies of the EU and their *bad* effects on people living in Britain. In Examples (1) to (5), the amount of £350m is highlighted, which, according to the tweeters, Britain contributes to the EU budget per week. This claim was put forward by the right-wing populist UK Independence Party (UKIP) party and was subsequently discredited as totally inaccurate. Nevertheless, the number played an important part in the rhetoric of Leavers and was quickly taken up in public discourse. In our corpus, 37 influential Leave tweets make reference to this amount and present it as a fact. The *we*-identity foregrounded in these tweets is that of people from whom a significant amount of money is being taken away (topos of loss) and as a result they are disadvantaged because they lose out on vital services like the National Health Service (NHS). This makes them weak and lacking control (*we don't control, we suffer*). The only possible way to stop this *suffering* and to turn the disadvantage into an advantage is to take back control and exit the EU. This is emphasised by the use of the two hashtags #VoteLeave[5] and #TakeControl, which almost always occur at the end of the tweets acting as a kind of coda.

Coda is an important tool of narrativity which creates a bridge between the story world and the present moment and mostly conveys evaluative and moral messages (De Fina and Georgakopoulou 2012). The tweets can be indeed seen as examples of stories but not the kind of fully-fledged narratives studied by Labov and Waletzky (1967) but as instances of digital narrative activities which Georgakopoulou (2007) calls "small stories". Similar to other examples of small stories explored in social media, for example, Facebook updates (Page 2012; Georgakopoulou 2014), the tweets analysed in this study demonstrate narrativity through explicit references to time, place and events and a proposal how the events should be understood and acted upon. The latter is mostly evoked in the hashtags at the end acting as codas. In this sense, #VoteLeave and #TakeControl are not only the solutions to the outlined problems but also moral imperatives and commands for action as signalled through the use of the imperative mood. The underlying image of disadvantaged people is changed to those who can have agency and be in control again (*we can #TakeControl*) by voting to leave the EU. The hashtags also act as a narrative glue in the story of Leave which creates a sense of collectivism and ambient bonds between those who are in favour of leaving the EU.

Not surprisingly, another cause of disadvantages is the EU policy of free movement, which in the view of those in support of Leave contributes to uncontrolled immigration, as shown in Examples (6) to (10).

(6) @MarkFoxNews Free movement rules while we are in EU means permanent, uncontrolled immigration from the EU. #VoteLeave

(7) If we stay in the EU, with free movement of millions of immigrants, our quality of life in Britain will deteriorate.

5. Vote Leave was the name of the official campaign in favour of leaving the EU involving members of parliament from the Conservative and Labour Party as well as UKIP. Many members are prominent political figures including Boris Johnson, Michael Gove and Gisela Stuart.

(8) Cameron continues to use #ProjectFear to cover up the fact that as a member of EU we cannot control immigration #VoteLeave

(9) The EU's free movement rule, means we pay EU immigrants £860 million in benefits. If we remain, this will rocket, along with immigration #VoteLeave

(10) German Islamist terrorist has free movement to the UK. We are NOT safer in EU. Controlling our borders makes us safer

Here *we* refers mostly to Britain as a member of the EU (*we are in EU, a member of EU, we stay in the EU*), which due to this membership cannot control immigration and hence is again portrayed as being disadvantaged. This is directly emphasised in Examples (7), (9) and (10), in which immigration is linked to deterioration of quality of life (*our quality of life*), terrorism and seen as a financial burden (*we pay*). As Example (10) shows, the EU is portrayed as not a safe place because it allows free movement of terrorists and hence puts Britain in danger. Such portrayals of immigration are archetypal for right-wing propaganda and have been documented in a variety of contexts (Wodak et al. 1999; Baker et al. 2008). As in the previous examples, most of the tweets finish with #VoteLeave, implying that leaving the EU is the solution to the specified problems and the moral imperative.

One of the facets of the *we*-identity that emerged in the influential Leave tweets is that of dissatisfaction with established political elites, specifically with politicians who were in favour of Remain such as David Cameron and George Osborne (Examples (11)–(13)).

(11) If we vote to leave we will be needing a new Prime Minister and Chancellor so George Osborne will not have a say in any post-Brexit budget

(12) George Osborne's announcement re an 'emergency budget' is simply a threat to punish voters if we #VoteLeave because it would not pass

(13) #Project Fear becomes Project Threat as Cameron and Osborne threaten to lay waste to our economy when we #Brexit

Examples (11) to (13) project the outcome of the referendum in favour of the Leave campaign (*if we #VoteLeave*), at times with a high degree of certainty (*when we #Brexit*). They also project a future without two political figures, the Prime Minister and the Chancellor, who both are constructed as posing a threat to the British economy. This argument may have appealed to those who might not essentially be against the EU but wished for a change in the government.

Although the image of the disadvantaged *we* is prominent in the corpus, more often than not the influential Leave tweets project a positive image of British people and the country. The tweets in Examples (14) to (16) are indicative of such positive identity constructions:

(14) Amazing point by Andrea, we exclude some of the best nurses in the world from working in the UK because of the EU #VoteLeave

(15) Who do we want to be in control? People we elect or EU bureaucrats?

(16) We are the fifth biggest economy in the world – we will prosper outside the EU #VoteLeave

Here the *we*-identity of Leavers is based on the image of open citizens who strive for democracy and global trade but are constrained by the EU. In order to break free from these constraints, it is necessary to vote leave. The Leavers also self-present as active agents of change (*we can, we elect*) and supporters of British families, farmers and consumers (see Examples (17)–(19)). A brave new future is projected for these groups after Brexit.

(17) Michael Gove explains how after we #VoteLeave we can help family budgets by cutting VAT on energy bills

(18) After Brexit we can pay our farmers more direct to promote what is best for British agriculture, British consumers; animal welfare

(19) Gove: After we #VoteLeave we can #TakeControl and spend the £350m we send to the EU every week on our priorities

Finally, the influential Leave tweets cite voices of those who fought in the Second World War and, in doing so, link the Leave campaign with the heroic past and fight for freedom (see Example (20)). Here the *we*-identity of Leavers acquires yet another moral dimension; voting to leave the EU is a moral gratitude for the sacrifices of those who fought in WWII.

(20) WWII Heroes: 'don't give away what we fought for' urging voters to back Brexit #VoteLeave

Turning to the corpus of influential Remain tweets, we identified among them a smaller range of Remainers' identities including: the British people, Labour party, the opponents of the Conservative party and economists. Most of the instances of *we* refer to Remainers but in the context of leaving since the pronoun occurs mostly in the conditional construction *if we leave* (42 occurrences). This projects the possibility of leaving the EU (see Examples (21)–(24)), which is rather striking in the context of the Remain campaign.

(21) IFS said there would be a £20–40b budget black hole if we leave by 2019/20 – meaning sharp cuts and tax rises #StrongerIn

(22) Just spoke to UK CEO of a major US bank who says 2,000 jobs in his firm will go to mainland Europe if we leave. The cost of BREXIT is real

(23) Former heads of Tesco/Sainsbury's/M&S/Asda/Waitrose/Morrisons and B&Q warn that 'prices will rise' if we leave the EU #Strongerin #EUref

(24) Excellent by @ChrisLeslieMP on the real economic risks facing us if we leave EU #Remain #StrongerIn

As can be seen in Examples (21) to (24), the projection of leaving serves as a warning of dire consequences to which Brexit might lead, especially in financial terms. It is striking that almost all the instances of the construction *if we leave* are followed or preceded by projections of economic risks including job losses, cuts and rising prices, at times emphasised by bringing in the voices of economists, as in Example (25).

(25) IFS: Leavers may not like economists, but we are right about dangers of BREXIT

Similar to #VoteLeave and #TakeControl, #StrongerIn acts as a coda but it is less of an imperative to vote Remain and rather a form of evaluation of the current status. Whereas the influential Leave tweets project the current state as negative, blame the EU for it and then hail the exit from the EU as the key to a bright new future, the influential Remain tweets present the future after Brexit as something to be afraid of, while the benefits of being in the EU are rarely emphasised. Against this background it is perhaps not surprising that the Remain campaign was branded *Project Fear* by the Leavers and this theme features prominently in the corpus of influential Leave tweets in the form of the hashtag #projectfear (206 occurrences).

Another *we*-identity identified in the influential Remain tweets is that of supporters of the Labour party and opponents of the Conservative wing. As Examples (26) to (29) show, this is signalled through the use of the hashtag #LabourInForBritain or the description of Brexit as a Tory Brexit:

(26) To avoid a Tory Brexit we need more Labour voices to save this campaign over the next week and a half #remain

(27) Today we face one of the most important decisions. To protect jobs, growth and the economy we must vote to remain in EU #LabourInForBritain

(28) Yeah, because all these Brexit right wing Tories/kippers really want to improve workers rights if we leave the EU

(29) The Labour Party wants to stay in the EU so we can create a real social Europe #LabourInForBritain

Those who speak in these influential Remain tweets align themselves with the Labour party and self-present as supporters of workers' rights and of people who are in poverty. This is a potentially difficult alignment given that a number of central political figures within the Labour party were in favour of Brexit as were many Labour voters. The traditional divide between left and right did not cleanly apply to the Brexit vote. In contrast, the Leavers did not create allegiances with the Conservative party in their influential tweets and instead targeted political figures on both sides who were in favour of remaining in the EU.

5. Discussion

The quantitative analysis of personal pronouns in the corpora of influential Leave and Remain tweets shows a more group-centred style as indicated by the prominence of plural self-references compared to the other types. This suggests that, in this context, creating affiliations and allegiances was generally more important than self-promotion and publicising personal political goals. Examining the differences between the two camps, it appears that a group-centred style was preferred by both camps but emerged as more prominent in the influential Leave tweets. Not only did the Leavers use *we* more, they also did so in a much more versatile manner mobilizing diverse memberships.

The qualitative analysis shows that the influential Leave tweets exploited the wide referential range of *we*. While *we* was used mostly to refer to Leave voters, supporters of Leave were speaking on behalf of and for diverse *memberships of discontent*. Not only the British were evoked within this large group; also people dissatisfied with political elites or the level of public services, people who feel disadvantaged, British families, farmers were all included in the campaign messages. Interestingly, allegiances to the established political parties were avoided and instead political figures who were in favour of Remain singled out and criticised. Although the *we*-identities foregrounded in the Leave tweets were often portrayed as being disadvantaged, which is typical for populist rhetoric (Kazin 1995; Van Kessel and Castelein 2016), this was possibly a strategic move to give more weight to the Leave vote, since #VoteLeave was often added at the end of the tweet in form of a coda signalling that leaving the EU was the only way to turn the identified disadvantages into advantages. It also projected audiences who vote to leave as agents of change, thus turning them from passive *sufferers* to those who can do good for themselves and their country. Conversely, the influential Remain tweets were less persuasive in that a smaller range of *we*-identities could be identified. When *we* was used as a marker of group identity, rather exclusive membership was mobilised (economists) or groups that were already divided by the Brexit campaign such as the main political parties.

Also, and possibly unintentionally, the influential tweets projected the possibility of leaving the EU through the frequent use of the construction *if we leave* mostly preceded or followed by outlining a negative outcome, for example, "there would be a £20–40b budget black hole if we leave". Although the use of *if* implies a hypothetical scenario, nevertheless it introduces the possibility of leaving the EU and doomsday-like events that will affect the population at large as indicated by the inclusive *we* as in *if we leave*. Thus, not leaving the EU is legitimised through the strategy of scaremongering. On the contrary, the Leavers often presented Brexit as a done deal (*after we #VoteLeave, when we #Brexit*), which was mostly preceded or followed by positive outcomes as in "after we #VoteLeave we can help family budgets". Thus, leaving the EU was premised and legitimised through an optimistic outlook suggested not as a hypothetical scenario but as a real prospect.

Essentially what such constructions do is to create a sense of hope for the discontented people, who were included in the influential tweets. They conjure a kind of feel good experience. The plea to *#VoteLeave* is not just a mobilisation to take political action and vote to leave the EU, it is a vote for a supposedly better future. Giving prominence to hope and a better future is a rhetorical theme that has been employed by many successful politicians, including the former president of the USA Barak Obama (e.g. Clayton 2007). Thus, the rhetoric of hope is not essentially a trait of a conservative or liberal stance. It is a powerful theme that is employed across the political spectrum to mobilise and persuade voters. Its persuasive power lies in the potential to generate enthusiasm, which, as shown in experimental studies on political campaigning, can increase interest and bolster voters (Brader 2006). Hope was conspicuously absent from the influential tweets produced by the Remainers which seemed to be dominated by fearful scenarios. Fear too is a powerful emotion that has been shown to persuade voters in that it can increase their responsiveness. Yet, it can also create some dissonance (Brader 2006).

While it would be too simplistic to assume a causal link between the use of personal pronouns in influential tweets and the political success of the Brexit campaign, the way in which the Leave campaign used personal references on Twitter, specifically *we*, points to a more involved and persuasive style. Strengthened by the rhetoric of hope, this persuasive style could have appealed to, engaged, and connected publics that were not on the same political wave before the EU referendum, for example both Conservative and Labour voters.

Although this study points to salient uses of personal pronouns as rhetorical devices to mobilise political support, it is based on a small set of tweets collected just before the referendum, and generalisations beyond this sample cannot be made. Also, this study analysed in depth the use of just one pronoun, and, although *we* appears to be a significant distinguishing factor in the rhetoric of the two camps, examining other personal pronouns as well as other linguistic devices such as metaphors could offer further valuable insights into the ways in which collective identities are imagined on Twitter and how these imagined collectives are exploited for the purpose of political agitation online.

6. Outlook

The analysis of the two corpora of influential Leave and Remain tweets has shown that, similar to political communication offline (e.g. Tyrkkö 2016), political campaigning on Twitter adopts a group-centred style to mobilise voters and build support. Messages that were retweeted at least 100 times prior to the EU Referendum demonstrated more involvement as reflected in greater use of inclusive pronouns and fewer self-promotional ones. The politicians and organisations from whose accounts the tweets were originally sent demonstrated a great deal of awareness concerning the affordances and constraints of Twitter. Specifically, the constraint of 140 characters was strategically exploited by crafting simple messages centred on one key point using devices such as hashtags to reinforce a point, offer an evaluation or include an imperative to take action, and create affiliations around the two main campaigns, #VoteLeave and #StrongerIn.

In his analysis of the use of Twitter as a narrative tool for political campaigning, Sadler (2018) makes the point that Twitter does not require the development of wholly new literacies. At the same time, he argues that the site demands higher interpretative skills; readers are presented with short factual messages *en masse* that they need to mentally order, classify, link to events, question, etc. The question arises: to what extent do audiences actually engage with high level interpretation? After all, the messages are short, with the use of personal pronouns, specifically *we*, reinforcing the understanding that the messages are facts that affect us all (*we send, we pay, we don't control*). In doing so, *we* gives the messages personal appeal and legitimacy, which many might not see as in need of questioning. In this sense, the constraint of Twitter's word limit seems a *useful* affordance for political persuasion, especially the kind used in a populist style (Moffitt 2016). In his seminal analysis of populist persuasion in the history of the USA, Kazin (1996) shows that the language of populists, whether on the right or left of the political spectrum, is mostly based on simple, hortatory and repetitive messages with few complex elaborations. Twitter's affordance to reach wider audiences, most notably common people, and the character limit that does not leave much space for complex explanations or elaborations indeed makes Twitter a *good* medium for populist performances (Moffitt 2016).

References

Baker, Paul, Costas Gabrielatos, Majid KhosraviNik, Michał Krzyżanowski, Tony McEnery, and Ruth Wodak. 2008. "A Useful Methodological Synergy? Combining Critical Discourse Analysis and Corpus Linguistics to Examine Discourses of Refugees and Asylum Seekers in the UK Press." *Discourse & Society* 19 (3): 273–306. https://doi.org/10.1177/0957926508088962

Biber, Douglas, and Susan Conrad. 2009. *Register, Genre, and Style*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511814358

boyd, danah, Scott Golder, and Gilad Lotan. 2010. "Tweet, Tweet, Retweet: Conversational Aspects of Retweeting on Twitter." Paper presented in Proceedings of *43rd Hawaii International Conference on System Sciences*, 2010.

boyd, danah. 2011. "Social Network Sites as Networked Publics: Affordances, Dynamics, and Implications." In *A Networked Self: Identity, Community, and Culture on Social Network Sites*, ed. by Zizi Papacharissi, 39–58. New York, NY: Routledge.

Brader, Ted. 2006. *Campaigning for Hearts and Minds: How Emotional Appeals in Political Ads Work*. Chicago, London: The University of Chicago Press.

Charteris-Black, Jonathan. 2004. *Corpus Approaches to Critical Metaphor Analysis*. Basingstoke: Palgrave Macmillan. https://doi.org/10.1057/9780230000612

Clayton, Dewey. 2007. "The Audacity of Hope." *Journal of Black Studies* 38 (1): 51–63. https://doi.org/10.1177/0021934707305057

Coesemans, Roel, and Barbara De Cock. 2017. "Self-Reference by Politicians on Twitter: Strategies to Adapt to 140 Characters." *Journal of Pragmatics* 116: 37–50. https://doi.org/10.1016/j.pragma.2016.12.005

Cramer, Jennifer. 2010. "'Do We Really Want to Be Like Them?': Indexing Europeanness through Pronominal Use." *Discourse & Society* 21 (6): 619–637. https://doi.org/10.1177/0957926510381217

De Fina, Anna, and Alexandra Georgakopoulou. 2012. *Analyzing Narrative: Discourse and Sociolinguistic Perspectives*. Cambridge: Cambridge University Press.

De Fina, Anna. 1995. "Pronominal Choice, Identity, and Solidarity in Political Discourse." *Text* 15 (3): 379–410. https://doi.org/10.1515/text.1.1995.15.3.379

Di Fatta, Guiseppe, James J. Reade, Sylvia Jaworska, and Anupam Nanda. 2015. "Big Social Data and Political Sentiment: The Tweet Stream during the UK General Election 2015 Campaign." Paper presented in Proceedings of *2015 IEEE International Conference on Smart City/SocialCom/SustainCom (SmartCity)*, Chengdu, 293–298. https://doi.org/10.1109/SmartCity.2015.87

Dunning, Ted. 1993. "Accurate Methods for the Statistics of Surprise and Coincidence." *Computational Linguistics* 19 (1): 61–74.

Fetzer, Anita, and Peter Bull. 2008. "'Well, I Answer It by Simply Inviting You to Look at the Evidence': The Strategic Use of Pronouns in Political Interviews." *Journal of Language and Politics* 7 (2): 271–289. https://doi.org/10.1075/jlp.7.2.05fet

Fetzer, Anita. 2014. "'Judge *Us* on What *We* Do': The Strategic Use of Collective *We* in British Political Discourse." In *Constructing Collectivity: 'We' across Languages and Contexts*, ed. by Theodossia-Soula Pavlidou, 331–350. Amsterdam, Philadelphia: John Benjamins. https://doi.org/10.1075/pbns.239.21fet

Georgakopoulou, Alexandra. 2007. *Small Stories, Interaction and Identities*. Amsterdam, Philadelphia: John Benjamins. https://doi.org/10.1075/sin.8

Georgakopoulou, Alexandra. 2014. "Small Stories Transposition and Social Media: A Micro-Perspective on the 'Greek Crisis'." *Discourse & Society* 25 (4): 519–539. https://doi.org/10.1177/0957926514536963

Grčar, Miha, Darko Cherepnalkoski, Igor Mozetic, and Petra Kralj Novak. 2017. "Stance and Influence of Twitter Users Regarding the Brexit Referendum." *Computational Social Networks* 4 (6). https://doi.org/10.1186/s40649-017-0042-6

Humphreys, Lee, Phillipa Gill, Balachander Krishnamurthy, and Elizabeth Newbury. 2013. "Historicizing New Media: A Content Analysis of Twitter." *Journal of Communication* 63 (3): 413–431. https://doi.org/10.1111/jcom.12030

Jungherr, Andreas. 2014. "The Logic of Political Coverage on Twitter: Temporal Dynamics and Content." *Journal of Communication* 64 (2): 239–259. https://doi.org/10.1111/jcom.12087

Kaufmann, Eric. 2016. "It's NOT the Economy, Stupid: Brexit as a Story of Personal Values." *LSE European Politics and Policy (EUROPP)* (blog). 9 July 2016. Available online at http://eprints.lse.ac.uk/70413/ (accessed 15 April 2019).

Kazin, Michael. 1995. *The Populist Persuasion. The American History*. Ithaca, London: Cornell University Press.

Kilgarriff, Adam, Pavel Rychlý, Pavel Smrz, and David Tugwell. 2004. "The Sketch Engine." Paper presented in Proceedings of *EURALEX 200*, Lorient, France, 105–116.

Kleinke, Sonja, and Birte Bös. 2018. "Indeterminate *Us* and *Them*: The Complexities of Referentiality, Identity and Group Construction in a Public Online Discussion." In *The Discursive Construction of Identities On- and Offline*, ed. by Birte Bös, Sonja Kleinke, Sandra Mollin, and Nuria Hernández, 153–176. Amsterdam, Philadelphia: John Benjamins.

Labov, William, and Joshua Waletzky. 1967. "Narrative Analysis: Oral Versions of Personal Experience." In *Essays on the Verbal and Visual Arts*, ed. by June Helm, 12–19. Seattle: University of Washington Press.

Marwick, Alice, and danah boyd. 2010. "I Tweet Honestly, I Tweet Passionately: Twitter Users, Context Collapse, and the Imagined Audience." *New Media & Society* 13 (1): 114–133. https://doi.org/10.1177/1461444810365313

Moffitt, Benjamin. 2016. *The Global Rise of Populism: Performance, Political Style, and Representation*. Stanford: Stanford University Press. https://doi.org/10.11126/stanford/9780804796132.001.0001

Mühlhäusler, Peter, and Rom Harré. 1990. *Pronouns and People*. Cambridge, MA: Basil Blackwell.

Musolff, Andreas. 2015. "Dehumanizing Metaphors in UK Immigrant Debates in Press and Online Media." *Journal of Language Aggression and Conflict* 3 (1): 41–56. https://doi.org/10.1075/jlac.3.1.02mus

Page, Ruth. 2012. *Stories and Social Media*. London: Routledge.

Papacharissi, Zizi. 2015. *Affective Publics: Sentiment, Technology, and Politics*. New York, NY: Oxford University Press.

Pavlidou, Theodossia-Soula. 2014. "Constructing Collectivity with 'We'. An Introduction." In *Constructing Collectivity: 'We' across Languages and Contexts*, ed. by Theodossia-Soula Pavlidou, 1–19. Amsterdam, Philadelphia: John Benjamins. https://doi.org/10.1075/pbns.239.03pav

Quirk, Randolph, Geoffrey Leech, and Sidney Greenbaum. 1985. *A Comprehensive Grammar of the English Language*. London: Longman.

Sadler, Neil. 2018. “Narrative and Interpretation on Twitter: Reading Tweets by Telling Stories.” *New Media & Society* 20 (9): 3266–3282. https://doi.org/10.1177/1461444817745018

Tumasjan, Andranik, Timm O. Sprenger, Phillipp G. Sander, and Isabell M. Welpe. 2010. “Election Forecasts with Twitter: How 140 Characters Reflect the Political Landscape.” *Social Science Computer Review* 29 (4): 402–418. https://doi.org/10.1177/0894439310386557

Tyrkkö, Jukka. 2016. “Looking for Rhetorical Thresholds: Pronoun Frequencies in Political Speeches.” In *The Pragmatics and Stylistics of Identity Construction and Characterisation* (Studies in Variation, Contacts and Change in English 17), ed. by Minna Nevala, Ursula Lutzky, Gabriella Mazzon, and Carla Suhr. Helsinki: VARIENG. Available online at http://www.helsinki.fi/varieng/series/volumes/17/tyrkko/ (accessed 15 April 2019).

Van Kessel, Stijn, and Remco Castelein. 2016. “Shifting the Blame. Populist Politicians’ Use of Twitter as a Tool of Opposition.” *Journal of Contemporary European Research* 12 (2): 594–614.

Van Leeuwen, Theo. 1995. “Representing Social Action.” *Discourse & Society* 6 (1): 81–106. https://doi.org/10.1177/0957926595006001005

Wilson, John. 1990. *Politically Speaking: The Pragmatic Analysis of Political Language*. Oxford: Blackwell.

Wodak, Ruth, Rudolf de Cillia, Martin Reisigl, and Karin Liebhart. 1999. *Discursive Construction of National Identity*. Edinburgh: Edinburgh University Press.

Zappavigna, Michele. 2014. “Enacting Identity in Microblogging through Ambient Affiliation.” *Discourse & Communication* 8 (2): 209–228. https://doi.org/10.1177/1750481313510816

PART III

Public discourses around the world

To be or not to be … a patient

Identity construction of healthcare professionals and patients in public online diabetes-related interaction

Barbara De Cock
UCLouvain

This study offers an analysis of medical and emotional identity features put forward by participants asking questions and providing answers in three types of public online diabetes-related interaction in Spanish (frequently asked questions, chat sessions and fora). The results show that many writers present themselves not as diabetics but as relatives of a diabetic, or else refrain from specifying their relation to diabetes. Furthermore, as no shared patient-professional identity is constructed, the healthcare relationship profiled is a rather traditional one. The differing natures of the platforms and genres are reflected in the identities constructed. In the FAQs a more impersonal or collective identity is apparent, whereas that constructed in the forum tends to be more personal, with more emotional aspects; the chat data occupy an intermediate position.

Keywords: identity, reference, online discourse, healthcare interaction, pragmatics, deixis, Spanish

1. Context and research aims

In this chapter, I present an analysis of identity construction in public online interaction on the topic of diabetes, with particular attention being paid to the identity construction of patients and healthcare professionals in this type of interaction. I also study whether this identity construction reflects a traditional patient-healthcare professional relationship or rather a relationship of co-responsabilisation (Salvador and Macián 2009). This chapter thus focuses on a very specific context in which identity and reference play a crucial role in constructing a professional relationship and in attributing responsibility.

https://doi.org/10.1075/pbns.306.08dec

In this section, I will briefly discuss the nature of diabetes and I will introduce the emergence of e-health and the co-responsabilisation of patients (1.1), leading into a discussion of the study's research aims (1.2). In Section 2, I will then present the methodology and data before, in Section 3, examining the representation of patients (3.1) and their relatives (3.2), and of healthcare professionals (3.3). Finally, I will offer some concluding remarks about identity construction in online interactions on diabetes (Section 4).

1.1 Context

Diabetes can take three forms. Type 1 diabetes typically appears at a young age. Patients play an active part in their treatment and in decision-making about it by, for example, self-controlling glucose levels in the blood and self-administering insulin injections. Type 2 diabetes is typically developed later in life and may be treated with pills or injections. Both these types are chronic health conditions which will affect patients for the rest of their lives. As a result, the condition may become part of their identity. Gestational diabetes, which occurs during pregnancy, does not necessarily become chronic but may do so. It has been shown in sociological research on chronic illness that adapting one's own biography (including one's conception of self) is one of the crucial factors in accommodating to such a condition (in addition to dealing with medical and emotional issues) (Corbin and Strauss 1988: 50). Thus, creating one's own biography and constructing a personal identity form part of the management of chronic illness. This makes the study of self-representation particularly relevant in the case of diabetes.

Additionally, in view of the significant changes the relationship between patient and healthcare professional seems to be undergoing, the study of other-representation in medical interaction is becoming increasingly important. Indeed, a generally increased co-responsibilisation (Salvador and Macián 2009) of the patient can be observed in many medical interactions, as well as a rising demand for patient participation (Keel and Schoeb 2017). Similarly, it has been shown that treatment can be constructed as a "joint purpose" (Antaki and Crompton 2015). This is particularly relevant since, as well as the professional's communication with the patient (Haskard Zolnierek and DiMatteo 2009), the active involvement of the patient in decisions (Drew et al. 2001; Rollnick et al. 2008) plays a role in treatment adherence. The very active role that diabetics assume in their treatment on a daily basis underlines this and contributes to the significance of their co-responsibility. In view of its link with treatment adherence, the topic of self- and other-representation in the medical context is important not only from a discursive or linguistic point of view, but also from a medical or psychological one.

The data studied in this paper are taken from public online interactions, given the increasing importance of online communication for self-expression and identity construction, and for communication concerning healthcare issues. However, the role that the internet can legitimately play in healthcare is currently the subject of discussion. Some warn of the negative effects of 'Doctor Google'; concerns in this regard relate to the ready availability of unmonitored and not necessarily reliable information, and to self-diagnosing by laypersons. And, in fact, studies on professionals' practices to enhance therapy adherence have revealed a "low reported use of technology and other resources to support patients with medicines used in routine practice" (Clyne et al. 2016: 8). Others, however, say that digital media have advantages in terms of facilitating access to information (Kerr et al. 2006), in public education and awareness raising, and in offering assistance for self-management (Pal et al. 2018). Indeed, the use of websites and apps in support of treatment or self-management is being studied with a view to meeting the ever-increasing demand for information and to reaching non-native speaker publics (De Wilde et al. 2019). Moreover, the internet also allows patients to contact other patients and to exchange experiences and ideas (Liu et al. 2017), which has been shown to be a way of obtaining more healthcare information and of supporting one another (Loader et al. 2002).

The healthcare professionals providing information in the data analyzed in this study work for official healthcare institutions, and two of the data sets were posted directly on the websites of official healthcare agencies, showing that they have thus opted for using new forms of digital communication in order to spread information. They therefore (at least implicitly) support the tendency towards valuing online communication as a useful extra tool in healthcare communication. The online interaction studied in this paper is public in nature, yet with varying degrees of general scope and (a)temporal scope, as we will see in Section 2, which introduces the data studied.

1.2 Research aims

This research is concerned with how the self and other are represented in public digital interactions on the topic of diabetes. It investigates the identity categories, of both patients and healthcare professionals, made relevant by participants asking questions in (the pre-expansion to) their queries and those categories subsequently exploited and made consequential by participants providing answers. In other words, it will look into those identity features put forward and commented upon by participants that play a role in the development of the interaction concerned. These include not only identity features strictly relevant from a medical perspective, but also more personal or emotional information. Furthermore, I will study whether the participants contribute jointly to constructing the identity of patients and healthcare professionals as a collective identity, or rather as separate entities, which are not co-responsible for the treatment. I will also look into differences in identity construction between the three interaction types discussed, namely an FAQ section, chat data and a medical discussion forum.

2. Methodology and corpus

The focus of this paper is on the way participants formulate and answer questions in the data studied and how this contributes to the construction of the self and other. In particular, I analyze how identity is constructed throughout entire interactions, rather than in individual utterances, by studying how the participants asking and answering questions present themselves – and, if they do so, others – indexically in the context of the interaction. In doing so, I draw on insights from pragmatics, conversation analysis, discourse analysis and discursive psychology, as well as on sociological research into healthcare interaction.

Identity being a multi-layered concept (De Fina 2010), the focus of this analysis will be on those identity features that are ascribed to the interaction's participants by themselves and by others. Thus, I will study those features that, in conversation analytic terms, are made relevant and consequential in the interaction. They are made relevant if they are somehow under discussion or taken up in the ongoing interaction, and they are consequential if they have "some visible effect on how the interaction pans out" (Antaki and Widdicombe 1998: 5), for instance, by being commented upon explicitly in the unfolding interaction. Of course, this also implies that there are some features which participants decided not to make relevant, but which may be considered interesting for the interaction at hand. Not all participants, for instance, specify their personal relationship to diabetes, although this information may be useful in the interaction. The absence of such features will also be discussed. Indeed, as has been shown in studies on self-disclosure in blogs (see Salonen 2018), the information disclosed contributes to constructing an identity. In the data analyzed in this chapter, part of this disclosure is of course motivated by the wish to receive an adequate answer to a specific question, for which the participants may need to disclose identity features. Given the type of interaction under study, the relational pair (Sacks 1995: 327) patient-healthcare professional seems to be the most evident one by which to categorize the participants, but we will see that it is by no means the only one appearing in the data.

As regards self-representation, I will look into self-definition and self-description through constructions with 1st person deictic forms. For other-representation, I will take into account interlocutor references but also address forms and ratification of an interaction participant by means of 3rd person forms. Spanish being a pro-drop language, both person reference expressed in the verb morphology and in explicit subject pronouns will be taken into account (yet, as we will see, the latter are extremely rare in these data).

These forms allow access to the construction of the contextual identity, as understood by Kerbrat-Orecchioni (2005: 157), namely the elements of the global identity that are relevant in a given context of interaction. The mobilization of these features may be realized through linguistic and non-linguistic behavior. In the data used, the communication platform in itself implies that a focus on diabetes-related features is evident. Thus, various chat participants start talking about their treatment immediately, without explicitly specifying that it is for diabetes. Given that the chat is presented as exclusively dedicated to dealing with questions concerning diabetes, this contextual identity is in part predefined by the chat session.

The data used for this study, summarized in Table 1, are drawn from a public question-and-answer page (FAQs), a public chat session organized for World Diabetes day (subsequently published online as a pdf), and a public forum on medical topics. In the three genres, healthcare professionals answer questions from the general public (both patients and non-patients). In all cases, the contributors and readers were making a conscious effort to obtain information (cf. Antaki et al. 2006 concerning forum data). Initiating a question is thus always a "permitted" new action (Antaki et al. 2006: 118).

All the interactions studied can be considered public. In the case of the chat and forum data, participants knew at the time of posting that their question (and answers to it) would be publicly available, which leads me to consider these interactions as public (cf. Spilioti 2017: 194). Admittedly, those who asked a question that appears in the question-and-answer data may not have known that their question would end up on the FAQ page. However, the FAQs do not feature any information that allows us to trace them back to the original writer (name, pseudonym, place…). These data could therefore be made public while respecting the privacy of writers who may not have wished to make their questions publicly available. As regards the professionals figuring in the three data sets, all work for officially recognized organizations and/or hospitals. Any possible commercial bias inherent in interaction with private healthcare professionals is thus eliminated.

As I wanted to include comparable data (European Spanish, public online interaction involving healthcare professionals linked to official institutions), the data set is relatively limited. Consequently, the aim of this paper is to show different mechanisms at work and to illustrate varieties of identity construction rather than to offer quantitative comparisons of these phenomena.

Table 1. Summary of the data used in this study

Genre	Q&A (FAQs)	Public chat for World Diabetes Day	Medical forum
Source	Diabetes Madrid	Facebook page of the Valencian autonomous region healthcare agency (*Cuídate CV*)	Onmeda (diabetes thread and diabetes-related posts in other threads)
Healthcare professionals	Several Madrid-based doctors	Endocrinologist	Doctors and nutritionist
Words	2,193	5,148	8,937
Questions	15	13	17

The data sets differ mainly in two respects. The first is synchronicity, which is high for the chat session and can be low for the other genres. The second relates to the presence of hearers and bystanders in Goffman's terms. In the FAQ session, the question writer is presented as de-individualized, allowing anyone with the same concern to identify with that position. In the public chat session, the participants who ask and answer questions know that there is a group of unaddressed recipients, namely others logged in to the chat session at the same time, who are following the interaction and who may join it actively by formulating their own questions. While there is also a potentially large group of bystanders in the forum, these are not necessarily simultaneously present and do not necessarily make their presence known by posting or answering questions. They can choose simply to read the threads without participating. The forum is moderated, as is evidenced by the fact that messages containing publicity for medicines are explicitly condemned and their originators warned that repeatedly posting such messages will lead to them being blocked.

Since the questions and answers on the FAQ page are selected for their assumed representativeness, it is possible that the samples forming part of my data were edited prior to publication. However, the rather varied formulation and structure of the questions in this section (e.g. some are written in the 1st person, others not) suggest that they are at least based on real questions, though their brevity (as compared to the questions in the two other data sets) suggests that some editing has taken place. Even if edited, they still represent the way in which patient and healthcare professional identities are typically constructed according to the authors/editors of the FAQs.

While we might consider time constraints to have played a role in the chat session (as opposed to the two other sets), the answers given there are fairly elaborate and written in standard language, suggesting that time pressure did not have a major impact, at least as far as the extent and style of answers are concerned. The question-and-answer page and the chat session explicitly state that answers will be given by healthcare professionals, whereas the forum allows for any registered participant to contribute. Nevertheless, replies there are typically given by a healthcare professional, who is the most ratified participant (Goffman 1979) and often even explicitly selected by the question writer through the choice of address forms or by a requirement for specialist attention (see Section 3). Indeed, only two threads result in an exchange that goes beyond the initial question and answer to include responses from non-professionals (patients sharing experiences). Thus, while the forum would allow for more prolonged interactions, in practice it works in a fairly similar way to the chat session and to FAQs. The forum data included in the current study therefore differ from diabetes fora where the participants are mainly or exclusively patients and which have been shown to fulfill a supportive function (Loader et al. 2002), being used also to seek healthcare information (Liu et al. 2017).

The data were anonymized according to the guidelines described by Antaki (2002), which involves the use of pseudonyms for the names of participants asking questions and providing answers. They may contain non-normative language use, which has not been edited for the purposes of this study. Translations aim to convey the register of the original message.

3. Analysis

In this section, I first discuss patient representation (3.1), as realized both through self-representation and through other-representation by professionals or by non-patients commenting on patients. I will consider whether the patients are presented as individuals or as part of a group, the identity features made relevant, the way in which they describe their identity in relation to diabetes, and the degree to which they are represented as (co)responsible for their treatment. I will then proceed to discuss how patients' relatives (3.2) and healthcare professionals (3.3) present themselves and how they are represented by the other participants.

3.1 Patient identity

Participants posing questions seem to opt for presenting their queries either as resulting from a personal concern (as a diabetic or a close relative of one) or as generic and abstract (using impersonal constructions or generalizing descriptions). While in the FAQs three questions (of the sample of 15) are formulated using the 1st person singular and two the 1st person plural, impersonal or generalizing constructions prevail in this data set, as illustrated by Example (1). Self-representation is much more present in the chat and forum data, where the vast majority of questions are formulated in the 1st person singular, ranging from a merely metadiscursive presence (*querría preguntar…* 'I would like to ask…') to the author explicitly representing him-/herself as a diabetic, as in Example (2).

(1) *¿Cómo se puede diagnosticar de forma precoz la diabetes?* (FAQs)
'How can diabetes be diagnosed at an early stage?'

(2) *Hola, soy un chico con diabetes de 28 años, qué actividad física me recomendarías?* (chat)
'Hi, I'm a 28-year-old guy with diabetes, what type of physical activity would you recommend to me?'

In only two cases, both among the FAQs, do questioners present themselves as representing the whole group of diabetics. In both questions, a construction specific to Spanish, Catalan and Occitan is used, namely a full subject NP immediately followed by a 1st person plural verb phrase (De Cock 2010), as in (3). This construction allows the speaker, on the one hand, to express membership of a group through the use of the 1st person plural and, on the other, to explicitly mention the group referred to (*las personas con diabetes* 'people with diabetes'). It is thus a way of creating an ingroup as opposed to an outgroup ('people who do not have diabetes') (cf. Duszak 2002: 6). Given its low frequency of less than 1.4% of 1st person plural forms in general corpora (De Cock 2014: 156), it is noteworthy that precisely this rare construction is used. Clearly, the fact that it combines group membership with a fairly detailed description of the group makes it particularly suitable for the genre of FAQs.

(3) *¿Las personas con diabetes podemos utilizar viagra?* (FAQs)
'Can people like us with diabetes use Viagra?'

Note, though, that the clear presence of the writer signified by a 1st person singular form/pronoun does not necessarily mean that the questioner identifies him/herself as a diabetic. While on the forum six out of 17 writers of initial posts present themselves with the diagnostic label of diabetes, four threads were started by people using a 1st person singular form but identifying themselves as close relatives of a diabetic, as in (4).

(4) *Una persona muy cercana a mi tiene diabetes del tipo que se tiene que aplicar inyecciones de insulina 3 veces al día […] Estoy preocupada porque su tratamiento lo lleva su médico de cabecera pero me pregunto, ¿dónde podría pedir una segunda opinión para no tener los niveles tan altos?* (chat)
'Someone very close to me has diabetes of the type where she has to have insulin injections 3 times a day […] I'm concerned because her family doctor is in charge of her treatment but I'm wondering where she could ask for a second opinion in order to not have such high levels?'

In the chat data, participants identify themselves as diabetics in five threads and as a close relative in four. The remaining four threads do not include any specification concerning a possible personal involvement. While the degree of detail included in some of these posts might lead us to think that the author does have a concrete person in mind, he/she does not make relevant any identity features linking him/her to diabetes and seems not to wish to present him/herself as (a close relative of) a diabetic. No declared close relatives feature in the FAQ data.

Since the FAQ data barely present individual patients, they do not offer any further identifying features. In the chat and forum data, however, further patient identity features are sometimes introduced by the patient or relative. These include, for instance, their name, gender (on some occasions inferable from the name), type of diabetes, age, time since diagnosis, or other health issues. In the forum data, the medical information provided may even be fairly specific, including glucose values in the blood, type of treatment, etc. Nine of 17 forum threads contain fairly specific medical information. In some cases, the writer also adds information concerning their personal attitude towards diabetes.

In the following, I will study in more detail how the relationship with diabetes is defined by the participant asking a question. The data show that both diabetics and close relatives use various constructions, ranging from *ser diabético* 'to be a diabetic', *tener diabetes* 'to have diabetes', or *con diabetes* 'with diabetes', to verbs that more clearly evoke the idea of an illness, such as *sufrir* 'suffer' or *padecer diabetes* 'suffer from diabetes'. Finally, some question authors refer to the diagnosis, stating *le diagnosticaron* 'they diagnosed him/her'. Unlike other expressions, especially *ser diabético*, which present the condition of being a diabetic as an intrinsic part of patient identity, constructions with *diagnosticar* rather suggest the idea of a label used by third parties. No clear correlation was found between the use of these two constructions and the author's status (patient or relative).

In the forum and chat data, some contributors expand on their (emotional) relationship with diabetes or their emotional state when posting their question. Thus, both diabetics and close relatives sometimes explicitly mention that they are worried, as in Example (5). Such personal or emotional information concerning illness identity is absent from the FAQ data, probably due to the fact that these were selected as prototypical questions of potential relevance for a variety of readers, and not merely for the person who asked the question.

(5) *Es la primera vez que escribo en el foro. estoy preocupada.* (forum)
'It's my first time writing in the forum. I'm worried.'

In the forum messages, which are typically longer, some members give an even more detailed account of their attitude towards diabetes itself. However, this appears always as a pre-expansion to a question that will concern more social or emotional aspects of being a diabetic, such as relationships or the possibility of moving abroad for professional reasons. Example (6) reflects the patient's struggle with her illness identity. It shows how a new event in her life (in this case the opportunity to work abroad) has led her to reconsider the impact of the chronic illness on her identity, which she had initially considered to be unaffected by her illness. This illustrates that the construction of the biography of the chronically ill person is an ongoing process, and that new events (such as this job opportunity) may bring patients to reconsider their identity with respect to their chronic disease even when they thought such biographical work had been completed (cf. Charmaz 1995: 660). In (6), moreover, this reconsideration occurs at a stage in life when development of both overall and illness identity is crucial (Oris et al. 2016).

(6) *Desde que me diagnosticaron esta enfermedad, siempre me he dicho que no me impediría llevar una vida "normal" y poder hacer cosas como el resto del mundo, pero estoy viendo que al final va a resultar así. De verdad que me gustaría muchísimo aprovechar esto, ya que únicamente tengo 19 años y algo así es difícil que vuelva a pasar por mi vida, soy joven y tengo muchas experiencias por vivir aún.* (forum)
'Since they diagnosed me with this illness, I've always said to myself that it wouldn't prevent me from leading a "normal" life and being able to do things like everybody else, but I'm coming to realize that that's how it will be in the end. I'd really, really like to take this [opportunity] up since I'm only 19 years old and something like this is unlikely to happen in my life again, I'm young and I still have lots of things to experience.'

As we will see in the analysis of the answer given in (7), the healthcare professional explicitly acknowledges that this question is not merely about medical aspects of diabetes but also about the patient's personal relationship with the condition, as he also does in some other interventions. Indeed, when we turn to the way in which patients (and their relatives) are represented in the answers, we will see that the professionals fully respect their choice whether to self-identify as a diabetic or not. In other words, they only make the patient identity relevant in their answer if it was made relevant in the question in the first place. This is logical from the perspective of medical ethics, since it respects patients' privacy.

In this regard, the forum and chat data vary. While the chat answers may refer to personal medical information (e.g. patient's age, how long he/she has been diagnosed with diabetes), they do not explicitly take up personal emotional information such as the expression of worries. The forum answers, by contrast, do address the emotional aspects of patients' and relatives' self-representation. In the case of patients, this may involve explicit discussion of aspects of illness identity made relevant by patients in the original questions. Thus, the doctor's reply (see

Example (7)) to the concerns voiced in Example (6) begins by explicitly acknowledging Cristina's worries (*entiendo tu inquietud* 'I understand your worries'). Her desire to live a 'normal' life is also taken up in the answer, which starts by pointing out the problems faced by anyone with a medical condition working in America, whether or not they are diabetic (*Todos, diabéticos o no* 'Everyone, diabetic or not'), before proceeding to more specific diabetes-related information. The professional thus underlines an identity Cristina shares with others. After extensive recommendations, the doctor closes the message by again addressing her desire to seize opportunities in life. In general, answers to forum posts such as (7) not only address the patient's medical condition but also the more emotional and personal concerns related to their condition, in acknowledgment of the identity struggle they are experiencing.

(7) *Hola Cristina,*
Entiendo tu inquietud por las incertidumbres que genera viajar a un país, como Estados Unidos, en el que ningún trabajador tiene derecho a seguro médico público antes de los 65 años. Todos, diabéticos o no, tienen que suscribir un contrato de asistencia médica con una entidad privada. […] Animo y no dejes que las dificultades que encontrarás en cualquier momento en tu vida te condicionen e impidan que hagas aquello que te ilusiona y motiva.
Un saludo.
Dr. Miquel González
'Hi Cristina,
I understand your worries about the uncertainties that travelling to a country like the US entails, where no employee has the right to public health insurance before the age of 65. Everyone, diabetic or not, has to take out health insurance with a private company. […] Think positive and don't let the difficulties that you may encounter at any time of your life define you and stop you from doing what thrills and motivates you.
Kind regards,
Dr. Miquel González'

Joint construction of responsibility or action with patients is very rare, but we do find examples such as (8). In this case, the specialist seems to oscillate between augmented inclusive uses of *we* (including the speaker, hearer and third persons) and pseudo-inclusive *we* (which actually addresses the hearer but does not include the speaker) (Cysouw 2003: 85; see also De Cock 2011, 2016). Insofar as the professional is answering a specific query about transgenerational transmission, posed by a writer whose father has diabetes 2, we can interpret the 1st person plural forms as pseudo-inclusive, i.e. they actually refer to the non-professional but are used to create empathy. Although they occur in many interactional situations (De Cock 2016), such uses have typically been associated with medical contexts (Brown and Levinson 1987: 119). The use of the 1st person plural also allows the answer to be broadened from a specific case to the more general issue of preventing type 2 diabetes (*hay algunas cosas que sí podemos hacer para intentar evitar que eso suceda* 'there are some things we can do to try to prevent that happening').

(8) *En cualquier caso esto [posibilidad de heredar diabetes tipo 2] es algo sobre lo que no podemos influir, al igual que con la edad las posibilidades son mayores. Sin embargo hay algunas cosas que sí podemos hacer para intentar evitar que eso suceda, y que si ocurre que sea más tarde y más leve: lo que podemos hacer es mantener el peso normal, sin engordar, hacer algo de ejercicio y comer equilibradamente una dieta mediterránea. […] Así que ya sabes: a cuidarte, que es lo que debemos hacer todos en realidad, para mantenernos sanos.* (chat)
'In any case, that [the chances of inheriting type 2 diabetes] is something we can't influence, just like the fact that the chances [of getting diabetes type 2] increase with age. However, there are some things we can do to try to prevent that happening, and to ensure that, if it happens, it is as late and mild as possible: we can maintain a normal weight, avoid getting fat, take some exercise and eat a balanced Mediterranean diet. […] So now you know what to do: look after yourself – which is what all of us should be doing, really, in order to stay healthy.'

In the closing utterance, the dichotomy between addressing the questioner's specific situation and giving general information is resolved by the use of more explicit person reference; the professional first uses a direct 2nd person form of address (*ya sabes: a cuidarte* 'now you know what to do: look after yourself') but then makes explicit that the recommendation holds for everybody (*que es lo que debemos hacer todos en realidad* 'which is what all of us should be doing, really'). Thus, both through the pseudo-inclusive *we* and through the explicitly generic closing statement (*todos* 'all (of us)'), the professional downplays the specific label of the questioner as a person with a more elevated risk of developing type 2 diabetes, representing both her- and himself as part of one big community which should try to live a healthy life. This then seems to be a joint construction, involving not only the professional and the questioner, but society at large. Indeed, the identities of the participants originally asking the question (relative of a diabetic) and responding to it (the professional) are both backgrounded. The professional steps out of his expert role and presents himself as part of a larger community, where not only (potential) patients but everyone should look after themselves. Interestingly, as we will see in the analysis below, such joint responsibility is often constructed when close relatives of diabetics ask questions. They are accordingly addressed as playing an active role in treatment, especially in helping to maintain a healthy lifestyle.

3.2 Relatives' identity

As mentioned above, on various occasions the participants asking questions identify themselves as relatives of a diabetic. In such cases, self-description tends to be limited to the degree of kinship (e.g. sisterhood in Example (9)). On the forum, the specialists systematically take up this relationship in their answer and acknowledge the writer as a concerned and valued interlocutor, emphasizing the importance of their support and concern for their diabetic family member, as in (10) (*tu preocupación por el bienestar de tu hermana* 'your concern for your sister's wellbeing'). In this sense, forum answers pay more attention to the emotional role of the family members, going well beyond a strictly medical answer as in the Q&A and the chat. Moreover, the healthcare professionals tend to construe the family members as involved persons with a role to play in the treatment or, at least, in supporting the patient by helping him/her to maintain a healthy lifestyle (e.g. eat healthy food, take exercise; *espero que esta información te sea útil para que […] la ayudes […]* 'I hope that this information is useful for you […] in helping her […]'). In this way, the professionals construct a relationship of co-responsibility between patients and their family members.

(9) *Buenos días,*
a mi hermana embarazada de 6 meses le han diagnosticado diabetes gestacional. ¿Qué medidas generales tiene que seguir? ¿Qué normas de alimentación y dieta tiene que tener en cuenta? (forum)
'Hello,
My sister, who is 6 months pregnant, has been diagnosed with gestational diabetes. What general measures must she take? What guidelines on nutrition and diet must she take into account?'

(10) *Buenos días, Maria Teresa,*
Muchas gracias por tu pregunta que pone de manifiesto tu preocupación por el bienestar de tu hermana. Espero ser capaz de ayudarte para que tú puedas ayudarla a ella. […] pero para ella es una suerte contar con tu interés y tu ayuda. Espero que esta información te sea útil para que, más allá de los aspectos teóricos, la ayudes compartiendo, por ejemplo, una comida o un paseo "saludables", siempre que os sea posible. (forum)

> 'Hello Maria Teresa,
> Thank you very much for your question, which shows your concern for your sister's wellbeing. I hope I'll be able to help you so you can help her. [...] but she's lucky to be able to rely on your interest and help. I hope that this information is useful for you, apart from the more theoretical aspects, in helping her by, for instance, sharing a "healthy" meal or walk, whenever possible.'

Interestingly, in the rare cases of longer forum threads to which – in addition to the professionals – other patients also contribute, these patients similarly make the family relationship consequential by further discussing and emphasizing the importance of family members' support, as in Example (11). This is also the only example in the data where a patient uses the explicit first person singular subject pronoun *yo*. While the Spanish subject pronoun is generally much more frequently absent than present (see e.g. Enríquez 1984; Comajoan 2006; Travis 2007), this is even more true for the interactional settings analyzed in this chapter. Indeed, given the specific format, contributors already have the turn and do not need to claim it. Moreover, there is no need to affirm their identity or contrast it with others since they are often the sole interlocutor of the healthcare professional in a specific thread. It is then not surprising that this single use of *yo* by a patient appears in a thread with various contributors. By creating a contrast through *yo*, this contributor highlights that she is a diabetic herself (and intervenes in that capacity), as opposed to the participant asking the opening question of the thread, who is a relative of a diabetic.

In Example (12), the doctor involved in the thread explicitly refers to the information provided by the patient in (11) and acknowledges her as qualified to comment on the topic due to her personal experience.

(11) [in reply to a worried son of an elderly diabetic father]
[response by another patient]
Yo soy diabética desde los 9 años y tengo 23. Tu apoyo es fundamental para tu padre, necesita saber que cuenta con tu apoyo incondicional. (forum)
'I'm 23 years old and I've been a diabetic since I was nine. Your support for your father is essential, he needs to know that he has your unconditional support.'

(12) [subsequent response by professional]
Como muy bien te indica Lola, con experiencia en el manejo de su diabetes, tu apoyo puede ser muy importante para contribuir a concienciarle de la importancia de la alimentación para controlar mejor su enfermedad; (forum)

'As Lola, with her experience in managing her own diabetes, has pointed out very well, your support can be very important in raising his awareness of the importance of diet in keeping his illness better under control;'

This identity construction of the relative in (11) and (12) as not only an emotionally involved person but also as someone who has a concrete role to play in helping the patient manage diabetes is pervasive in the forum but absent in the chat data.

3.3 Healthcare professional identity

In this section, I study the way in which the healthcare professionals construct their own identity and how it is constructed by the other participants (patients and relatives). In the FAQs, professionals are not mentioned at all, perhaps because questions were edited before publication, although participants were explicitly informed that questions would be answered by a healthcare professional. So too were chat participants, who were told that a specific doctor would be answering questions during a particular session, making him/her the single, ratified interlocutor. Presumably as a result, the act of addressing a professional is frequently present in the chat data. In fact, most chat questions start with a greeting such as *hola* 'hello', and four out of 13 explicitly address the doctor by means of a vocative or ratify them in a more indirect way, e.g. by thanking them for devoting time to the session.

The situation on the forum is rather different. Here, any registered user may reply to a post, but users are informed that healthcare professionals participate in this particular sub-forum on diabetes – and, in fact, only two of 17 threads contain an answer by a layperson. The forum questions are not typically addressed to a specific interlocutor, although indirect strategies are sometimes used, such as the explicit mention of the need for a nutritionist in Example (13).

(13) *quisiera saber que dieta puedo comer hace poco me dio pancreatitis y soy diabetico necesito la orientacion de un nutriólogo* (forum)
'I would like to know what foods I can eat. Recently I got pancreatitis and I am diabetic. I need the guidance of a nutritionist.'

(14) *Mi hija de 13 años se tomo una curva de insulina y esta arrojo en ayunas 24.3 y en 120 min 453,es diabetes ????? cabe señalar que ella ya esta diasnostiada con resistencia de insulina* (forum)
'My 13 year old daughter did an insulin curve, and it showed 24.3 on an empty stomach and 453 after 2 hours; is this diabetes????? I should point out that she's already been diagnosed with insulin resistance.'

Alternatively, as in (14), questions may be formulated in a sufficiently specific way, including detailed information about the results of blood tests, as to require a fairly high degree of specialization on the part of the interlocutor. Consequently, one could argue that these questions are implicitly addressed to professionals. Other than through the use of greetings, however, the participants asking questions do not tend to construct the identity of these, generally limiting their representation to the purely professional role.

In light of the above, the following discussion focuses on how healthcare professionals represent themselves in their own contributions. The vast majority of these are constructed in an impersonal way, describing treatment and diagnosis with a focus on the processes involved, not on the participants. References to the collectivity of healthcare professionals, as in Example (15), where the author distinguishes family doctors from psychiatrists and psychologists, are rare. In cases such as (16), the reference made in describing the state of the art as regards diabetes research is vaguer; it may include all of society rather than just the medical profession.

(15) *Los compañeros psiquiatras y psicólogos clínicos son conscientes de que […] somos los médicos de familia quienes ejercemos la mayor labor de detección […] de una gran parte de los casos de ansiedad y depresión.* (forum)
'Our colleagues who are psychiatrists and clinical psychologists are aware that […] it is us family doctors who do the most work in detecting […] a large proportion of cases of anxiety and depression.'

(16) *Los estudios con células madre en diabetes están actualmente en fase de experimentación animal, y por tanto, estamos aún muy lejos de que sea una posibilidad cercana para tratar esta enfermedad […]* (FAQs)
'Studies of diabetes using stem cells are currently in the phase of animal testing, so we are still very far away from a possible treatment for this illness any time soon […]'

As shown in the discussion of (8), emphasizing the need of society at large for a healthy lifestyle is a way of construing shared responsibility and shared identity between diabetics, non-diabetics and professionals, and of blurring the professional identity. Interestingly, the clearest example of creating a shared identity with a questioner is not realized within a patient-professional dyad. Rather, Example (17) is part of a forum healthcare professional's reply to a participant who is worried about his diabetic father. The professional utilizes an inclusive *we*, creating a shared identity not in terms of a doctor-patient relationship but in terms of the shared identity of two sons of (elderly) parents.

(17) *Con los años nos vemos obligados a desarrollar hacia los padres una "piedad filial" que debe contribuir a entenderles, aceptarles y ayudarles mejor.* (forum)
'Over the years, we find ourselves obliged to develop a kind of "filial devotion" towards our parents, which must contribute to understanding them, accepting them and helping them better.'

Moreover, this is the only instance where a personal identity feature of the professional, albeit a very general one, is made relevant. It is also the clearest case of shared responsibility constructed between a questioner and the responding participant. This makes the fact that it does not pertain to the patient-professional relationship particularly striking.

4. Conclusion

In this paper, I have discussed how patients and healthcare professionals construct the identity of the self and other in three types of public online interaction about diabetes. The three data sets – FAQ, chat and forum – have in common that the contextual identity is to a large extent predefined by the platform. My results suggest that participants asking questions in fora, chats or FAQs do not tend to present themselves as diabetics but usually adopt a wording that barely stresses identity features relating to diabetes. When participants do refer to such identity features, they may pertain to a diabetic or to the relative of a diabetic. In all data sets, writers overwhelmingly appear to use either impersonal and generic constructions or highly individual representations, rather than constructing collective identities.

When participants asking questions explicitly portray themselves as diabetics or as close relatives of diabetics, they often include further detailed medical information and, especially on the forum, more emotional identity features, such as their attitude towards the illness. The healthcare professionals only refer to features that have previously been made explicitly relevant by the participants, thereby respecting their privacy choices. The specifically emotional features of both diabetics and their relatives are highlighted primarily in the forum data. This fact may be related to a difference in perceptions of the platforms' public nature; the forum does not aim at being generalizable (as is the case for FAQs) nor does it imply the simultaneous presence of a large group of readers (as in the chat session). Overall, few collective identities are being constructed and even fewer with the aim of creating the impression of co-responsibilisation between diabetics and professionals. Such joint responsibility is, however, constructed with relatives of diabetics, emphasizing that they can actively play a part in helping the patients with their treatment.

The healthcare professionals present themselves, and are referred to by other participants, in their professional capacity only. On the very rare occasions when professionals stress collective identity features that they share with patients, they try to construct the image of a more general collective societal responsibility for a healthy lifestyle, thus going well beyond the relationship between patient and healthcare professional. All in all, like the patients, the healthcare professionals are presented in a fairly traditional way.

The differences in identity construction between the three types of public online discourse studied seem to be related to their different degrees of atemporality and generalizability. Thus, the FAQs prefer impersonal constructions and are the only genre where at times a collective identity of diabetics is constructed. Other personal identity features remain absent as they would interfere with the aim of offering generalizable answers to frequently asked questions. The forum, on the other hand, includes the most personal information as well as the most indications about participants' emotional identity features, which are also more clearly taken up by the participants providing responses. The chat session occupies an intermediate position. Writers clearly identify as individuals and make personal identity features relevant. However, the nature of the interaction implies a large group of 'bystanders' and a short time span, which may explain why fewer emotional characteristics are mentioned.

While the healthcare professionals in these data use online communication as a means of spreading information about diabetes, the focus seems to be mainly on more generalizable information and less on a personalized approach. With the exception of the forum data, professionals emphasize above all medical identity features and pay much less attention to personal or emotional ones, despite their importance for the patient's personal biography and, therefore, for his/her diabetes management.

Acknowledgements

I would like to thank Ursula Lutzky and two anonymous reviewers for valuable comments on an earlier version of this manuscript. This paper has furthermore benefited greatly from discussions concerning parts of the analysis during the panel *Knowing me, knowing you: reference and identity markers in public discourse*, organized by Ursula Lutzky and Minna Nevala at the 2017 International Pragmatics Association conference, and during the 2017 Conversation Analysis Day in Loughborough. I would like to thank colleagues for their interesting suggestions and comments. Any mistakes remaining are, however, entirely my own.

References

Antaki, Charles. 2002. "An Introductory Tutorial in Conversation Analysis." Available online at http://ca-tutorials.lboro.ac.uk/sitemenu.htm (accessed 15 April 2019).

Antaki, Charles, and Sue Widdicombe. 1998. "Identity as an Achievement and as a Tool." In *Identities in Talk*, ed. by Charles Antaki, and Sue Widdicombe, 1–14. London: Sage Publications.

Antaki, Charles, Elisenda Ardévol, Francesc Núnez, and Agnès Vareyda. 2006. "'For She Who Knows Who She Is:' Managing Accountability in Online Forum Messages." *Journal of Computer-Mediated Communication* 11: 114–132. https://doi.org/10.1111/j.1083-6101.2006.tb00306.x

Antaki, Charles, and Rebecca J. Crompton. 2015. "Conversational Practices Promoting a Discourse of Agency for Adults with Intellectual Disabilities." *Discourse & Society* 26 (6): 645–661. https://doi.org/10.1177/0957926515592774

Brown, Penelope, and Stephen C. Levinson. 1987. *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511813085

Charmaz, Kathy. 1995. "The Body, Identity and Self: Adapting to Impairment." *The Sociological Quarterly* 36 (4): 657–680. https://doi.org/10.1111/j.1533-8525.1995.tb00459.x

Clyne, Wendy, Comfort Mshelia, Sarah McLachlan, Peter Jones, Sabina de Geest, Todd Ruppar, Kaat Siebens, Fabienne Dobbels, and Przemyslaw Kardas. 2016. "A Multinational Cross-Sectional Survey of the Management of Patient Medication Adherence by European Healthcare Professionals." *BMJ Open*. https://doi.org/10.1136/bmjopen-2015-009610

Comajoan, Llorenç. 2006. "Continuity and Episodic Structure in Spanish Subject Reference." In *Functional Approaches to Spanish Syntax: Lexical Semantics, Discourse and Transitivity*, ed. by J. Clancy Clements, and Jiyoung Yoon, 53–79. Basingstoke: Palgrave Macmillan. https://doi.org/10.1057/9780230522688_3

Corbin, Juliet M., and Anselm Strauss. 1988. *Unending Work and Care. Managing Chronic Illness at Home*. San Francisco: Jossey Bass Publishers.

Cysouw, Michael. 2003. *The Paradigmatic Structure of Person Marking*. Oxford: Oxford University Press.

De Cock, Barbara. 2010. "La funcionalidad discursiva del esquema construccional: Los españoles hemos hecho historia en Europa." *Lingüística Española Actual* 32 (2): 173–195.

De Cock, Barbara. 2011. "Why *We* Can Be *You*: the Use of 1st Person Plural Forms with Hearer Reference in English and Spanish." *Journal of Pragmatics* 43: 2762–2775. https://doi.org/10.1016/j.pragma.2011.04.009

De Cock, Barbara. 2014. *Profiling Discourse Participants. Forms and Functions in Spanish Conversation and Debates* (Pragmatics & Beyond New Series 246). Amsterdam: John Benjamins.

De Cock, Barbara. 2016. "Register, Genre and Referential Ambiguity of Personal Pronouns: A Cross-Linguistic Analysis." *Pragmatics: Quarterly Publication of the International Pragmatics Association* 26 (3): 361–378. https://doi.org/10.1075/prag.26.3.02dec

De Fina, Anna. 2010. "Discourse and Identity." In *Discourse Studies: A Multidisciplinary Introduction*, ed. by Teun A. Van Dijk, 263–282. London: Sage.

De Wilde, Julie, Ellen Van Praet, and Yves Van Vaerenbergh. 2019. "Language Discordance and Technological Facilitation in Health Care Service Encounters: a Contrastive Experiment." In *Mediated Service Encounters*, ed. by Pilar Garcés-Conejos Blitvich, Lucía Fernández-Amaya, and María de la O. Hernández-López, 17–44. Amsterdam: John Benjamins.

Drew, Paul, John Chatwin, and Sarah Collins. 2001. "Conversation Analysis: a Method for Research into Interactions between Patients and Health-Care Professionals." *Health Expectations* 4: 58–70. https://doi.org/10.1046/j.1369-6513.2001.00125.x

Duszak, Anna. 2002. "Us and Others: an Introduction." In *Us and Others: Social Identities Across Languages, Discourses and Cultures*, ed. by Anna Duszak, 1–28. Amsterdam: John Benjamins. https://doi.org/10.1075/pbns.98.01dus

Goffman, Erving. 1979. "Footing." *Semiotica* 25 (1–2): 1–29. https://doi.org/10.1515/semi.1979.25.1-2.1

Haskard Zolnierek, Kelly B., and M. Robin DiMatteo. 2009. "Physician Communication and Patient Adherence to Treatment: a Meta-Analysis." *Med Care* 47 (8): 826–834. https://doi.org/10.1097/MLR.0b013e31819a5acc

Keel, Sara, and Veronika Schoeb. 2017. "Patient Participation in Action: Patients' Interactional Initiatives During Interdisciplinary Goal-Setting Meetings in a Rehabilitation Clinic." *Text & Talk* 37 (2): 213–241. https://doi.org/10.1515/text-2017-0004

Kerbrat-Orecchioni, Catherine. 2005. *Le discours en interaction*. Paris: Armand Colin.

Kerr, Cicely, Elizabeth Murray, Fiona Stevenson, Charles Gore, and Irwin Nazareth. 2006. "Internet Interventions for Long-Term Conditions: Patient and Caregiver Quality Criteria." *Journal of Medical Internet Research* 8 (3): e13. https://doi.org/10.2196/jmir.8.3.e13

Loader, Brian D., Steve Mucer, Roger Burrows, Nicolas Pleace, and Sarah Nettleton. 2002. "Medicine on the Line? Computer-Mediated Social Support and Advice for People with Diabetes." *International Journal of Social Welfare* 11: 53–65. https://doi.org/10.1111/1468-2397.00196

Liu, Na, Yu Tong, and Hock Chuan Chan. 2017. "Information Seeking in Online Healthcare Communities: the Dual Influence from Social Self and Personal Self." *IEEE Transactions on Engineering Management* 64 (4): 529–538. https://doi.org/10.1109/TEM.2017.2703121

Oris, Leen, Jessica Rassart, Sofie Prikken, Margaux Verschueren, Liesbet Goubert, Philip Moons, Cynthia A. Berg, Ilse Weets, Koen Luyckx. 2016. "Illness Identity in Adolescents and Emerging Adults with Type 1 Diabetes: Introducing the Illness Identity Questionnaire." *Diabetes Care* 39 (5): 757–763. https://doi.org/10.2337/dc15-2559

Pal, Kingshuk, Charlotte Dack, Jamie Ross, Susan Michie, Carl May, Fiona Stevenson, Andrew Farmer, Lucy Yardley, Maria Barnard, and Elizabeth Murray. 2018. "Digital Health Interventions for Adults with Type 2 Diabetes: Qualitative Study of Patients' Perspectives on Diabetes Self-Management Education and Support." *Journal of Medical Internet Research* 20 (2): e40. https://doi.org/10.2196/jmir.8439

Rollnick, Stephen, William Richard Miller, and Christopher C. Butler. 2008. *Helping Patients Change Behavior*. New York: The Guilford Press.

Sacks, Harvey. 1995. *Lectures on Conversation*. Oxford: Blackwell. https://doi.org/10.1002/9781444328301

Salonen, Elise. 2018. "Constructing Personal Identities Online. Self-Disclosure in Blogs." In *The Discursive Construction of Identities Online and Offline*, ed. by Birte Bös, Sonja Kleinke, Sandra Mollin, and Nuria Hernández, 57–79. Amsterdam: John Benjamins.

Salvador, Vicent, and Cecili Macián. 2009. "Primera relación entre paciente y profesional de salud: rituales de interacción y estrategias discursivas en la entrevista podológica." *Estudos linguísticos* 3: 439–457.

Spilioti, Tereza. 2017. "Media Convergence and Publicness: Towards a Modular and Iterative Approach to Online Research Ethics." *Applied Linguistics Review* 8 (2–3): 191–212. https://doi.org/10.1515/applirev-2016-1035

Travis, Catherine E. 2007. "Genre Effects on Subject Expression in Spanish: Priming in Narrative and Conversation." *Language Variation and Change* 19: 101–135. https://doi.org/10.1017/S0954394507070081

What's in a diminutive?

The pragmatics of the Spanish diminutive in a televised political interview and its reverberations in online comments

Bettina Kluge
University of Hildesheim

This contribution attempts to reconstruct the different values associated with the diminutive *-ita*, which was repeatedly added to the first name of Spanish journalist Ana Pastor by Ecuadorian president Rafael Correa in a 2012 televised interview, and against which she protested explicitly. While Correa has claimed that his use of *Anita* was intended to show respect and affection, some online followers interpret it as a sign of *machismo*. The contribution reflects on these interpretations, also in light of two other interviews between Correa and Pastor and of the continuing online discussion among commentators from Spain, Ecuador and other countries. Ultimately, different conceptions of journalist practices can be shown to be at the heart of the discussion.

Keywords: Spanish, diminutive, nominal address, political news interview, professional identity, intercultural misunderstanding, *machismo*, gender, online comments, term of reference

1. Introduction

On March 19, 2012, the Spanish journalist Ana Pastor interviewed Rafael Correa (president of Ecuador from 2007 to 2017) in her morning show *Los desayunos de TVE* on Televisión Española (TVE), the national public-service broadcaster. Both she and her guest were well-known for being outspoken and not accepting criticism lightly, so a lively debate was to be expected. In the thirty-minute interview, Pastor makes abundant use of the title *Presidente* (or *Presidente Correa*) to address her interlocutor. Meanwhile, Correa repeatedly addresses his interviewer as *Anita*, despite the latter's protests that her name is *Ana*. In Spanish, use of the

https://doi.org/10.1075/pbns.306.09klu

diminutivized form of the first name is considered a sign of familiarity, tenderness, and/or endearment, but can also be interpreted as a sign of condescension (Zuluaga 1970; Hummel 1995 and 1997), especially when used asymmetrically and in public. Ana Pastor clearly saw the danger this posed to her professional identity as a seasoned political journalist and to her local identity as an interviewer (Goffman 1959; Brown and Levinson 1987), as is shown by her protests. After the interview, and in view of the ensuing comments in various social media channels, Correa reframed his use of the diminutive *Anita* as a typical Latin American custom, recategorizing the incident as an intercultural misunderstanding, while Pastor said that she had felt *incómoda pero no ofendida* ('uneasy but not offended') during the interview. In fact, a close analysis of the eight instances of *Anita* reveals a certain intentionality by the Ecuadorian president, at least in the second half of the interview.

The use of the vocatives *Anita* and *Presidente* has already been analyzed by Cuenca (2013), and I will follow her analysis in great part.[1] However, I feel that Cuenca's discussion does not take adequate account of the role played by long-term face considerations related to the public identities Correa and Pastor claim for themselves, especially in view of the dispersed audience in Spain, Ecuador and other parts of the world. At the same time, the use of the diminutive itself merits further consideration, especially in light of the subsequent online discussion among commentators from Spain, Ecuador and other Spanish-speaking countries. This paper will briefly discuss two subsequent interviews (in 2013 and 2014) after Pastor's switch to *CNN en español*. More importantly, however, it will also turn to the reactions of the general public in social media comments, with commentators voicing strong opinions on the incident even today, in 2018, and often adopting either Correa's or Pastor's stance. Here, a strong instrumentalization of the incident becomes visible. It will be shown how the incident becomes instrumentalized, with those users who are admirers of Correa making fun of, or criticizing Pastor, while his critics jump to Pastor's defense, with Latin Americans often taking the first stance and Spaniards the second.

The remainder of the paper is structured as follows. By way of background to the 'Anita incident', Section 2 briefly discusses the use of the Spanish diminutive with personal names. Sections 3 and 4 then reflect on the dynamics of televised news interviews, with special emphasis on the use of address terms during them, and the facework done on a local and more long-term scale. In Section 5, I turn to a description of the original interview, while Section 6 considers the two

1. I only became aware of Cuenca's article in 2018 when updating my previous analysis dating back to 2014. In a sense, it is an encouraging sign that we reached similar conclusions about the data.

subsequent interviews and the reaction of online users to one YouTube video of the entire original one. The final Section draws together my main conclusions.

2. The Spanish diminutive

The meaning and pragmatic functions of Spanish diminutives are notoriously hard to grasp. Most scholars agree that besides conveying a smaller than normal size for an object, another aspect of the meaning of diminutives is the expression of endearment and affection.[2] Use of the diminutive can convey familiarity, tenderness, and/or endearment, but also condescension (Zuluaga 1970; Hummel 1995, 1997), especially if it is used asymmetrically and in public. Diminutives are more often used with women's first names than men's, and they are far more common in Latin America and on the Canary Islands (Travis 2005) than in mainland Spain.

Travis (2005) finds that, in her corpus of conversational, colloquial Colombian Spanish, the diminutive is very frequent, occurring approximately 600 times in the 70,000-word corpus. She shows that, in addition to its core uses in relation to children and to express small size, it has taken on the pragmatic functions of expressing not only affection, but also contempt, as well as hedging speech acts. She asserts that "[t]he diminutive proves to be an essential linguistic resource in Colombian culture, embodying many of the interpersonal ideals central to everyday interaction. Its role in encoding good feelings allows speakers to show *calor humano* 'human warmth' to others, and maintain *confianza* 'trust'" (Travis 2005: 269).

While these observations were based on Colombian data, they should be more or less applicable to Ecuador, as is shown by Placencia's (2010) very similar observations on address terms there. She notes that first names, and their abbreviated forms, often have intensifying forms to express tenderness (*cariño*) between interlocutors who know each other well and trust each other. In the public sphere, especially in the Ecuadorian Sierra, address is much more formal, e.g. *señor ingeniero* (Placencia 2010: 368), whereas many service encounters are marked by forms like *mijita* and *mamita* (*mijita* < *mi hija* 'my daughter', *mamita* 'Mommy'), even in initial encounters between strangers. Placencia (2010) also notes, however, that some middle-class interviewees in her study qualified these uses as *meloserías*, i.e. abuses of confidence. Thus, it seems safe to say that the diminutive is a powerful tool to negotiate interpersonal relations regarding the proximity and

2. In Spanish, it is also possible to add a diminutive suffix to parts of speech other than nouns, especially to adjectives (*feo* > *feito* 'ugly') and adverbs (*ahorita* 'right now'). This is more common in Latin America than in Spain. However, since my reasoning is mostly concerned with proper names, I will leave out this aspect of diminutive suffixation.

distance of interlocutors, and an important means of developing and displaying *confianza* 'trust'.

3. Power relations and roles in political interviews

The dynamics of the political news interview are well understood. According to Rendle-Short (2011: 94), "the very design of the political news interview means that journalists are restricted to asking questions and politicians to providing answers". Power asymmetries in the news interview are thus two-fold. Politicians, especially in the case of democratically-elected presidents or prime ministers, wield more political power, yet they are constrained by the status of the 'fourth estate' and the conversational requirements of the media format. Moreover, thanks to the question-and-answer format, journalists enjoy local power as they are in a position to insist on answers to their questions or to initiate topic changes as they see fit.

The question of how aggressive a journalist's questions may be depends on national norms, and sometimes even on those of the broadcasting corporation or of the specific program in which the interview is shown. Are journalists expected to take a neutral stance, or can they press some issues they feel important to clarify? We will see later that this issue is especially relevant when a journalist interviews a politician from another country who might have different expectations about the question-and-answer format. For the Australian context, Rendle-Short (2011: 93) refers to the *Australian Broadcast Corporation's Code of Ethics*, which stresses four fundamentals: "fairness, independence, reliability and sensitivity". Similar guidelines are found in other countries. In the case of Spain's TVE, the broadcaster is required by law to provide objective, truthful news coverage and to give voice to a plurality of opinion (Ley de la Radio y la Televisión Estatal 17/2006, article 3; cf. Boletín Oficial de España 2006). TVE has gained international acclaim for doing so on various occasions.[3]

3. However, in the last few years there have been repeated accusations that TVE's news programs have favored Prime Minister Mariano Rajoy and his People's Party by not reporting extensively on accusations of fraud and by giving less coverage to political adversaries. In 2015, the channel's own journalists posted a formal complaint to the European Parliament (Buck 2015), and, shortly before Rajoy's government was swept away among accusations of corruption on June 5, 2018, journalists announced that they would wear black every Friday to protest the broadcaster's perceived political bias (Trueba 2018). Accusations of political bias are not new, however; in 2012, after the PP's general election victory, several journalists were laid off – one of them being none other than Ana Pastor, the journalist involved in the 'Anita incident'. After losing her job in August 2012, Pastor claimed she was fired for asking difficult questions instead of 'playing it nice' and doing her job as a journalist (Burgen 2012).

4. Address terms in the political news interview

The use of terms of endearment and vocatives in political interviews or discussions has mostly been analyzed in the Anglophone context (e.g. Jaworski and Galasiński 2000; Rendle-Short 2007, 2011; Clayman 2010; Kerbrat-Orecchioni 2010). There, analyses by Clayman (2010) of U.S. data and by Rendle-Short (2007, 2011) of Australian data suggest that the public use of first names in interviews is becoming more common. Yet distanced terms of nominal address are also common, often in the combination of an honorary *Mr / Ms / Madam* followed by the current political function, e.g. *Mr President, Senator, Madam Secretary*.

As regards Spanish, some studies discuss the use of address terms in electoral debates, that is, among the candidates involved, and between them and the journalists moderating the debate. Examples are Blas Arroyo (2000, 2003) on the debates between Aznar and González, the two leading candidates in the Spanish general elections of 1989 and 1993, and Vázquez Laslop (in press) on Mexican debates between presidential candidates in the 1994 and the 2012 elections. Her Mexican data show a considerable increase in first name address. However, no study has analysed a political interview as such.

However, Clayman (2010), working on interviews with U.S. politicians, raises a more fundamental question: Why do address terms appear in the news interview in the first place? After all, in dyadic news interviews, address terms should not account for too many occurrences of turn allocation, since the selection of the next speaker is "already established by the activity structure and participation framework" (Clayman 2010: 163). Additionally, directionality of talk is aided by the interviewee's bodily orientation and gaze direction. Therefore, "an explicit mode of address" would seem to be "thoroughly redundant and unnecessary" (ibid.).

On the basis of U.S. data, Clayman (2010: 163) concludes that address terms in news interviews seem to serve as indicators either of disalignment, "including topic shifts, non-conforming responses, and disagreement", or of a particularly sincere answer ("responses that are portrayed as particularly significant and sincere", ibid.). Address terms are preferably turn-initial, and, especially when marking disalignment, are often combined with discourse or hesitation markers (*uh, well*). In the second case, where address terms are supposed to mark a specific argument as especially sincere, they can range more freely "across the action and occur frequently at the end of the turn constructional unit through which it is launched" (Clayman 2010: 179). In fact, they are often found at the end of a turn-constructional unit. Using an address term in this context conveys the image of a person who is so entrenched in their line of argumentation that they forget about the mass media audience watching the interview, for the sake of convincing the interlocutor of the argument just expressed.

> For now it may be noted that address terms, which cast the talk in progress as directed to a singular co-present recipient, have a distinctly personal and intimate quality in the context of broadcasting to a mass audience. They *explicitly* disattend the media audience for whom the interaction is ostensibly being conducted. Perhaps 'speaking personally' is a way of indicating that one is not 'playing to the audience', which may in turn further the interviewee's self-presentation as genuine and sincere. (Clayman 2010: 180)

To my knowledge, all the cited analyses examine the use of first names or titles such as *Mr President*. To date, Cuenca's (2013) analysis of *Anita* is the only one to have considered the use of diminutive forms in a political interview, to address either a politician or a journalist.[4]

5. Analysis of the 'Anita incident': Public roles and media format

Throughout the interview, the interlocutors address each other with polite *usted* to show mutual respect and distance in their public roles as interviewer and interviewee. However, use of nominal address is asymmetric. Pastor calls her interviewee *Presidente* on 24 occasions but never *Rafael*. In contrast, Correa addresses her by her first name ten times: eight times as *Anita*, including one *mi querida Anita* 'my dear Anita', and only twice as *Ana*.

As is apparent from the public reactions after the interview, the reasons behind Correa's use of the diminutive *Anita* remain unclear. It can be interpreted either as a deliberate sign of condescension (or *machismo*, as some commentators allege), or as an unintentional sign of affection and respect (see Section 1). But is the interpretation the same for all occurrences of *Anita*? Clayman's (2010) observations on the positioning of first names in political interviews (turn-initial, intermediate, turn-final) and their correlation with topic-changing activities are especially important here. However, before we turn to analyzing the occurrences in the interview itself, it is pertinent both to reflect on the local and long-term face needs of Pastor and Correa, and to consider the media format itself, that is, the setting of *Los desayunos de TVE*.

4. This does not apply to forms of habituation, for example, the use of *Bill* for former U.S. President Bill (< William) Clinton, or *Joschka* for Germany's former Foreign Minister Joschka (< Joseph) Fischer, both of which are unmarked.

5.1 The long-term face needs of Pastor and Correa

Ana Pastor García, born in 1977, is a Spanish journalist who worked for TVE from 2006 to August 2012, when she was dismissed from *Los desayunos de TVE*. She then went on to work for the program *Frente a frente* ('Face to face') broadcast on the international channel CNN en español, which acclaimed her as one of the most prestigious Spanish journalists due to her rigor and independence ("*una de las periodistas españolas más prestigiosas, gracias a su rigor y a su independencia*", CNN en español 2012). Since 2013, she has also fronted a program on the private Spanish channel La Sexta called *El objetivo*, which specializes in fact checking. She defines her role as a journalist as an intermediary between the public and the people she interviews, her aim being to reveal their intentions, convictions, and beliefs. Her rather aggressive questioning techniques used against both sides of the political spectrum (the ruling party and the opposition alike) have been labeled "*periodismo 'punch*'"; over the years, she has acquired the reputation of never letting a question go unanswered.

Rafael Correa, born in 1963, was a university professor of economics before serving as Ecuador's Finance Secretary for four months in 2005. In 2007, he was elected his country's president, a position he held until 2017 (having been reelected in 2009 and 2013). His political orientation is fairly leftist, and, while in power, he was a political ally of fellow Latin American presidents Hugo Chávez, Evo Morales, and Lula da Silva. His policies significantly reduced poverty levels in Ecuador, and he gained political sympathy among many left-leaning people in Latin America and Europe alike. Others, though, labeled him a populist and demagogue, pointing, among other things, to his troubled relation with the Ecuadorian press.

Thus, we can assume different long-term face claims that depend on the previously established public image and status of the two interlocutors. Pastor will have wanted to establish her public image as a seasoned, popular and critically acclaimed journalist used to asking critical questions, while perhaps hoping to extract from her interlocutor an assertion that makes the headlines. Correa, as president of a small country and 'darling' of leftist movements in Latin America and Europe, will have hoped to convince a very diverse public of his status as a seasoned, popular politician who is independent in his thinking. It should also be borne in mind that, before the interview, he had visited the Ecuadorian immigrant community in Spain. In fact, Correa repeatedly addressed these viewers (and potential voters in elections to be held in 2013), which suggests he was well aware of the interview's international dimension. It is unclear, though, if he was aware of Pastor's self-image as a critically acclaimed journalist, and of her particular questioning style.

It is unclear whether age and gender identities also came into play as relevant face categories. At least the fact that several commentators in the online debate accuse Correa of *machismo* show that gender is (perceived as) a factor. As regards age, Correa may be somewhat older, but the difference (14 years) is not particularly significant. At stake are thus the professional identities of both Pastor and Correa.

5.2 The media format *Los desayunos de TVE*

Los desayunos de TVE has been on air since 1994 in a morning slot (9:00–10:30) and can be seen in full in an archived version on TVE (*Los desayunos de TVE* 2012). Ana Pastor was the program's main presenter from 2009 until July 2012. The format involves three to four journalists from other media (mostly newspapers) coming together in the television studio to discuss the day's most important news items. Its highlight, however, is the main interview, about 20 to 30 minutes in length, usually with a national or international politician, but sometimes with writers, artists, or actors. This interview, which is often recorded prior to the show, is subsequently analyzed by the journalists in the studio. In the case of the interview with Rafael Correa, which was pre-recorded, the three journalists present were Marisa Cruz of *El Pais*, Luis Aizpeolea of *El Mundo*, and Fernando Berlín of *Radiocable*.

During the *Anita* interview, the interlocutors are seated in swivel chairs about two meters apart, as can be seen in the archived video of the morning show (*Los desayunos de TVE* 2012). Ana Pastor's movements are very limited. Holding some notes and with her legs crossed, she sits on the edge of her chair, possibly in order to reduce the intervening distance or to communicate alertness. Correa leans comfortably in his chair, adopting various leg positions. His gestures become very lively during the interview, as do his facial expressions.

The studio setting is complemented by two multimodal devices; two large screens can be used to show photos and video clips previously prepared by the team of *Los desayunos de TVE*. The material shown on them becomes a relevant topic during some parts of the interview and is referenced by both Correa and Pastor. Occasionally, the screens display inserts showing the most important quotes attributed to Correa. These were produced post-interview but prior to broadcast, as can be seen when some inserts are superimposed before Correa actually makes the statements in question.[5] The inserts are visible for several seconds and are thus important for viewers unable to devote their full attention to the broadcast.

5. Some commentators in the two YouTube videos (Sucreranda Hugo Chávez Venezuela 2012; IUAVILA1 2012) note that the inserts sometimes misrepresent Correa's assertions. This later led to an official complaint by the Ecuadorian government, e.g. for mispresenting Correa as having

Hence, the political interview in question was a simulation of a dialogue inserted in a complex mediatized event. Before it, both interlocutors had acquired a certain public image that now needed to be upheld via facework. This is all the more interesting in the case of Rafael Correa as he was visiting Spain, and the Ecuadorian expat community there, but was also aware of his popularity with left-leaning circles in Europe. Thus, the audience he was addressing was somewhat broader than Pastor's.

5.3 *Anita* and *Presidente* in the political interview

In her analysis of the *lucha de vocativos* 'battle of the vocatives', Cuenca (2013: 529) points to the role of the pre-interview phase, before the cameras were turned on. As is made clear in Ana Pastor's opening statement, interviewer and interviewee apparently talked in broad terms about the topics of the interview, one of them being the freedom of the press in both Ecuador and Spain.

Cuenca (2013) interprets the role of the frequent use of vocatives in the interview as indicating an ongoing fight for discursive power between interviewer and interviewee (and here I would add, a struggle for the conception of how a political interview should be conducted). In Pastor's case, outside the opening and closing sections, where her use of the honorific title is clearly polite, her use of *presidente* has a slightly aggressive undertone. Cuenca observes how the interviewer often makes use of *presidente* in moments of disalignment, when she questions the veracity of Correa's statements or is fighting for the floor to finish her turn. These uses often come in a turn-final position or apply to questions that Correa had previously indicated he was not willing to answer, such as whether he would confirm his attendance at the Summit of the Americas in Colombia (Cuenca 2013: 532). This is much in line with Clayman's (2010) findings, cited above, that turn-final vocatives explicitly disattend the media audience for whom the interaction is ostensibly being conducted, by conveying the image of someone focused on the activity performed (here, the interview).

Cuenca (2013: 535) considers Pastor's general conduct of the interview to be extremely impolite, since she often fails to apply the generally accepted genre conventions (i.e. a journalist asks questions but does not offer their own opinions and/or judgements). Cuenca counts only 55 questions during the 120 turns initiated by Pastor, some of which are truncated because they occur while the interlocutors are

implied that the NGO Human Rights Watch was financed by the Mexican Sinaloa drug cartel (he had meant this as a rhetorical question, clearly ironic). In the officially archived version of the show (*Los desayunos de TVE* 2012), most inserts have been removed, unlike in the YouTube videos cited above.

fighting for the floor. Thus, Cuenca finds Pastor's conduct "*claramente trasgresor*" 'clearly transgressive' (ibid.), reminding us of Heritage and Clayman's (2010: 218) assertion that "the obligation to question is […] such a pervasive constraint on journalists' conduct that the vast majority of their contributions are indeed limited to questions." That is not the case here.

Cuenca (2013: 538) also observes that Correa interrupts Pastor more frequently than he is interrupted by Pastor. She seems to put the blame for this on Pastor, who often asks very elaborate questions, with Correa taking advantage of short pauses and other transition-relevant places within these. However, Correa's answers are also very elaborate. He also has a tendency to expand on his own answers after a short pause during which Pastor is preparing to ask another question or comment, or has actually begun to do so.

The 24 tokens of *presidente* (cf. Cuenca 2013: 547–549 for a complete list) are spread throughout the entire interview, but they occur especially in a segment on Correa's troubled relations with the press. There, while formally a sign of respect and seemingly enhancing Correa's positive face, the use of *presidente* emphasizes Pastor's positive face as a person who wants to be taken seriously as a journalist asking tough questions. Furthermore, as Cuenca observes, Pastor's use of this vocative is often in turn-final position (in 13 of 24 cases). Together with its overuse (24 tokens in 120 turns), this gives it an aggressive undertone, in the sense of demanding an answer, especially in those cases where her interviewee has previously tried to avoid the answer.

Turning to the instances of *Anita*, I will analyze two of these in greater detail as they occur at decisive moments in the interview (and are also those most commented on later by journalists and on social media). The first of these occurs during the introduction and the second one in the last third of the interview, immediately before Pastor protests at being called *Anita*.

The first example of *Anita* appears at the very beginning of the interview, in the initial greeting (see Example (1)).

(1) 'buenas tardes a usted anita' (min. 2:06)[6]

```
001  AP:  hoy saludamos al presidente correa (.) al
002       presidente de ecuador (. ) presidente (.) buenas
003       tardes
004  RC:  (--) buenas tardes a usted anita y buenas tardes a
005       toda españa
006  AP:  ((inhaling)) bueno ya ven que las formas son
007       diferentes aquí (.) vamos a hablar mucho (.)
008       estábamos [antes de]
009  RC:            [((laughter))]
010  AP:  empezar hablando mucho de las relaciones con la
011       prensa pero empezamos por lo que ha hecho en
012       españa en estas horas
```

Translation:

```
001  AP:  today we welcome president correa (.) the
002       president of ecuador (. ) president (.) good
003       afternoon
004  RC:  (--) good afternoon to you anita and good afternoon
005       spain
006  AP:  ((inhaling)) well you can see already that the conventions
007       are different here (.) we will talk a lot (.)
008       we were [before the]
009  RC:          [((laughter))]
010  AP:  interview we were talking a lot about relations with the
011       press but let's begin with what you have been doing in
012       spain for the last couple of hours
```

Pastor's formal greeting *presidente buenas tardes* (lines 2/3 'President, good afternoon') sets up a professional frame dominated by mutual respect and formality. This contrasts with Correa's much more jovial greeting *buenas tardes a usted anita y buenas tardes a toda españa* (line 4 'Good afternoon Anita, and good afternoon Spain'). Here, Correa explicitly addresses the mediatized nature of the interview situation. There is no indication that the question of how to address each other in the interview was discussed in advance, so the Ecuadorian president most likely uses the diminutive according to his own communicative norms, to reduce the emotional distance from his interlocutor. For Pastor, however, this is too intimate in the professional context, with her professional identity as an impartial and tough

6. The examples were transcribed by my students Mayra Condemarín Cellio (Example 1) and Melissa Herrera (Example 2), following the GAT2 transcription conventions (Selting et al. 2009). Note that Melissa pays greater attention to prosodic features.

interviewer at stake. She therefore immediately directs a comment towards the audience that marks the diminutive as something noteworthy and unusual (*ya ven que las formas son diferentes aquí*, 'well you can see already that the conventions are different here', lines 6/7). This is the only time she directly addresses the audience during the interview.[7] She also reminds the audience of the intercultural situation she and her guest find themselves in (as mentioned above, Spaniards know, at least in theory if not always from first-hand experience, that Latin Americans use more diminutives). To Correa, the incident should have made clear that the diminutive has irritated Pastor. Had he used *Ana* instead of *Anita*, addressing her by her first name would probably have gone unremarked, especially in his combined greetings to the interviewer and the audience (cf. Clayman's 2010 observation that in a polylogue the vocative serves to indicate who is addressed). Note that while Pastor may address Correa by the use of bare *presidente*, Spanish does not allow Correa to address Pastor in an analogous manner by referring to the professional identity she is currently assuming: it would be impossible to call her *periodista* 'journalist'.

Various further tokens of *Anita* (at minutes 8:53, 22:34, 24:33 and 25:00) are met by Pastor's seemingly ignoring them, although she still notes them, as is palpable from her later reaction. In one case (min. 24:32–24:37), Correa explicitly reverses the local roles of interviewer and interviewee by asking *le puedo hacer una pregunta anita. usted cree que soy un represor que- atento derechos humanos?* 'May I ask you a question, Anita, do you think that I am an oppressor, that I violate human rights?'. Here, Pastor does not react to being called *Anita*, but she does comment *lo siento presidente decirle que yo no respondo preguntas (.) yo las hago* 'I am sorry, Mr. President, to tell you that I do not respond to questions, I ask them'. The use of *Anita* increasingly seems to have a certain intentionality, designed to provoke the journalist whose professional identity is challenged by the suggested reversal of roles. Pastor's attempt to regain control is successful in the sense that she is able to continue her questions. However, she still loses face as she is forced to make the activity structure of the interview explicit, thus departing from the illusion of a friendly dialogue.

The second instance of *Anita* I wish to analyse in detail (see Example (2)), the seventh occurrence in all, comes in the final section of the interview (on Ecuadorian press laws; Correa wants the press not to openly endorse any one candidate). It is the one after which Pastor asks not to be called by the diminutivized form of her first name.

7. Pastor raises her eyebrows while addressing the audience, giving a questioning or surprised impression, but staying professional by turning to Correa again and talking about the topics planned for the interview. As noted above in Section 5.2, it must be kept in mind that her main audience is the Spanish one, while Correa is addressing both Ecuadorian and Spanish viewers.

(2) 'aNIta no se equiVOque' (min 26:44)

326 RC: [quién] lo va a decidir (.) NO (.) un conSEjo nacional
327 electoral indepenDIENte que esTÁ converSANdo con los
328 medios pero lo que quieren es ser los grandes electores
329 **aNIta (.) no se equiVOque** y ESO es lo antidemocrático
330 AP: **no estoy acostumbrada ni a responder a preguntas**
331 **[en las entrevistas] ni a que me llamen anita** ya se lo digo (.)
332 RC: [((risas))]
333 AP: [[tiene otra polémica]]
334 RC: **[[cómo se llama]]**
335 AP: **ana ana**
336 RC: **ana** (.) ok con todo gusto
337 AP: hay otra polémica […]

Translation:

326 RC: [who] will decide this (.) NOT (.) an independent national
327 electoral council that talks to the
328 media but which wants to be somehow above the voters
329 **aNIta (.) make no mistake:** that is the undemocratic aspect
330 AP: **I'm not used to answering questions**
331 **[in interview situations] nor to being called anita** I already told you (.)
332 RC: [((laughter))]
333 AP: [[there's another issue]]
334 RC: **[[what's your name]]**
335 AP: **ana ana**
336 RC: **ana** (.) ok my pleasure
337 AP: there's another issue […]

Since Correa argues very passionately against media influence, employing lively gestures and facial expressions, the instance of *Anita* in line 329 appears as very sincere, with the speaker disattending the media audience entirely. But clearly, by linking it to an imperative *anita no se equivoque* 'Anita, make no mistake', the president directly threatens Pastor's face. Her reaction is swift: *no estoy acostumbrada ni a responder a preguntas ni a que me llamen anita* ('I'm not used to answering questions nor to being called Anita'). This is apparently intended as a side sequence similar to the previous one, but she is interrupted by Correa who (at minute 26:53) asks her first name, and repeats it in a ratifying manner, almost as if memorizing it. In fact, at minutes 27:17 and 28:44, he calls her *Ana*, in the first case in turn-final position and with a lot of emphasis.

The last instance of *Anita* occurs in the closing section of the interview, at minute 30:21, after Pastor begins to thank Correa for his time. The president interrupts her to also thank her and pauses slightly before adding her first name – in a diminutivized form and combined with a broad smile: *a usted mi querida (.) Anita* ('and thanks to you too, my dear (.) Anita'). This final occurrence of *Anita* has a very intentional feeling to it, especially after Correa has used her first name *Ana* as requested twice and because of his pausing and smiling as he says *Anita*. Pastor immediately adds her name *ana (.) pasTOR* and speaks her closing words; remaining in her professional role, she thanks Correa for having given his time for the interview.

In a sense, the interview, the use of the various vocatives, and especially of the diminutivized form of the journalist's first name, are all very noteworthy because they do not suggest an easy and clear-cut interpretation. Both interlocutors are communications experts who knew what they were doing. What they might not have been prepared for – at least not in practice – was the fact that the use of Spanish, an inherently pluricentric language, is governed by different norms in different countries. Moreover, the diminutive form, with its connotations of tenderness and respect as well as condescension, is inherently ambiguous, as was noted in Section 1. Thus, as Cuenca (2013) points out, the vocatives *Anita* and *Presidente* need to be considered as a pair, and in conjunction with other factors: the dynamics of political interviews, the genre itself, its mediatized form, and the image claimed by both participants.

5.4 Post-interview reframing of the incident by Ana Pastor, her colleagues, and Rafael Correa

While it is not entirely possible to reconstruct the perceptions of what went on during the interview itself, it was reframed immediately afterwards in various ways in light of the public's reaction (see below). Indeed, Correa's first public reaction occurred on Twitter, in response to a user's question about how he perceived the interview (see Example (3)).

(3) Tweet by Rafael Correa (@MashiRafael):

@MarioTroya No recuerdo, pero nos entendimos muy bien. Fue una entrevista muy agradable. Anita (costumbre Latina) es muy professional.

20/03/2012, 03:06am

'I don't remember, but we got along very well. It was a very enjoyable interview. Anita (Latino usage) is very professional.'
(https://twitter.com/MashiRafael/status/181939642562646016, accessed 15 April 2019)

In his tweet, Correa cites generally known findings according to which the diminutive is used much more often in Latin America than in Spain, and that it has a positive value showing *confianza* 'trust' and *calor humano* 'human warmth'. The use of *Anita* is thus reframed as an intercultural misunderstanding by the President of Ecuador.

Meanwhile, Ana Pastor is quoted as having felt uneasy, but not offended. During a visit to a Master's course offered by the newspaper *ABC* four days after the interview, she is cited by the paper as "finding it difficult to believe that Correa would have attempted this closeness with a man. I cannot imagine him calling Mariano Rajoy – whom he had met earlier during the day – 'Marianito'" (López Trujillo 2012).[8] Here Pastor argues, firstly, that a gender differential is involved; that closeness and tenderness would not have been seen as appropriate in a professional encounter between two men. Secondly, she also introduces a power differential; while Correa would not have used the diminutive in a meeting with another president, he portrays it as normal in an interview with a journalist. In this latter situation, however, Spanish communicative norms would demand a symmetrical form of address, or perhaps one recurring to the use of first names in their full forms.

It would seem that both interlocutors were very conscious of the interview's public reverberations; they wanted to downplay the 'Anita incident' and rather concentrate on the political side of the interview. In their roles as interviewer and president, they met again on two occasions, in May 2013 and May 2014, on Ana Pastor's new program *Frente a frente* on CNN en español. In the first of these subsequent interviews, Pastor says "*siempre es un placer escucharle*" 'it's always a pleasure to listen to you' (CNN en español 2013). While this may have been formulaic politeness, the interview shows signs of both interlocutors knowing better what to expect from the other. Cuenca (2013: 545), in a postscript to her article, notes that there are far fewer vocatives (six instances of *Ana*, none of *Anita*, and ten of *presidente*, mostly in turn-initial position), as well as far less overlapping of turns and fighting for the floor, although in the later, more heated part of the interview each reprimands the other for interrupting them. The final interview, in May 2014 (CNN en espanol 2014b), is interesting since, two days before the broadcast, CNN en español – possibly in an attempt to attract a bigger audience – posted a video called *Detrás de cámaras: Rafael Correa y Ana Pastor* (CNN en español 2014a). It gives a glimpse of the preparations for the interview. Correa, at min 1:13, can be heard saying (while seeming to look into the camera) "*tengo que concentrarme a decirle (.) ANA (.) se me va a quitar eso?*" ('I must concentrate on calling you (.)

8. "*Me cuesta creer que hubiese tenido esa cercanía con un hombre. No me imagino a Correa llamando a Mariano Rajoy – con quien el presidente ecuatoriano acababa de reunirse – 'Marianito'*", ha afirmado Pastor.

ANA (.) will I ever learn to do that?'). Ana Pastor and the surrounding people briefly laugh at this remark, and Pastor responds "*Qué va, si ya somos viejos conocidos*" ('Oh well, we're old friends, aren't we?').

6. Public reaction to the interview, 2012–2018

The public reframing of the *Anita* incident is of particular importance as the interview continues to attract comments even now, with repercussions for the public status accorded to both Pastor and Correa. Many comments have been made on social media, especially on the internet platform YouTube. The most commented YouTube video of the interview itself was uploaded on the day of the interview by a user called *Sucreranda Hugo Chávez Venezuela*, apparently an admirer of late Venezuelan President Hugo Chávez. Their YouTube channel is very influential, and this particular video has been seen over a million times; up to July 27, 2018, when the URL was last verified by the author, it had scored 1,004,197 hits, with 5,122 'thumbs up' reactions and 635 'thumbs down'. By the same date, other users had left almost 7,000 comments on the video, some of which had themselves received further comments. The flow of comments has been very steady from the day of posting. The same video was then also posted on March 25, 2012, by 'IUAVILA1' (2012), where IU denotes Izquierda Unida 'United Left', a Spanish political party (as of June 2018 it had recorded 300,000 views and over a thousand comments) and by a user called *Santiago Aguirre* (2012) on November 19, 2012 (over 41,000 views and 117 comments).

Most comments on the videos take sides either for Correa or Pastor, often in a very general way but frequently also in obvious reflection of a particular political stance. For instance, *Sucreranda Hugo Chávez Venezuela* and *IUAVILA1* gave their videos the same title: *Correa vapulea a Ana Pastor* 'Correa trashes Ana Pastor'. In doing so, they make clear their perception that Correa 'won' the interview, as well as revealing their own political convictions. Most users who post a comment share the view of Correa as the 'winner'. Many mention their national origin, e.g. "*Ojalá algún día en España tengamos un presidente así*", "*Saludos desde Colombia, tengo envidia de la Buena que ustedes tengan un presidente tan bueno*", "*Soy peruano y admirador de Rafael Correa*."[9] A large number of users place the incident in a larger communication pattern of Latin Americans being discriminated against by Spaniards, accusing Ana Pastor of a Eurocentric perspective, of being right-wing

9. 'I just hope that one day we'll have a president like him in Spain', 'Greetings from Colombia, I am envious (in a good way) that you have such a good president', 'I'm Peruvian and an admirer of Rafael Correa.'

and anti-Chávez, or of being ill-prepared. There are also users (with both male and female user names) who see this more as a pattern of *machismo*.

For lack of space, it is impossible to offer a detailed analysis of the entire discussion. In any case, our main concern here is the meaning attributed to the diminutive *Anita*. I therefore collected a small subsample (*N* = 75) of the 7,000 comments left on the video posted by *Sucreranda Hugo Chávez Venezuela*, consisting of those that mentioned either *Ana* or *Anita* or the word *diminutivo* 'diminutive'. *Ana* / *Anita* in most cases are used as a term of reference, but there are also instances where *Anita* is used as a term of address (purportedly to the journalist herself). The subsample was then culled of simple insults, such as the following posted in 2016 by *Geronimo*: "*jajajjaj Mi querida Anita jajajjaja :) Rafael Correa es un lider innato :)*" ('hahahah My dear Anita hahahah :) Rafael Correa is a natural leader :)'). The comments picked were then broadly spread over the six-year time span, with seven to thirteen per year.

As was to be expected, more comments in the subsample were posted by Latin Americans than Spaniards, at least among those users who categorize themselves by giving their nationality/domicile, by their adoption of a nickname indicating a certain national or regional identity, or by their use of personal pronouns for plural address (2nd person plural in the case of Spaniards, 3rd person plural in the case of Latin Americans). Fourteen users either self-align as Ecuadorians or as someone living in Ecuador (*acá en Ecuador* 'here in Ecuador'); fifteen users do something likewise on a Latin American level or name another specific country of origin; five can be identified as Spaniards. Forty-one user names suggest a male identity, and six users are probably female, while the remaining names chosen give no indication as to gender.

Of the 75 comments, 52 indicate that that the diminutive is used in a tender and affectionate manner in Latin America, implying that Correa used it this way throughout the interview. In a sense, the majority of Latin Americans share the view: 'What's the problem? It's a term of endearment!' Example (4) is a good representative of this sort of comment.

(4) *en América latina anita o pedrito …etc no es falta de respeto es una apreciación . no se como lo vean en España ,,, debió tener cuidado correa . pero el entrevistador , y estos opinologos deben aprender a diferenciar la diversidad cultural … el castellano peruano ejemplo no es similar a Un castellano venezolano , o mexicano.* (User *no a keyko*, 2018)
'In Latin America Anita or Pedrito … etc is not a lack of respect, it's a sign of appreciation. I don't know how you see it in Spain … Correa should have been more careful. But the interviewer and these self-appointed experts

must begin to understand cultural diversity … Peruvian Spanish, for example, is different from Venezuelan, or Mexican.'[10]

Here, the key term is *diversidad cultural*, 'cultural diversity'; users are asserting the right of a Latin American, even when interviewed in Spain and by a Spanish journalist, not to accommodate their use to the linguistic norms of *la madre patria*, 'the Spanish motherland'.

Fifteen users consider the diminutive as a sign of *machismo*, at least in this particular interview situation; mostly female, they tend also to be Spaniards. Some suggest, as Pastor herself had already done, imagining the interview with the roles reversed.

(5) *¿Y si a él le hubieran tratado de "rafaelito"? En este escenario el uso del diminutivo no es en absoluto una expresión de cariño sino de desdén. […]* (User *kentauros54*, 2014)
'What if she had called him "Rafaelito"? In this setting the use of the diminutive is absolutely not a sign of tenderness but of disdain.'

(6) *si fuera un hombre, le diría pepito? un respeto, que es una periodista seria y no una cría a la que mostrar "cariño"* (User *nacamo17*, 2013)
'If she were a man, would he call her Pepito? More respect: she's a serious journalist and not a little girl to whom one can show "tenderness".'

In Example (5), Pastor is envisioned calling Correa 'Rafaelito'. The user *kentauros54* argues that this would show disdain, not tenderness, and thus that the diminutive should be avoided 'in this setting' (probably meaning the interview setting). In Example (6), user *nacamo17* imagines an interview between Correa and a male journalist José, the diminutive of which, Pepito, they argue would very likely not be used – once again suggesting that use of the diminutive must be interpreted as a lack of respect. This argument resembles Zuluaga's (1970) observation that one of the diminutive's main functions is condescension, in this case for a female journalist, who is after all a journalist, not a little girl. The professional interview setting is thus seen as decisive for the opinion that Correa ought to have chosen a more respectful term of address, no matter the argument of cultural diversity.

The professional identities put on display by Correa and Pastor during the interview are barely discussed. Where they are, as in Example (7), this often leads to attacks on Pastor's self-conception as a journalist. She is judged to be unprofessional for letting her personal feelings interfere with her work.

10. The user's name *no a keyko* indexes them as Peruvian since it refers to Keiko Fujimori, politician, presidential candidate in 2011 and 2016, and daughter of former President Alberto Fujimori.

(7) *ANITA ANITA ANITA………. Ridicula…! ¿Como se puede decir que ella es profesional?, si en plena entrevista internacional, le dice me llamo Ana Pastor, por favor… ahí está mezclando sus sentimientos o gustos, con el trabajo… Eso no es ser profesional… ¿Cómo puede ser posible?, toda España se enteró que a ella no le gusta que le llamen Anita… Pero aún así el presidente Rafael Correa se portó muy respetuoso… Simplemente el presidente le hizo quedar mal..! Correa es más inteligente..!* (User *Guido Armas*, 2012)
'ANITA ANITA ANITA… ridiculous…! How can they say she's a professional? To say, in the middle of an internationally televised interview, I'm called Ana Pastor, please… there, she's mixing up her feelings or her personal preferences with her work… That's not being professional… How is that possible? All of Spain now knows that she doesn't like to be called Anita… But nevertheless President Rafael Correa behaved very respectfully… The president simply made her look bad…! Correa is more intelligent…!'

In a similar vein, the user *Guido Armas* accuses Pastor of having made a fool of herself by insisting, 'in the middle of an internationally televised interview', on her 'real' first name. In fact, the persistence with which this interview receives deprecating comments suggests that many users – Spaniards, but even more so Latin American commentators – prefer a more neutral, less prominent role for a journalist, as is also palpable in the subsequent assertion that Pastor's emotions interfered with her professional role.

Meanwhile, Example (8) points out the professionalism of both interlocutors, even if the user *María Marta y Luisa* is not in agreement with Pastor's political opinions.

(8) *Es el estilo habitual de entrevista de Ana Pastor. Da igual que sea de un partido político o del contrario. Cree hacer un periodismo valiente. Pero siempre detrás hay manipulación y falacia que pueden llevar al engaño a gente no informada que ve sus entrevistas. Correa sabe defenderse de ese tipo de periodismo. Bien hecho! desde Madrid.* (User *María Marta y Luisa*, 2016)
'It's Ana Pastor's normal interviewing style. It doesn't matter if someone is from one political party or the other. She thinks she's being a courageous journalist. But behind it all there's always manipulation and lies that might deceive ill-informed people who watch her interviews. Correa knows how to defend himself against that type of journalism. Well done! From Madrid.'

María Marta y Luisa compares Pastor's style in this interview with other interviews conducted by her and points out that Pastor is a tough interviewer for all politicians – unlike other journalists (see my remark in footnote 3 on TVE's reputation for catering to the ruling party). However, the user also recognizes manipulative elements behind this tough stance, although unfortunately they do not give examples.

In a sense, the entire discussion of whether or not a diminutive of a journalist's first name is indicative of *machismo* must be seen in the broader context of the long-term self-image claimed by both interlocutors, as well as public expectations of how a journalist should behave towards his or her interviewee. The non-Spanish spectators of the event perceive Pastor as too critical and too condescending; her critical questions seem to irritate many commentators, most, but not all from a Latin American background. It would be interesting to compare the interviews with Correa to others carried out by Pastor, and also to investigate whether the public's perception has changed since she switched to CNN en español, which tends to favor the more assertive U.S. manner of questioning in political interviews.

7. Conclusion

This contribution has attempted to trace the 'Anita incident' from the first interview on Spanish television in 2012, through the next interviews on CNN en español, to the reverberations of the interview on social media, using the example of comments on one YouTube video. So, what's in a diminutive? I have tried to show that the first instance of *Anita* in the introduction to the interview was most likely a product of differing national norms and, in that sense, a cultural misunderstanding. By contrast, the last occurrence was clearly intended to provoke, while the second to seventh instances probably also contained a provocative element. Nonetheless, they are framed in such a way that Correa was able to excuse his use as a *costumbre latina* 'Latino custom'.

The use of the diminutive *Anita* instead of *Ana*, while it admittedly made the journalist uneasy, did not challenge her professional identity. Neither Correa nor Pastor seems to have had any bad feelings about the other after the first interview; instead, they appear to have shrugged off the incident as a rather localized power struggle and been astonished by the global reaction. Indeed, the interviews with Correa advanced Pastor's global reputation in the sense that they extended her profile beyond Spain. If she is ridiculed and/or vilified by many YouTube users, especially those indexing a Latin American origin, this is more often than not due to their political preferences, and possibly to having particular expectations about the genre of the mediatized political interview.

As we have seen, the values associated with the diminutive differ considerably between Latin American and Spanish speakers of Spanish. The vast majority of Latin American speakers see the diminutive as a sign of affection, while Spanish users tend to be more aware of its potential for condescension, especially if used asymmetrically, in public and in a situation of a male-female interaction. In this particular situation of an internationally viewed political interview, the use of *Anita* diverged from the conventions normally underlying the interchange between interviewer and interviewee, to the point that the journalist considered her long-term face needs under threat. This overrode her local obligations as an interviewer. Also, we need to consider that for Pastor, the main audience was the Spanish public, while Correa was addressing a more global, more diverse audience. In the end, their 'conflict' did not damage them. On the contrary, both seem to have profited from the public attention accorded to the interview.

Acknowledgements

An earlier version of this study was presented at the Workshop *Political discourse in the Romance-speaking countries* (Stockholm, October 2014). I am indebted to those in the audience who provided comments. I am particularly grateful to my former students Katarina Bösner, Mareike Busch, Mayra Condemarín Cellio, Melissa Herrera, and Rebecca Reiber, who analyzed the 'Anita incident' in a course on communicative strategies in the mass media during winter term 2013/14 at Hildesheim University. I would also like to thank Pia Roser for her aid in constructing a corpus of online commentaries and their subsequent annotation. Naturally, all interpretations and any mistakes remaining are my own.

References

Blas Arroyo, José Luis. 2000. "Mire usted Sr. González … Personal Deixis in Spanish Political-electoral Debate." *Journal of Pragmatics* 32 (1): 1–27. https://doi.org/10.1016/S0378-2166(99)00040-5

Blas Arroyo, José Luis. 2003. "'Perdóneme que se lo diga, pero vuelve usted a faltar a la verdad, señor González': Form and Function of Politic Verbal Behaviour in Face-to-Face Spanish Political Debates." *Discourse & Society* 14 (4): 395–423. https://doi.org/10.1177/0957926503014004001

Boletin Oficial de España. 2006. Ley de la Radio y la Televisión Estatal 17/2006. Available online at https://www.boe.es/buscar/pdf/2006/BOE-A-2006-9958-consolidado.pdf (accessed 15 April 2019).

Brown, Penelope, and Stephen Levinson. 1987. *Politeness: Some Universals in Language Usage*. Cambridge: Cambridge University Press. https://doi.org/10.1017/CBO9780511813085

Clayman, Steven. 2010. "Address Terms in the Service of Other Actions: The Case of News Interview Talk." *Discourse and Communication* 4: 161–183. https://doi.org/10.1177/1750481310364330

Cuenca, María Josep. 2013. "Uso del vocativo en la entrevista política: género discursivo y (des) cortesía." *Discurso & Sociedad* 7 (3): 522–552.

Goffman, Erving. 1959. *The Presentation of Self in Everyday Life*. New York: Anchor Books.

Heritage, John, and Steven Clayman. 2010. *Talk in Action: Interactions, Identities and Institutions*. Oxford: Blackwell. https://doi.org/10.1002/9781444318135
Hummel, Martin. 1995. "Diminutive als Apreziativa. Zur Theorie der Diminutive im Spanischen." *Romanistisches Jahrbuch* 45: 243–261.
Hummel, Martin. 1997. "Para la lingüística de vuestro diminutivo: Los diminutivos como apreciativos." *Anuario de Estudios Filológicos* 20: 191–210.
Jaworski, Adam, and Dariusz Galasiński. 2000. "Vocative Address Forms and Ideological Legitimization in Political Debates." *Discourse Studies* 2 (1): 35–53. https://doi.org/10.1177/1461445600002001002
Kerbrat-Orecchioni, Catherine (ed). 2010. *S'adresser à autrui. Les formes nominales d'adresse en français*. Chambery: Université de Savoie.
Placencia, María Elena. 2010. "El estudio de formas de tratamiento en Colombia y Ecuador." In *Formas y fórmulas de tratamiento en el mundo hispánico*, ed. by Martin Hummel, Bettina Kluge, and María Eugenia Vázquez Laslop, 348–383. Mexico-City: El Colegio de México.
Rendle-Short, Johanna. 2007. "'Catherine, You're Wasting Your Time': Address Terms Within the Australian Political Interview." *Journal of Pragmatics* 39 (9): 1503–1525. https://doi.org/10.1016/j.pragma.2007.02.006
Rendle-Short, Johanna. 2011. "Address Terms in the Australian Political News Interview." In *Talking Politics in Broadcast Media: Cross-cultural Perspectives on Political Interviewing, Journalism and Accountability*, ed. by Mats Ekström and Marianna Patrona, 93–111. Amsterdam: John Benjamins. https://doi.org/10.1075/dapsac.42.09ren
Selting, Margret, Peter Auer, Dagmar Barth-Weingarten, Jörg Bergmann, Pia Bergmann, Karin Birkner, Elizabeth Couper-Kuhlen, Arnulf Deppermann, Peter Gilles, Susanne Günthner, Martin Hartung, Friederike Kern, Christine Mertzlufft, Christian Meyer, Miriam Morek, Frank Oberzaucher, Jörg Peters, Uta Quasthoff, Wilfried Schütte, Anja Stukenbrock, and Susanne Uhmann. 2009. "Gesprächsanalytisches Transkriptionssystem 2 (GAT 2)." *Gesprächsforschung – Online-Zeitschrift zur verbalen Interaktion* 10: 353–400. http://www.gespraechsforschung-ozs.de/heft2009/px-gat2.pdf
Travis, Catherine. 2005. "The Ethnopragmatics of the Diminutive in Conversational Colombian Spanish." *Intercultural Pragmatics* 1 (2): 249–274. https://doi.org/10.1515/iprg.2004.1.2.249
Vázquez Laslop, María Eugenia. In press. "Some Pragmatic and Grammatical Categories for the Analysis of Forms of Address in Presidential-election Debates." In *It's Not All about* You: *New Perspectives on Address Research*, ed. by Bettina Kluge, and María Irene Moyna, with the assistance of Horst Simon and Jane Warren. Amsterdam: John Benjamins.
Zuluaga Ospina, Alberto. 1970. "La función del diminutivo en español." *Thesaurus* 25 (1): 23–48.

Journalistic sources

Buck, Tobias. 2015. "Spanish State Broadcaster TVE Accused of Political Bias." *Financial Times*, April 27, 2015. Available online at https://www.ft.com/content/4625b188-e818-11e4-9960-00144feab7de (accessed 15 April 2019).
Burgen, Stephen. 2012. "Spanish Government Accused of Purging Critics from National Radio and TV." *The Guardian*, August 5, 2012. Available online at https://www.theguardian.com/world/2012/aug/05/spanish-government-critics-national-tv (accessed 15 April 2019).

CNN en español. 2012. "La periodista española Ana Pastor se suma a CNN." *CNN español*, September 11, 2012. Available online at https://cnnespanol.cnn.com/2012/09/11/la-periodista-espanola-ana-pastor-se-suma-a-cnn/ (accessed 15 April 2019).

Correa, Rafael (@MashiRafael). 2012. "@MarioTroya No recuerdo, pero nos entendimos muy bien. Fue una entrevista muy agradable. Anita (costumbre latina) es muy profesional." *Tweet*, March 19, 2012. https://twitter.com/MashiRafael/status/181939642562646016 (accessed 15 April 2019).

López Trujillo, Noemí. 2012. "Ana Pastor: Con Rafael Correa me sentí incómoda, no ofendida." *ABC.es*, March 23, 2012. Available online at https://www.abc.es/20120320/medios-redes/abci-pastor-201203202253.html (accessed 15 April 2019).

Trueba, David. 2018. "De negro. Al que no le guste TVE, que cambie de canal. Ese es el descaro inmoral que rige entre los políticos españoles." *El País*, May 22, 2018. Available online at https://elpais.com/elpais/2018/05/21/opinion/1526904701_430261.html (accessed 15 April 2019).

YouTube videos and TVE website

CNN en español. 2013. Rafael Correa "Frente a frente" con Ana Pastor. Part 1. September 28, 2013. https://www.youtube.com/watch?v=pE2UeKAQQQ4 (accessed 25 June, 2019)

CNN en español. 2014a. Detrás de cámaras mostrando los preparativos para la entrevista. May 2, 2014. http://www.youtube.com/watch?v=YEBEWg-O_nc (accessed 15 April 2019)

CNN en español. 2014b. Ana Pastor entrevista a Rafael Correa. May 4, 2014. https://edition.cnn.com/videos/spanish/2014/05/05/cnnee-frente-a-frente-pastor-correa-1.cnn and https://edition.cnn.com/videos/spanish/2014/05/05/cnnee-frente-a-frente-pastor-correa-2.cnn (accessed 15 April 2019)

Izquierda Unida Avila (IUAVILA1). 2012. "Rafael Correa vapulea a Ana Pastor." YouTube video, March 25, 2012. http://www.youtube.com/watch?v=BlSfLjzBn6s (accessed 15 April 2019)

Los desayunos de TVE. 2012. Los desayunos de TVE – Rafael Correa, presidente de Ecuador. Archived version of the morning show. March 19, 2012. http://www.rtve.es/alacarta/videos/los-desayunos-de-tve/desayunos-tve-rafael-correa-presidente-ecuador/1352796/#aHR0cDovL3d3dy5ydHZlLmVzL2FsYWNhcnRhL2ludGVybm8vY29udGVudHRhYmxlLnNodG1sP3BicT0yJm1vbnRoRmlsdGVyPTMmb3JkZXJDcml0ZXJpYT1ERVNDJm1vZGw9VE9DJnllYXJGaWx0ZXI9MjAxMiZsb2NhbGU9ZXMmcGFnZVNpemU9MTUmY3R4PTE2NTUmYWR2U2VhcmNoT3Blbj1mYWxzZQ== (accessed 15 April 2019).

Santiago Aguirre. 2012. "Entrevista a Rafael Correa por Ana Pastor [TVE, 19/03/2012]." YouTube video, November 19, 2012. https://www.youtube.com/watch?v=sEdLqi2wMC0 (accessed 31 July 2018)

Sucreranda Hugo Chávez Venezuela. 2012. "Rafael Correa vapulea a Ana Pastor." YouTube video, March 19, 2012. https://www.youtube.com/watch?v=zMQEvZ4itoo (accessed 15 April 2019)

From Poland to #SanEscobar

On strategies subverting political discourse on Twitter

Nadine Thielemann
Vienna University of Economics and Business

Using the example of the hashtag #SanEscobar, which was triggered by a slip of the tongue made by the then Polish Minister of Foreign Affairs Witold Waszczykowski, the chapter shows how Twitter provides a discourse arena for the subversion of political discourse. A multimodal discourse analysis of tweets reveals how participants jointly devise the virtual country of San Escobar. The conceptual structure of the unfolding discourse uses elements from frames which comprise knowledge about South-America, Pablo Escobar and about the national conservative PiS party. Conceptual elements and structure from these frames are combined in a way that associates PiS-governed Poland with the stereotype of a South American narco-democracy (San Escobar) and thus humorously criticises the current political situation in Poland.

Keywords: multimodal discourse analysis, frames, Grounded Theory, social media communication, Polish, political discourse

1. Introduction

The chapter focuses on Twitter as a discursive arena for a *counter-public* (Fraser 1990) in the context of the worsening political situation and societal climate in Poland. This is characterised by a strong and increasing antagonism between a national conservative camp headed by the *Law and Justice* party (*Prawo i Sprawiedliwość* – PiS) and a somewhat more liberal camp. The former has been constantly gaining influence since the noughties, especially after winning the parliamentary and presidential elections in 2015. Meanwhile, the latter has become rather weakly represented on the political stage due to substantial losses in those elections.

https://doi.org/10.1075/pbns.306.10thi

The government formed by PiS in 2015 has been seeking to use its newly heightened political power to strengthen its influence on various Polish institutions, thus reinforcing the impact of its policies on society. Among the first measures taken in this direction was a law granting the government the right to make appointments to key positions in the public media. Subsequently, several journalists and presenters of leading programmes on Polish public-service broadcasting were dismissed or left for private stations of their own accord. Another controversial law passed in 2017 allows the government to exercise control over the constitutional court. Meanwhile, several new, PiS-affiliated constitutional judges have been appointed. These actions, which substantially undermine the separation of powers (including the media as the *Fourth Estate*), have triggered protests and rallies on the streets of major Polish cities.

Given the weakness of the parliamentary opposition, which is still suffering from some minor political scandals during the prior parliamentary term, the main organiser of these actions has been the *Committee for the Defence of Democracy* (*Komitet Obrony Demokracji* – KOD). This extra-parliamentary, oppositional civic movement, which is not linked with any specific party, is generally playing an important role in public criticism of the national conservative government's political actions and the main organiser of rallies and protests. Owing to PiS's rising influence on the publicly financed media and its dominant role within political institutions, however, alternative positions and oppositional views are accorded limited media coverage and thus have limited opportunities to spread to a wider audience.

In this context, Twitter provides a discourse arena in which users critical of PiS, and of its political and societal projects, can promote their views and mutually affiliate in an *online* public sphere (cf. Choi and Cho 2017). Fraser describes this societal capacity (then still *offline*) in terms of *counter-publics*. These she defines as "parallel discursive arenas where members of subordinated social groups invent and circulate counterdiscourses, which in turn permit them to formulate oppositional interpretations of their identities, interests, and needs" (1990: 67). The interactive and participatory character of social media such as Twitter not only represents the advancement of Web 2.0 (cf. Herring 2013). It also equips social groups and civic movements with an easily accessible tool to express their discontent about what is going on in the mainstream media and/or on the institutionalised political stage. Moreover, Twitter allows them to publicly display (at least on the Internet) their "collaboration and networking […] and a sense of belonging and connectivity" (Choi and Cho 2017: 15) and thus shares essential features with counter-publics as stressed by Choi and Cho. This chapter will show how *counter-discourses* can emerge on Twitter in reaction to offline events on the political stage, and how users online can jointly construe alternative conceptualisations which humorously undermine public discourse in the mainstream media and on the political stage.

1.1 San Escobar as an Internet phenomenon

Specifically, this chapter analyses the discourse labelled by the hashtag #SanEscobar, which emerged, and quickly started trending, in January 2017 in reaction to a slip of the tongue by the then Polish Minister of Foreign Affairs, Witold Waszczykowski (PiS). This was Waszczykowski's report, made in the context of Poland's efforts to be elected to the UN Security Council, of successful talks 'with countries such as Belize and San Escobar' (*z takimi krajami jak San Escobar albo Belize*), when he meant to refer to the Caribbean island state of St. Nevis & Kitts (http://www.polsatnews.pl/wiadomosc/2017-01-10/san-escobar-wpadka-ministra-waszczykowskiego-mowi-o-spotkaniu-z-ministrem-nieistniejacego-panstwa/).[1] On 10 January 2017, this mishap received a brief mention on Polish TV, which marked the beginning of San Escobar's rise as an Internet phenomenon.

Among other actions, users created social media accounts (e.g. on Twitter or Facebook) or memes (e.g. image macros or demotivators) featuring San Escobar, thus starting a process of joint online creativity in which a virtual country took shape and went viral. Furthermore, San Escobar's rising popularity on the Internet attracted the attention of offline media, which then reported about the Internet poking fun at Waszczykowski (on convergence in the context of participatory online-media cf. Fraas et al. 2010 and Androutsopoulos 2010). Shortly afterwards, several advertising campaigns also drew on the idea of San Escobar as an exotic country and holiday destination (cf. Issel-Dombert 2017). A fast food chain, for example, promoted Mexican flavoured pizzas, a whisky company advertised its product served in a glass with a cocktail umbrella and exotic fruits under palm trees, and a taxi app launched a sales discount for clients going to the airport for their holidays and included #SanEscobar in their online order.

The discourse on Twitter in reference to San Escobar represents a clearly distinct extract of this Internet phenomenon. It first appeared on 10 January in reaction to the TV broadcast of Waszczykowski's slip, reached its activity peak on 12 January, continued running at a lower level of activity until summer 2017 and is still active at the time of writing (https://www.signs.pl/artykul.php?sid=31468&fo=1). The users contributing to this hashtag on Twitter jointly devised the virtual country of San Escobar. In their joint conceptualisation, it resembles a stereotypical South American *narco-democracy*, that is, a country whose economic, social, judicial and political situation is profoundly affected by drug traffickers – as was the case of Colombia during the times of the drug baron Pablo Escobar. Topics and issues from Polish politics are interwoven in this discourse in a way that mostly ridicules PiS

1. Note that all links mentioned in this chapter were last accessed in April 2019, unless otherwise indicated.

politicians, their ideological positions and their political projects. The contributors thus quite clearly display their affiliation with oppositional views to a wider online public. In this context, the hashtag serves to express users' "ambient affiliation" and facilitates the construction of a community as stressed by Zappavigna (2011: 83–99).

1.2 Research questions

This chapter pursues two main research questions. How can we analyse the emergent conceptualisation of Twitter discourses? And how can we adequately detect their specific evaluative and emotive tone (cf. Evans 2016)? In doing so, it employs an analytic approach that combines ideas and concepts from Cognitive Linguistics with an explorative coding procedure in order to determine and describe the emergent conceptualisation (cf. Fraas et al. 2010; Fraas 2013; Fraas and Pentzold 2015). Using the example of the #SanEscobar discourse, I show how the creative online conceptualisation relies on discourse-specific elaborations and combinations of frames which organise the multimodal content of the discourse (cf. Ziem 2008a and b; Fraas 2011). As the analysis shows, it is the combination and contiguity of otherwise disparate frames that lends the discourse a humorous tone, which enables users to subvert the political positions and projects of the national conservative government. By participating in this process of collaborative conceptualisation, users affiliate with the emerging collective stance of the discourse and identify with the virtual group, or discourse community, critical of PiS. In other words, they express their belonging, while attributing emotive and evaluative significance to their membership, thus enacting their ingroup identity (cf. Tajfel 1979).

Identity politics, however, is at work in several ways (cf. Bucholtz and Hall 2005; Bös et al. 2018). By using the hashtag #SanEscobar, users enact their ingroup identity, in which joint opposition towards and ridicule of the PiS government and its political projects play a key role. In the emerging discourse, they jointly create the fake country (cf. Kirner-Ludwig 2018 on the online construction of fictional identities) of San Escobar whose identity draws strongly on elements of South American culture and geography and on the political and socio-economic features of a narco-democracy. In the process of the joint conceptualisation of this fake country, many implicit and explicit comparisons to Poland are drawn. This includes a discursively shaped identity of the leader of San Escobar that is in multiple ways reminiscent of Jarosław Kaczyński, the leader of PiS, on the one hand, and of Pablo Escobar or a stylised South American drug lord, on the other. The analysis thus also aims at revealing how the hybrid identity of San Escobar and its leader emerges in the unfolding social media discourse and at uncovering the linguistic and semiotic practices and resources which contribute to their construction (cf. Bucholtz and Hall 2005: 588).

1.3 Structure of the chapter

The remainder of the chapter is structured as follows. First, in Section 2, I address issues of data collection. In view of the multimodal character of social media communication, I adopt an on-screen mode of data collection and a qualitative approach to analysis. In Section 3, the framework of the multimodal discourse analysis is introduced. It relies on a frame-analytic account to describe the discourse-specific conceptualisation (cf. Ziem 2008a and b) and on a preceding, computer-assisted coding procedure which allows for bottom-up identification of the frames employed. Then, in Section 4, I present the results of this methodological procedure, that is, the conceptualisation of San Escobar in terms of a specifically configured [COUNTRY]-frame which organises the multimodal content accessible via the hashtag #SanEscobar. I conclude by summarising the benefits of this approach for the analysis of emerging discourses on social media which lend the counter-public a voice.

2. Analysing hashtags and collecting data

Twitter is a micro-blogging service which allows users to post, reply to or forward (*retweet*) messages of up to 140 (now also 280) characters, which may include graphics or video material (cf. Halavais 2014; Burghardt 2015). Users rely on hashtags (#) in order to mark a message (*tweet* or *retweet*) as a contribution to a specific topic emerging on Twitter. Hashtags such as #SanEscobar thus function as "ideational labels" (Zappavigna 2011: 91) which "represent a way of indicating textually keywords or phrases especially worth indexing" (Halavais 2014: 36), and by means of which "Twitter users communicate a desire to share particular keywords *folksonomically*" (ibid., italics added). This turns hashtagging into a form of social sharing (social tagging, collaborative tagging). Summing up, hashtags have a threefold function, "an ideational function (indicating a topic), an interpersonal function (construing a relationship), and a textual function (aggregating and organizing discourse)" (Zappavigna 2018: 12). By this very quality, hashtags resemble *discursive keywords*. These, too, emerge in the context of a debate and function as labels with the potential, whenever they are mentioned, to immediately evoke recipients' knowledge of and about the discourse in question (cf. Liebert 2003). Consequently, a hashtag marks a point of access to a given discourse and can thus serve as a starting point for a data collection strategy of "sampling by theme" (Androutsopoulos 2013: 238).

Social media such as Twitter are "discursively created spaces of human interaction, which are dynamically related to offline events" (Androutsopoulos

2013: 239). They provide sites for an *emerging semiotic network of meaning making* (Zappavigna 2011) through which a social online community constitutes itself. And, indeed, contributors to the hashtag #SanEscobar do rely on a multitude of semiotic resources. These range from linguistic input drawn from several languages (including tweets which mix languages) to pictorial elements (e.g. photos, image macros, caricatures) and videos such as GIFs, music or film clips (on the multimodal nature of social media communication cf. Siever 2015: 243–311). Androutsopoulos coined the term *spectacle* in order to describe the multimodal character of social media communication and the way in which users participate (i.e. contribute and receive).

> [T]he concept is meant to encompass other types of digital content such as music and photography, which may not involve language at all. The metaphor suggests that these spectacles are displayed to an audience; are viewed rather then read; are mainly perceived and consumed as entertainment; and prompt responses, which are usually expressed in comments or video responses.
>
> (Androutsopoulos 2010: 209)

Such a perspective on social media communication has consequences for both the data to be collected and the methodological approach required to analyse it. In order to include as many of the multimodal resources contributing to the conceptualisation of San Escobar on Twitter as possible, I created a screenshot of the scrolling screen (https://twitter.com/hashtag/sanescobar?lang=de using FireShot Pro). Although screenshots do not preserve some formats (videos, for example, have to be accessed separately), they remain the best way of accessing and thus analysing what the Twitter users themselves see. In order to back the multimodal analysis, my following discussions include both direct quotes from tweets and examples of images cited in unedited form. I here follow Hardaker and McGlashan (2016), who choose not to anonymise tweets, since Twitter in its privacy policy explicitly informs its users that the posted content as well as other information such as name/username, location etc. are instantly disseminated to a broad public. Hardaker and McGlashan (2016: 86) further stress that "anonymising accounts would contravene Twitter's policies on displaying their data in static publications", which should show the unmodified text. In view of time and space constraints, I focus on contributions made during the first three days of the hashtag's life. Since #SanEscobar passed its peak activity after two days, this period covers the most intense period of tweeting. The data analysed comprise almost 2000 contributions included in the screenshot, with few users contributing more than once.

In view of the multilingual and multimodal characteristics of the contributions, I employ a qualitative methodology because that enables me to better assess the impact of these diverse resources on the conceptualisation of San Escobar. Moreover, qualitative approaches cope better with discourses, like that of #SanEscobar, characterised by creativity and *non bona fide* communication (cf. Klemm and Michel 2014). In the next subsection, I introduce the methodological framework underlying the multimodal discourse analysis in the main section of this chapter.

3. Analytic framework and methodology

In order to analyse the *emerging semiotic network of meaning making* (Zappavigna 2011) on Twitter in which the virtual country of San Escobar is jointly conceptualised and in which its distinct identity emerges, I adopt an approach which resembles the one applied by Fraas and Pentzold (2008, 2015) and Fraas (2013) for convergent multimodal and multimedia discourses. This analytic framework essentially draws on ideas and concepts from Cognitive Linguistics, such as frames and their role in conceptualisation, and uses a bottom-up coding procedure in order to identify empirically the frames structuring the conceptualisation. I therefore begin by discussing the concepts of frames and framing, and showing how they can be employed in the analysis of discourse-specific conceptualisations, before going on to introduce the methodological procedure employed for the identification of frames.

3.1 Frames and frame-based discourse analysis

The concept of the frame emerged in the Cognitive Sciences as a model used to explain how our collective knowledge of the world is organised, stored and represented. It quickly made its way to linguistics, where it explains why our conceptualisation is mostly richer than the linguistic input triggering it (cf. Bednarek 2005; Busse 2012). Media and communication studies have adopted the concept and reinterpreted it in terms of *framing*. Their understanding stresses that frames lend a presentation a specific perspective, as succinctly explained by Entman:

> To frame is to *select some aspects of a perceived reality and make them more salient in a communicating text, in such a way as to promote a particular problem definition, causal interpretation, moral evaluation, and/or treatment recommendation* for the item described. (Entman 1993: 52; italics in original)

Fraas et al. (2010: 237) draw on the two ideas of frames and framing in promoting their understanding of frames as models for the contextualisation of concepts which play a crucial role in a given discourse ("Frames als Modelle für das Kontextualisierungspotential diskursiv zentraler Konzepte (Schlüsselkonzepte)"). Similarly, Ziem (2008a and b) determines basic discourse-semantic figures (*diskurssemantische Grundfiguren*) in terms of specifically configured frames and frame-complexes. Although the terminology they use to refer to discourse-specific conceptualisations differs slightly (*basic discourse-semantic figure* by Ziem vs. *discursive key concept* by Fraas et al.), both studies arrive at a similar discourse-analytic operationalisation of the frame-concept. According to both lines of research, frames organise the multimodal content of a given discourse. They thus play a triple role: as a cognitive format of knowledge representation, as an analytic tool and as a discourse-specific interpretational pattern ("1. Analytische Werkzeuge, 2. Kognitive Repräsentationsformate und 3. Diskursive Interpretationsmuster") (Fraas et al. 2010: 240).

How, then, do frames function as tools for the analysis of discourse-specific conceptualisations? They do so by organising and representing world knowledge in the form of "a network of nodes and relations" (Minsky 1975: 212) which defines conceptual slots. Frames can either assign default values to such slots, which then remain implicit, since they represent collectively shared information units, or they can provide slots for overtly realised fillers. Ziem (2008a: 369–375; 2008b: 100–110) ties in with this slot-filler structure. His approach is basically informed by the idea that frames which organise the content of a given discourse are structured in a specific way. In other words, the discourse-specific configuration of frames can be determined on the basis of a corpus of texts representative of a given discourse (Ziem 2008a: 406–420).

According to Ziem (2008a: 399–401), this specific configuration of frames can manifest itself in two ways, since frames are structured by both fillers and default values. On the one hand, conceptual slots of a frame may or may not occur with fillers. If slots remain largely void, this points to their rather marginal role in a given discourse. Meanwhile slots with overtly realised fillers are indicative of a more important discursive role. *Nota bene*, conceptual slots can be identified by questions. This means that overtly realised slot fillers provide answers for questions raised in a given discourse, and lacking slot fillers point to the discourse-specific irrelevance of a question.

On the other hand, the frequency of specific fillers shapes a given discourse. An observable strong recurrence of a conceptual slot and a specific filler contributes to the emergence of a new, discourse-specific default value and lends a slot a more central character as compared to others. Within a given discourse, knowledge is then reorganised in a way, which contributes to the entrenchment of new default values essentially characterising a given discourse.

Summing up, the discourse-specific configuration of frames is defined by the appearance of fillers for particular slots (and the absence of fillers for other slots), and by the entrenchment of discourse-specific default values. Both processes give conceptual slots cognitive prominence and centrality in a given discourse (Ziem 2008b: 110).

3.2 Coding procedure and identification of frames

In order to identify empirically the frame(s) organising or underlying the conceptualisation typical of the #SanEscobar discourse and their configuration, I adopted a coding procedure inspired by Grounded Theory and its adaption to multimodal data by Fraas et al. (2010: 240–250) and Fraas and Pentzold (2015). This procedure aims at identifying the smallest conceptual units serving as fillers and potentially emerging as new default values in the discourse-specific configuration of a frame. It subsequently allows for the identification of the frames organising the multimodal content.

Grounded Theory, despite its name, describes a heuristic mentality rather than a theory (Glaser and Strauss 1967; Strauss and Corbin 1990). It refers to a research process which reconstructs a theory from the material analysed. A coding procedure inspired by Grounded Theory does not rely on *a priori* categories. Instead, it determines the categories relevant in a given dataset in a *bottom up* procedure (Strübing 2014). Here, it is important to stress that Grounded Theory is a research paradigm from social science (cf. Mayring 2010; Kuckartz 2012), which was not developed for, and has not been tailored to, frame analysis. The same is true of the software used for coding (ATLAS.ti). Nevertheless, both, the methodological procedures and software offer features which are fruitful for a frame-based discourse analysis.

For this study, I used ATLAS.ti to identify and code the recurrent, and thus relevant, categories across diverse modalities and languages in an explorative fashion. Codes which proved to work well in annotating the screenshot correspond to the conceptual units which serve as slots in a discursively relevant frame. The material coded with a given category label, be it a pictorial, sound or video element or a text segment, then provides the discourse-specific fillers which emerge as potential default values in the discourse-specific configuration of a frame. In its code manager, ATLAS.ti lists all passages annotated with a particular code. A number of contributions, for example, elaborate on the geography and climate of San Escobar, so these proved to be meaningful codes and thus relevant aspects in the given discourse. The data extracts which were listed in the code manager under these codes – that is, the fillers (e.g. photos of tropical beaches, Spanish phrases) – then point to San Escobar being located in the South American tropics. Moreover, the high incidence of these fillers (i.e. their recurrence in these slots) points to them emerging as discourse-specific default values. In other words, the coded material serves as a linguistic or multimodal semiotic resource that represents conceptual slots within frames which are relevant for the construction of the identity of the virtual country of San Escobar.

In addition, ATLAS.ti provides a tool for grouping codes in families (family manager). This function proved useful in the identification and representation of frames since code families can be associated with frames. In this case, for example, the relevant codes can essentially be grouped into four families (cf. frames). They refer to conceptual slots within a [COUNTRY]-frame (e.g. geographic location, climate, culture, government, state symbols), a [PARTY]-frame (e.g. politicians, political projects, rhetoric), a [FICTION]-frame (e.g. works of fiction including their locations and main characters, genres) and a frame comprising ideas associated with Pablo Escobar (e.g. drug lords and cartels, Colombia as a narco-democracy, drug trafficking, cocaine). A code family identified in this way already represents a discourse-specific configuration of a particular frame since it includes only codes/ conceptual slots relevant within the dataset (i.e. only slots for which fillers occur in the coded material). When jointly shaping the identity of the fake country and its leader, users thus evoke and configure these frames contributing to the discourse-specific conceptualisation.

Lastly, ATLAS.ti offers the possibility to check and display which codes and/or code families co-occur in the data. Co-occurring codes or code families point to a contiguity or fusion of frames in the discourse-specific conceptualisation (cf. Ziem 2008a). This process is referred to as *conceptual blending* (Fauconnier and Turner 2003; Oakley and Coulson 2000), which essentially relies on the selective combination and mapping of conceptual elements from at least two different frames and which results in the emergence of a novel conceptual structure. Discursive conceptualisations relying on metaphor (i.e. explaining or interpreting a phenomenon from one domain in terms of a phenomenon from another domain) represent a subtype of blending. These discursive strategies thus account for the hybrid character of the identity of San Escobar and its leader that trace back to the mapping and blending of conceptual elements across different frames.

4. Conceptualising San Escobar on Twitter

The participants contributing to the discourse accessible via #SanEscobar are involved in a process of joint online creativity. In this process, they collectively shape and elaborate on the concept of San Escobar as a country while humorously incorporating critique of the PiS party, and of its political and societal projects. In this section, I explain this process in terms of discourse-specific configurations and combinations of, mainly, two of the frames which organise the multimodal content on Twitter: the [COUNTRY]-frame underlying the conceptualisation of San Escobar

(4.1) and the [PARTY]-frame underlying that of PiS (4.2). The other (discursively configured) frames – the [FICTION]-frame and the [PABLO ESCOBAR]-frame – are alluded to in discourse segments or contributions which ridicule the national conservative, right-wing political project promoted by PiS and its members. The conceptual structure which characterises these contributions can be analysed in terms of blends combining these two frames with elements and structure from the discourse-specifically configured [PARTY]-frame (i.e. [PIS]-frame). I therefore do not devote a separate subsection to them but rather comment on them insofar as they provide conceptual structure to blends ridiculing the PiS.

The following two subsections thus detect how the identity of San Escobar and its leader emerges in this process of conceptualisation and devotes special attention to three aspects playing a key role in identity works. These are

1. the linguistic and other semiotic practices and resources employed in order to evoke crucial concepts and frames (e.g. the [PIS]-frame or the [PABLO ESCOBAR]-frame) relevant for identity construction (cf. Bucholtz and Hall 2005: 588),
2. the discourse-specific configuration of the [COUNTRY]-frame which incorporates reminiscences of a narco-democracy and a tropical holiday destination with general cultural features of South America and which, in conjunction, characterise the identity of the fake country of San Escobar and
3. the multiple ways in which conceptual units from different frames involved in the conceptualisation are mapped onto each other. These suggest diverse links between Poland and San Escobar or between PiS and the upper echelon of San Escobar and thus account for the hybrid identity of San Escobar and its leader.

4.1 San Escobar as a country

The conceptualisation of San Escobar as a country, obviously, gets much of its conceptual structure from a general [COUNTRY]-frame, which represents our knowledge about the political and administrative organisation, institutions, culture and symbols, geographic and climatic characteristics, etc. that characterise a particular country. Contributions to #SanEscobar elaborate preponderantly on three aspects of the virtual country San Escobar: its geography (including its climate, landscape and culture), its political organisation (e.g. state symbols, political regime and head of state) and its international integration and relations (diplomatic, military or commercial). As a result, codes (cf. conceptual slots) for these aspects proved applicable in the annotation of the material. They point to conceptual slots relevant in the discourse-specific configuration of the [COUNTRY]-frame. The corresponding 'quotes', understood as coded extracts irrespective of their modality (cf. fillers), then shape the characteristics of the virtual country emerging in the discourse. As discourse-specific fillers, they provide information about how the concept of San Escobar is elaborated and thus shape the identity of the fake country. Recurrent content in a particular conceptual slot then points to the emergence of a discourse-specific default value, and thus to the cognitive and discursive prominence of this conceptual aspect.

The contributors collectively devise San Escobar as a southern country. Only a few tweets suggest that it is an island, but – judging by the pictures posted – it definitely has sandy beaches and tropical vegetation. Posts referring to Latin music, as well as Latin American dances or food and beverages (e.g. tequila), further suggest that it is located in South or Central America, or at least that its culture is Latin American. These features also turn it into a holiday destination, an aspect broached in more than hundred tweets.

The two pictures in Figure 1 are representative examples of these. The left one is an apparent offer of holidays in San Escobar at prices from PLN 1999 (the respective tweet text reads: 'Fly with us to #SanEscobar' (*Poleć z nami na #SanEscobar*).[2] The design of the right one is reminiscent of well-established internet booking sites, with the name of the supposed resort (*Comrade Waszczykowski Resort & Spa*) referring to the politician who triggered the Internet phenomenon and alluding to socialism by the form of address *comrade*. The accompanying text in the respective tweet suggests that the Polish parliament (*Sejm*) will not start its sessions until after 25 January, because members have booked a holiday in San Escobar (*Ale wiecie dlaczego Sejm wznawia obrady dopiero 25 stycznia?*).[3]

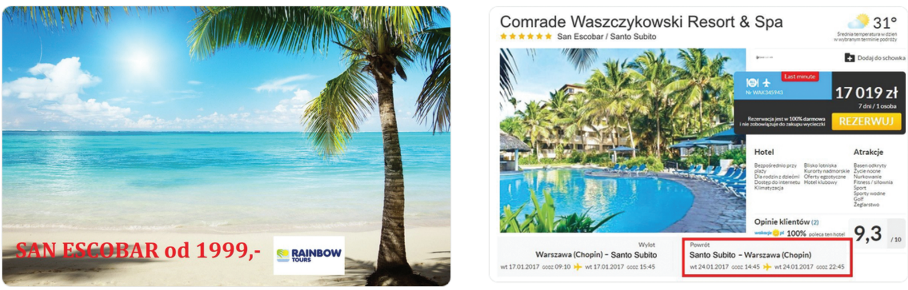

Figure 1. Travel and tourism advertisements[4]

Further recurrent themes relating to tourism are flights by the airline *El Niño* directly connecting Warsaw and Santo Subito, the capital of San Escobar, (left picture

2. https://twitter.com/szatko80/status/818858493058813952

3. https://twitter.com/Phil_Backensky/status/819587185976496131

4. https://twitter.com/szatko80/status/818858493058813952 (left) and https://twitter.com/Phil_Backensky/status/819587185976496131 (right)

in Figure 2). In one example, the beaches of Esperal Bay (*Bahia de Esperal*) have been turned into a popular destination for those fleeing the cold Polish January (right picture in Figure 2).

Figure 2. Examples giving information about flights, lifting of visa requirements and Esperal Bay[5]

The examples also illustrate a linguistic strategy frequently employed in the #SanEscobar discourse, which relies on giving places, companies or people from San Escobar names associated with the Spanish- (or more general Romance-) speaking world, or at least sound as if they were (cf. Badyda and Warda-Radys 2018). The capacity of these names to evoke specific social or cultural meanings, which are exploited for the construction of the identity of San Escobar, can be explained in terms of *enregisterment* (Johnstone 2016). This means that the linguistic items (mostly phonolexical features) which index the relevant cultural meanings are (at least perceived as) iconic representation of them (cf. Kirner-Ludwig 2018: 43–50 on the role of stylisation in the construction of fake identities). The recurring forms thus iconically index the Spanish-speaking world or at least the Romance south and thus help locating and characterising the virtual country. Here, contributors very often resort to names that establish links to other conceptual domains. Thus, *El Niño* is a meteorological phenomenon affecting the Pacific coast of South America and *Esperal* a drug applied in alcohol withdrawal therapy, while *Santo Subito* reminds us of calls for the immediate canonisation of Pope John Paul II (Karol Józef Wojtyła) after his death in 2005 (cf. right tweet in Figure 3).

The same strategy underlies a whole series of memes featuring different places in San Escobar. The image macros in Figure 3 are further memetic instantiations of the pattern presenting the beaches of Esperal Bay (cf. Figure 2). Another source for Romance-sounding place names is the Godfather series of movies alluded to

5. https://twitter.com/OlgaLetycja/status/819146015433490432 (left) and https://twitter.com/rpdsanescobar/status/819161154245316608 (right)

Figure 3. Image macros featuring locations in San Escobar[6]

via the names of actors and the characters they play (e.g. the city of *Al Pacino* in the district *De Niro*, or the *Sierra Corleone*).[6]

Several contributions even provide a rather detailed map of San Escobar (see Figure 4). It locates the virtual country not only in relation to real ones (*Meksyk* 'Mexico') but also to other imagined ones (*Westeros, Legoland*) and to nonsense countries given Spanish-sounding names (*San Serriff, San Pequeño, San Theodoros*). In doing so, the map also promotes the idea of San Escobar as a state with borders, symbols and – by extension – a political and administrative organisational structure.

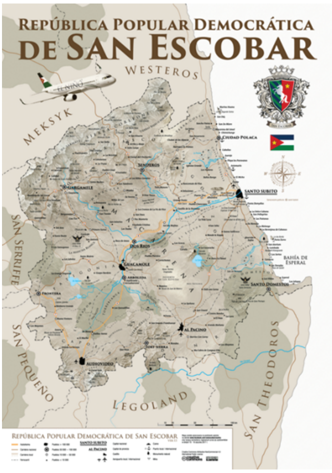

Figure 4. Map of San Escobar[7]

Other examples elaborate on San Escobar as a state with a leader, a specific political system, official symbols (flag, coat-of-arms, anthem, etc.), a capital (*Santo Subito*, see

6. https://twitter.com/adam_lelonek/status/819140098155544578 (left) and https://twitter.com/SanEscobarUN/status/819326194252398593 (right)

7. https://twitter.com/niko2win/status/825057623024222208

above) and a currency. In many regards, the state thus delineated references socialist regimes in South and Central America. In this regard, especially visual (e.g. collages/montages) as well as linguistic (e.g. naming strategies) resources help shaping the identity of San Escobar and its leader in evoking relevant concepts and frames. For example, the flag (cf. top right on the map in Figure 4) resembles that of Cuba with its star in the red triangle. Its 'official' name (*Republica Popular Democratica de San Escobar* – 'People's Democratic Republic of San Escobar') and its ideology (*popular socialism*) point in the same direction. Similarly, tweets refer to the leader either as *el presidente* or as *el comendante* [sic!]. The latter can either be interpreted as an allusion to *el comandante Che Guevara* or to the head of a South American military junta regime. As a rule, pictorial elements included in tweets about the leader of San Escobar display Jarosław Kaczyński, the former president and now head of the PiS party (cf. Figure 5). These tweets establish a conceptual link between the [PIS]-frame and the concept of San Escobar, mapping the roles of party leader and head of state onto each other and thus creating a blend ridiculing Kaczyński by implicitly comparing him to the leader of a South American state. Such contributions furthermore illustrate how blending (either linguistically or visually established in the form of photomontages and collages or by tagging pictures in specific ways) accounts for the hybrid character of the identity of San Escobar and its leader. They portray the leader of San Escobar as a persona combining features of Jarosław Kaczyński, Pablo Escobar and the head of a socialist South American state. In doing so, they also characterise the fake country as a state sharing qualities of PiS-governed Poland and of a South American narco-democracy or socialist regime.

Figure 5. Pictures of Jarosław Kaczyński as leader of San Escobar (*El Presidente, El Comendante*) and its currency[8]

8. https://twitter.com/MagdaDlugosh/status/819985925870653440 (left) and https://twitter.com/Chris_Kurdyla/status/819128672263307266 (right)

A few posts refer to San Escobar's currency as the *peso sanescobariano*. However, the name usually given to it is the *Pablo* (on the right in Figure 5), which establishes a further conceptual link to the [PABLO ESCOBAR]-frame and seems to interpret the leader of San Escobar / head of PiS in terms of a South American drug lord. Other aspects (i.e. conceptual slots) characterising San Escobar as a state are not as extensively elaborated on (i.e. filled) in the tweets but serve to complete the picture. Also featured, for example, are pictures of official San Escobar documents (photomontages of passports and ID cards) and embedded music clips (mostly Latin pop) tagged as the official anthem of San Escobar.

Lastly, and this can be regarded as another aspect of conceptualising San Escobar as a state, tweeters devote much attention to its integration in the international community and relations with other countries. Indeed, content posted points to the vigorous growth of cooperation with Poland following Waszczykowski's talks at the UN. Cooperation, and relations in the widest sense, concern four areas – diplomacy, defence, business and migration.

A large number of contributions deal with Waszczykowski's initial appearance at the UN, when his slip of the tongue created the virtual country. Tweets further elaborate on the deepening of diplomatic relations between San Escobar and Poland, for instance by the opening of embassies in each country and by the Polish Foreign Minister's visit to San Escobar and his meetings with representatives there. The latter aspect is extensively used in order to ridicule Waszczykowski. As illustrated in Figure 6, some tweets establish a conceptual link to the [PABLO ESCOBAR]-frame, suggesting that Waszczykowski initiated diplomatic relations with a narco-democracy. The example on the left shows a photo-shopped picture of a handshake between him and Pablo Escobar (labelled as prime minister of San Escobar *Pablo Cocainas*). On the right in Figure 6, a photo of Waszczykowski was embellished with the emblematic phrase 'silver or lead' (*plata o plomo*) attributed to Pablo Escobar, who used it in enforcing collaboration in illegal actions. The comment on the picture in the respective tweet specifies that 'Waszczykowski has already initiated diplomatic negotiations with representatives of #SanEscobar' (*Waszczykowski już podjął pierwsze rozmowy diplomatyczne s przedstawicielami #SanEscobar*).[9] Such examples thus further stress the hybrid identity of the leader of San Escobar who, on the one hand, (often also visually) resembles Jarosław Kaczyński, head of PiS, and, on the other hand, the unofficial leader of a narco-democracy, mostly Pablo Escobar.

Many contributions mention embassies and ambassadors who represent Poland in San Escobar or San Escobar in Poland (suggesting potential candidates for the job, providing information about visa application procedures, posting

9. https://twitter.com/JanMencwel/status/818846341971460097

Figure 6. Waszczykowski on diplomatic mission in San Escobar[11]

photos of buildings tagged as embassies, etc.). Figure 7 consists of two examples featuring supposed ambassadors. The left picture purports to show the 'newly appointed Polish ambassador and his assistant on San Escobar' (*Nowo mianowany ambasador RP s asystentką na San Escobar*),[10] actually a white male tourist and a traditionally dressed native woman probably of South American origin, thereby exploiting politically incorrect strategies to ridicule Poland's current political condition. The right example pursues a similar strategy. The photo of Polish president Andrzej Duda solemnly greeting an Asian looking official is described in the tweet text as showing the ambassador of San Escobar handing over his certificate of authentication.

Figure 7. Ambassadors of Poland in San Escobar and of San Escobar in Poland[12]

---

10. https://twitter.com/Vojtekus/status/818858255913013248

11. https://twitter.com/PhilTominski/status/819497052283539456 (left) and https://twitter.com/JanMencwel/status/818846341971460097 (right)

12. https://twitter.com/Vojtekus/status/818858255913013248 (left) and https://twitter.com/xybert/status/819251597205241858 (right)

Commercial relations between Poland and San Escobar also feature under #SanEscobar, with tweets in this area focusing mainly on Polish imports from San Escobar. Among these cocaine plays a key role, which again evokes the [PABLO ESCOBAR]-frame and devises the fake country as a narco-democracy, although some tweets also mention goods, such as coffee and bananas. Others point more generally to new business opportunities in San Escobar, such as the left example in Figure 8. The picture shows Waszczykowski enticing Polish companies to invest in San Escobar: 'Don't be afraid of foreign markets! Invest in San Escobar!' (*Nie obawiaj się zagranicznych rynków! Rozwiń biznes w San Escobar!*).[13]

Figure 8. Economic and military cooperation between Poland and San Escobar[14]

Tweets elaborating on the topic of defence cooperation allude to sales of arms (including other equipment) between Poland and San Escobar, as well to joint manoeuvres. The right example in Figure 8 illustrates this aspect. The photomontage is titled 'Poland and San Escobar in joint naval manoeuvres. St. Kitts & Nevis are concerned!!' (*Wspólne manewry morskie Polski i San Escobar. Saint Kitts and Nevis zaniepokojone!!*)[15] and owes much of its humour to the stark contrast between the warships involved.

Relative to the previously mentioned topics, that of migration plays a minor role among the tweets. It is, however, also used to mock the political situation created by the PiS government and its political protagonists. Although tweets mention migration in both directions, more of them feature Poles moving to San Escobar than refugees from San Escobar arriving in Poland. The reasons given for emigration are the cold winter and the political situation in Poland. Where the latter is concerned, participants either consider emigrating themselves or suggest that the PiS government, or at least particular members of it, should migrate to San Escobar.

13. https://twitter.com/IT_SERVAQ/status/819104614343643136

14. https://twitter.com/Chris_Kurdyla/status/819105854247104512 (left) and https://twitter.com/IT_SERVAQ/status/819104614343643136 (right)

15. https://twitter.com/Chris_Kurdyla/status/819105854247104512

By and large, the fake country of San Escobar owes its hybrid character to a number of strategies relying on visual and linguistic resources but also on the content of tweets which evoke and often mix features of South American socialist regimes, of narco-democracies and of tropical holiday destinations. Contributions furthermore elaborate on the fake country's relationships with Poland gaining momentum. The identity of its leader is even more hybrid in blending features of a South American drug lord (often specified as Pablo Escobar) or of the head of a socialist and/or military junta regime onto Jarosław Kaczyński, the leader of PiS.

4.2 The conceptualisation of PiS

Tweets marked by the hashtag #SanEscobar also refer to knowledge about the PiS party, especially those that ridicule its members or political projects. This knowledge is represented in a discourse-specific configuration of a more general [PARTY]-frame which organises the multimodal content alluding to PiS in the discourse on #SanEscobar. In this subsection, I discuss the aspects of a political party (i.e. the conceptual slots) made relevant in this discourse and present the way in which these are realised in form and content (i.e. the fillers). In this subsection, I thus adopt a different perspective on the fake country of San Escobar and its leader in looking at them through the lens of Polish politics. This perspective shifts focus to the features of PiS made relevant for the characterisation of San Escobar.

The discourse-specific conceptualisation of PiS focuses on Waszczykowski and other prominent party members, on several of its political projects and on its characteristic rhetorical style (cf. PiS speak). When these conceptual units are recruited from the [PiS]-frame, they mostly co-occur with elements from disparate frames (comprising, for example, knowledge about works of fiction, or Pablo Escobar and drug cartels), which results in humorous effects and is exploited to ridicule PiS in a similar way to the examples in Figures 5–7.

In this subsection, however, I change the point of view. In the preceding subsection, I adopted the concept of San Escobar as a point of departure and analysed diplomatic relations and international integration. Now, the concept from which I depart is that of PiS. Accordingly, I am interested in the discursive conceptualisation of PiS and its protagonists, and I analyse the discursive contexts in which they occur and which accordingly shape their presentation. Contributions elaborating on Waszczykowski initiating and maintaining diplomatic relations with San Escobar, for example, are examined here in terms of the co-occurrence of elements drawn from the [PIS]-frame (here Waszczykowski as a PiS politician) and from the discourse-specific configuration of the [COUNTRY]-frame characterising San Escobar (here diplomats / official representatives of San Escobar). Owing to the particular way PiS is conceptualised, however, other aspects also come to the fore.

Tweets to #SanEscobar deal with various PiS politicians: the president of the Polish Republic, the leader of the party, members of the government and prominent members of parliament. The most frequently mentioned is, not surprisingly, Witold Waszczykowski, the former Minister of Foreign Affairs who, as we have seen, plays a crucial role in the entire discourse. He is presented in two roles: as the discoverer of San Escobar and as a minister who establishes diplomatic relationships with virtual countries. Both pave the way for ridicule.

Thus a number of contributions mock Waszczykowski's bad knowledge of geography, as illustrated in Figure 9. The example on the right features a monkey probably searching for San Escobar on a globe, a popular motive. The comment on the picture in the respective tweet's text reads: 'Hey! Don't show this to #Waszczykowski!' (*Ej! Nie pokazujcie tego #Waszczykowskiemu!*).[16] Similarly, some tweets ironically propose to award Waszczykowski the Nobel Prize for his discovery. In other contributions, he is compared to great explorers such as Vasco da Gama (earning him the nickname *Waszczo da Gama*) or Christopher Columbus. The tweet text accompanying the picture shown on the left of Figure 9 humorously assumes that Waszczykowski 'fixed his gaze too much on Columbus' (*za bardzo zapatrzył się Kolumba*).[17] The picture is a classic portrayal of Christopher Columbus with Waszczykowski's face photo-shopped into it and subtitled 'Explorer/Discoverer of San Escobar'.

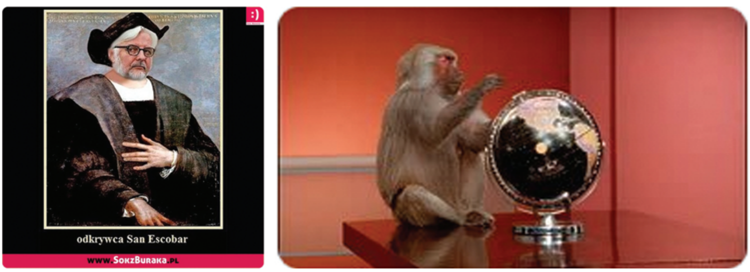

Figure 9. Examples ridiculing Waszczykowski for his discovery of a virtual country[18]

Another line of ridicule is realised by tweets suggesting that Waszczykowski establishes or maintains diplomatic contact with other virtual countries or meets their representatives at fictitious locations. The co-occurrence of elements from a specifically constructed [FICTION]-frame characterises the conceptual structure

16. https://twitter.com/phil_backensky/status/818874912932827136/photo/1

17. https://twitter.com/mat_poland/status/819268525315551232

18. https://twitter.com/mat_poland/status/819268525315551232 (left) and https://twitter.com/phil_backensky/status/818874912932827136/photo/1 (right) of these contributions, which often appear as blends. They allude to countries, places and characters from the works of genres such as fantasy and science fiction (e.g. Narnia, Lord of the Rings, Game of Thrones, Harry Potter, Star Wars, Marvel Universe), Hollywood films (e.g. Twin Peaks, Gotham) or Polish animation films (e.g. *Dolina Muminków, Wyspy Bergamuta*). Contributions also refer to Atlantis and 'as if country' (*Nibylandia*), as well as Intermarium (*Międzymorze*), a political project pursued by Józef Piłsudski after World War I for a federation of non-Russian Slavic countries ranging from the Baltic to the Adriatic Sea. Places or characters from these sources are mapped onto the role of 'real' states and their diplomatic representatives who meet Waszczykowski, who in turn represents Poland. The examples in Figure 10 are about the initiation of diplomatic relations with Narnia and the Galactic Empire (left) and plans to meet the ambassadors of Gondor and Rohan, two countries in Tolkien's *Lord of the Rings* (right).

Figure 10. Examples ridiculing Waszczykowski for diplomatic meetings with virtual countries[19]

Other members of the PiS government are also referred to in #SanEscobar, but not as often as Waszczykowski. Beata Szydło, Prime Minister during the period covered by this study, is shown mainly in a frequently posted caricature showing her with party leader Jarosław Kaczyński sniffing cocaine brought by Waszczykowski from San Escobar. The then Minister of Defence, Antoni Macierewicz, a renowned ideological hardliner within PiS, features in tweets about military cooperation with San Escobar (see Section 4.1). Polish president Andrzej Duda is, as a rule, shown meeting diplomats from San Escobar (see 4.1). The fact that Szydło and Duda are less prominent in the concept of PiS on #SanEscobar reflects a general assumption that they exert only marginal influence on the otherwise disputed ideology and political projects of PiS.

19. https://twitter.com/DagmaraPakulska/status/819462999685664768 (left) and https://twitter.com/AGajewska/status/818859104223510528 (right)

In fact, the PiS politician with the second highest number of mentions on #San Escobar is party leader Jarosław Kaczyński (also referred to as *prezes*). Although he holds no other political office, he is widely thought to pull the strings behind the scenes. His discursive salience thus reflects his assumed influential role in Polish politics. As already shown in Section 4.1., he is displayed as president of the virtual country San Escobar (cf. Figure 5) or as Pablo Escobar, roles often mapped onto each other in conceptual blends recruiting elements from a specifically configured frame comprising knowledge about Pablo Escobar. The hybrid identity of the leader of the virtual country of San Escobar is this time thus examined merely from a different perspective, namely from the vantage point of Polish politics and its leading figure behind the stage. The linguistic and semiotic practices involved in the construction of the identity of the leader of PiS / San Escobar and introducing relevant conceptions e.g. of Pablo Escobar resemble those discussed in Section 4.1.

Images, in this vein, often allude to portrayals of Escobar in film adaptations (e.g. Netflix's *Narcos, El patron del mal*). The example in Figure 11 establishes a conceptual link between Kaczyński and Escobar via the billboard, showing the actor playing the main character in the Netflix series, with Kaczyński's face photoshopped onto the actor's face. This link is fostered by the linguistic structure *Jarcos*, a morphological blend of *Jarek* (the short form of Kaczyński's first name *Jarosław*) and *Narcos*. Contributions relying and elaborating on the conceptual structure of this blend thus compare PiS's leader to a South American drug lord, which suggests an understanding of the party in terms of a drug cartel. Indeed, one tweet includes an organisational chart of the Medellin cartel with various PiS politicians blended in.

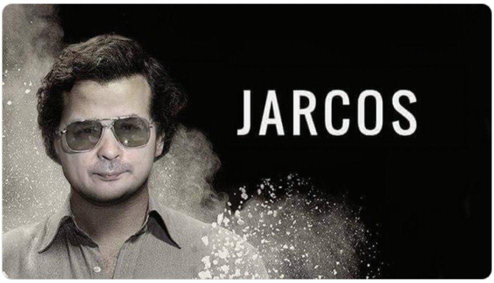

Figure 11. PiS politicians associated with the world of drugs (cocaine)[20]

Other PiS members also associated with drugs in tweets are Marek Kuchciński and Ryszard Terlecki. Both are known for their hippie past before their turn towards the political right, which probably evokes the issue of drug consumption. The examples in Figure 12 present them as interested in San Escobar's

20. https://twitter.com/MacSond/status/819321182046535681

main export: good cocaine. The image macro on the left shows Terlecki saying 'I'm content with the visit to #SanEscobar – for breakfast we had the best *kleik*' (*kleik*, meaning porridge, is also slang for cocaine) (*Jestem zadowolony z wizyty w #SanEscobar na śniadanie najlepszy był kleik*).[21] Another tweet is even more explicit: 'Newsflash! Poland increases imports of white powder from #SanEscobar after talks with Minister Waszczykowski. Parliament President Terlecki is content' (*Pilne! Po rozmowie min. Waszczykowskiego Polska zwiększa import białego proszku z #SanEscobar. Marszałek Terlecki zadowolony*).[22] The right picture in Figure 11 shows Waszczykowski explaining to the press, 'I've asked Escobar for some better stuff for Kuchciński and Kukiz… because lately they have been shaking too much' (*Prosiłem Escobara o lepszy towar dla Kuchcińskiego i Kukiza… bo ostatnio ich za bardzo trzepnie.*).[23]

Paweł Kukiz is a former rock musician and supporter of Tusk and the Civic Platform, who has also turned national conservative; he ran for president in 2015 and founded a right-wing party, the *Kukiz'15* movement. This group of tweets thus mocks PiS politicians and other representatives of the new right in suggesting that they are hypocrites in pretending to denounce their past lifestyles and opinions.

Figure 12. Further examples associating PiS politicians with the world of drugs (cocaine)[24]

In #SanEscobar, PiS is conceptualised as a party with a distinct political programme to be put into practice by changes in legislation. Specifically, three of the party's political projects are dealt with in the context of #SanEscobar: its plans for education reform, a new child benefit programme (paying each family PLN 500 per month and per child) and the reform of the constitutional court mentioned above.

21. https://twitter.com/jacek2929/status/818891673627480065

22. https://twitter.com/darbyd75/status/818842355725901824

23. https://twitter.com/planeta_swira/status/819149551890206721

24. https://twitter.com/jacek2929/status/818891673627480065 (left) and https://twitter.com/planeta_swira/status/819149551890206721 (right)

All three are extremely controversial and lampooned by associating them with San Escobar (as conceptualised in 4.1) or the realm of fiction (as represented in the discourse-specific [FICTION]-frame). Indirectly, these contributions also foster the doubtful image or reputation of the fake country in presenting it as a bad model.

For example, the educational reform aimed at re-establishing a two-level school system similar to that of the 1990s is ridiculed in tweets assuming that Poland is following San Escobar's lead. Other tweets rely on conceptual links to the discourse-specific [FICTION]-frame in order to ridicule the reform. One user, for example, suggests that *Dumbledore*, head of the witchcraft school in the Harry Potter book series, will be in charge of the reform, while defence matters are being negotiated with action heroes, 'Reportedly, talks about defence issues with Iron Man himself and Captain America are still going on. Dumblediore [sic!] will oversee the education reform.' (*Podobno trwają rozmowy nt obronności z samym Iron Manem i Kapitanem Ameryką. Reformę Edukacji nadzorować będzie Dumblediore*).[25] Another tweeter is afraid that talks between Macierewicz and Trump will lead to a joint attack on San Escobar aiming at the abolition of grammar schools (*Boję się że przyśni mi się Antoni Mać rozmawiający z Trumpem o wspólnym ataku na #SanEscobar a celem będzie likwidacja wszystkich gimnazjów*).[26]

The child benefit introduced by the PiS government (*500Plus*) attracts similar ridicule. Some tweets speculate about family and child benefits in San Escobar, suggesting that PiS will finance its *500Plus* programme with money earned through trade with San Escobar. Others speculate as to whether refugees from San Escobar qualify for the programme (*Pani Premier @BeataSzydlo, czy uchodźcy z #SanEscobar mogą liczyć na #500Plus?*).[27]

Thirdly, PiS's recurrent attempts to establish government control over the constitutional court are also mocked by tweets alluding to links between Poland and San Escobar. One user, for example, refers to an embedded photo of a group of dressed up people in front of a rather exotically looking coat-of-arms as, 'The last picture of the judges of the #TK [*Trybunał Konstytucyjny*, i.e. the Polish constitutional court] in #SanEscobar after their dismissal and before their appointment as judges in Poland' (*Ostatnie zdjęcie sędziów #TK w #SanEscobar przed odwołaniem i powołaniem na miejsce sędziów z Polski*).[28] Other contributions assume that the party's attempts to abolish the separation of powers follow San Escobar's lead.

25. https://twitter.com/MikolajNowak/status/819122596448976896

26. https://twitter.com/JaStaniszewski/status/819633764116611072

27. https://twitter.com/BWicinski/status/818839772290371584

28. https://twitter.com/K_Izdebski/status/818884035128791040

Finally, the concept of the PiS, as it is shaped by the contributions to #SanEscobar, also encompasses the party's distinctive rhetoric, which is characterised by buzzwords and catchphrases conveying crucial ideological concepts of the national conservative new right (cf. Czerwiński et al. 2010). Scholars such as Głowiński (2009) describe this rhetoric as PiS speak (*PiSomowa*) due to its at least functional resemblance to socialist newspeak. As a rule, the phrases concerned are rhetorically so salient that the moment of their emergence (and thus also their author) can still be recovered (cf. Thielemann 2016). This also turns them into reliable markers of PiS affiliation. In their contributions to #SanEscobar, participants take advantage of these features in *ironically echo-mentioning* (Sperber and Wilson 1981) the phrases in their tweets, thus mocking PiS ideology. Sperber and Wilson (1981: 311–315) argue that irony derives from dissociative or derogatory allusion to content. This is the case when texts – even extremely short but highly salient ones such as the elements of PiS speak quoted on #SanEscobar – are referred to in tweets characterised by a critical attitude towards the party and its political and societal programme.

Figure 13. Example ridiculing PiS rhetoric[29]

This group of tweets ironically renders or alludes to two rhetorically salient phrases indicative of PiS ideology and puts them in a context critical of the party. The picture in Figure 13 alludes to the phrase *najgorszy sort Polaka* ('the worst sort of Pole'), which was coined by Jarosław Kaczyński to refer to those holding oppositional views. During several public appearances, he elaborated on this concept, which is strongly linked to betrayal of Poland, patriotic values and deeds. Moreover, it seems to be inherited in the same way as its positive counterpart *genetic patriotism* (cf. Thielemann 2016: 82–87), to which the phrase *POLACO PRIMA SORT*

29. https://twitter.com/MagdaDlugosh/status/819140220100743169 shown on the shirt in the image alludes. Its author is shown as a South American drug lord or at least as a sleazy criminal with a thick necklace and moustache, surrounded by banknotes (seemingly 100 and 50 euro notes). The brand name *Polnos* printed on the shirt, which can be read as an acronym built from *polski* ('Polish') and *nos* ('nose'), also conjures up the world of drugs (specifically the consumption of cocaine). The conceptual structure underlying the picture can thus be analysed as a blend mapping the PiS leader onto the boss of a criminal organisation. In doing so, it insinuates some wider resemblance between the two domains (i.e. drug cartels and the PiS party). Two other elements also trace back to aspects of PiS, i.e. its leader and his ideological affiliates. While the cat is a reference to Kaczyński's strong emotional attachment to his pet, the figure shown on the shirt's right sleeve is Tadeusz Rydzyk, a priest and founder-boss of the Catholic broadcasting stations *Radio Maryja* and *TV Trwam*, both of which address an extremely conservative audience. The picture thus fosters the implicit comparison (i.e. mapping of conceptual units from both involved frames) in multiple regards and at least implicitly also reminds of the hybrid identity of the leader of San Escobar.

The second phrase lampooned in tweets is the PiS slogan *dobra zmiana*, which translates as 'good change' and frames the PiS programme as an attempt to transform Polish politics and society for the better. Authors of such tweets humorously suggest that cooperation between Poland and San Escobar 'enhances Poland's position on the international stage' (*Rośnie pozycja Polski na arenie międzynarodowej*),[30] thereby inducing 'good change' not only for Poland but also for San Escobar and the rest of the world (*Gracias #Waszczykowski for the meeting. I hope that our future cooperation brings Good Change to #SanEscobar and all countries of the world*).[31]

5. Conclusion

The Twitter discourse on San Escobar illustrates how a counter-public gathers on social media in order to voice its critical stance towards what is happening on the political stage. The interactive and participatory features of social media facilitate quick reaction to offline events (such as Waszczykowski's slip of the tongue) and the emergence of a *semiotic network of meaning making* (Zappavigna 2011). The unfolding discourse labelled by the hashtag #SanEscobar forms a discourse arena in which participants jointly conceptualise the country of San

30. https://twitter.com/d_olczykowski/status/818860440973443074

31. https://twitter.com/afiladodepato/status/819127803862011908

Escobar as a Latin American, rather tropical country, which maintains diplomatic relations with Poland.

The multimodal contributions to the hashtag allude to PiS (its political personnel, projects, opinions and rhetoric), fiction of highly diverse genres (e.g. fantasy, science fiction, Hollywood movies, animation films, TV series) and to Pablo Escobar and his criminal activities. They exploit these allusions in order to mock PiS and its government. The underlying conceptual structure of such intertextually allusive tweets often takes the shape of blends, which recruit material from disparate frames, map it onto each other, and thus associate them with each other. In this vein, for example, tweets recurrently instantiate a blend that suggests understanding the leader of PiS in terms of a South American drug lord or the leader of the virtual state of San Escobar. Other threads in the discourse insinuate that PiS-governed Poland maintains close relations with San Escobar and other fictitious countries (known from works of fantasy, science fiction and the like). In a further perspective, these conceptual structures implicitly and explicitly draw a comparison, or suggest an ideological and cultural affinity between PiS-governed Poland and a South American narco-democracy. The emergent hybrid identity of both, the fake country and its leader thus paves the way for ridicule. It suggests in multiple ways that Jarosław Kaczyński resembles a South American drug lord (e.g. Pablo Escobar) and PiS-governed Poland a narco-democracy.

This criticism, however, unfolds in a joint process of multimodal online creativity. It strongly relies on the entertaining power of humour, which enhances dissemination on the Internet. Both the medium (social media) and the discourse strategies (relying on multimodal creativity) thus ensure the impact of oppositional voices.

References

Androutsopoulos, Jannis. 2010. "Localizing the Global on the Participatory Web." In *The Handbook of Language and Globalization*, ed. by Nick Coupland, 203–231. Oxford: Oxford University Press. https://doi.org/10.1002/9781444324068.ch9

Androutsopoulos, Jannis. 2013. "Online Data Collection." In *Data Collection in Sociolinguistics: Methods and Applications*, ed. Christine Mallinson, Becky Childs, and Gerard van Herk, 236–250. London: Routledge.

Badyda, Ewa, and Lucyna Warda-Radys. 2018. "W żartach, ale na serio – odwołania do polskiej rzeczywistości politycznej w toponimii fikcyjnego kraju San Escobar." *Język Polski* 3: 49–64.

Bednarek, Monika. 2005. "Frames Revisited – the Coherence-Inducing Function of Frames." *Journal of Pragmatics* 37 (5): 685–705. https://doi.org/10.1016/j.pragma.2004.09.007

Bös, Birte, Sonja Kleinke, Sandra Millon, and Nuria Hernandez (eds). 2018. *The Discursive Construction of Identities On- and Offline. Personal – group – collective*. Amsterdam: John Benjamins.

Bucholtz, Mary, and Kira Hall. 2005. "Identity and Interaction: a Sociocultural Linguistic Approach." *Discourse Studies* 7 (4–5): 585–614. https://doi.org/10.1177/1461445605054407

Burghardt, Manuel. 2015. "Tools for the Analysis and Visualization of Twitter Language Data." *10plus1: Living Linguistics* 1. Available online at https://epub.uni-regensburg.de/35669/ (accessed 15 April 2019).

Busse, Dietrich. 2012. *Frame-Semantik. Ein Kompendium*. Berlin, New York: De Gruyter. https://doi.org/10.1515/9783110269451

Choi, Su Young, and Younghan Cho. 2017. "Generating Counter-Public Spheres Through Social Media: Two Social Movements in Neoliberalised South Korea." *Javnost – The Public* 24 (1): 15–33. https://doi.org/10.1080/13183222.2017.1267155

Czerwiński, Maciej, Paweł Nowak, and Renata Przybylska (eds). 2010. *Język IV Rzeczypospolitej*. Lublin: UMCS.

Entman, Robert. 1993. "Framing: Toward Clarification of a Fractured Paradigm." *Journal of Communication* 43 (4): 51–58. https://doi.org/10.1111/j.1460-2466.1993.tb01304.x

Evans, Ash. 2016. "Stance and Identity in Twitter Hashtags." *Language@Internet* 13 (1). Available online at https://www.languageatinternet.org/articles/2016/evans (accessed 15 April 2019).

Fauconnier, Gilles, and Mark Turner. 2003. *The Way We Think. Conceptual Blending and the Mind's Hidden Complexities*. New York: Basic Books.

Fraas, Claudia. 2013. "Frames – ein qualitativer Zugang zur Analyse von Sinnstrukturen in der Online-Kommunikation." In *Die Dynamik sozialer und sprachlicher Netzwerke*, ed. by Barbara Frank-Job, Alexander Mehler, and Tilman Stutter, 259–283. Wiesbaden: VS. https://doi.org/10.1007/978-3-531-93336-8_12

Fraas, Claudia, Stefan Meier, and Christian Pentzold. 2010. "Konvergenz an den Schnittstellen unterschiedlicher Kommunikationsformen: Ein Frame-basierter analytischer Zugriff." In *Neue Medien – Neue Formate. Ausdifferenzierung und Konvergenz in der Medienkommunikation*, ed. by Hans-Jürgen Bucher, Thomas Gloning, and Katrin Lehnen, 227–256. Frankfurt a. M., New York: Campus.

Fraas, Claudia, and Christian Pentzold. 2008. "Online-Diskurse – Theoretische Prämissen, methodische Anforderungen und analytische Befunde." In *Methoden der Diskurslinguistik. Sprachwissenschaftliche Zugänge zur transtextuellen Ebene*, ed. by Jürgen Spitzmüller, and Ingo Warnke, 287–322. Berlin, New York: De Gruyter. https://doi.org/10.1515/9783110209372.4.287

Fraas, Claudia, and Christian Pentzold. 2015. "Big Data vs. Slow Understanding? Voraussetzungen und Vorgehen computergestützter Analyse transmedialer multimodaler Diskurse." *Zeitschrift für Germanistische Linguistik* 43 (1): 112–133.

Fraser, Nancy. 1990. "Rethinking the Public Sphere: A Contribution to the Critique of Actually Existing Democracy." *Social Text* 25 (26): 56–80. https://doi.org/10.2307/466240

Glaser, Barney, and Anselm Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago: Aldine.

Głowiński, Michał. 2009. *Nowomowa i ciągi dalszy: szkice dawne i nowe*. Kraków: Universitas.

Halavais, Alexander. 2014. "Structure of Twitter: Social and Technical." In *Twitter and Society*, ed. by Katrin Weller, Axel Bruns, Jean Burgess, Merja Mahrt, and Cornelius Puschmann, 29–42. New York: Peter Lang.

Hardaker, Claire, and Mark McGlashan. 2016. "'Real Men Don't Hate Women': Twitter Rape Threats and Group Identity." *Journal of Pragmatics* 91: 80–93. https://doi.org/10.1016/j.pragma.2015.11.005

Herring, Susan. 2013. "Discourse in Web 2.0: Familiar, Reconfigured, and Emergent." In *Georgetown University Round Table on Languages and Linguistics 2011: Discourse 2.0: Language and New Media*, ed. by Deborah Tannen, and Amy Tester, 1–25. Washington, DC: Georgetown University Press.

Issel-Dombert, Sandra. 2017. "*M. le Président, la prochaine fois, évitez le scooter* – Zur Parodierung von Skandalen in der Werbekommunikation im Web 2.0." *Zeitschrift für Angewandte Linguistik* 66: 99–119. https://doi.org/10.1515/zfal-2017-0004

Johnstone, Barbara. 2016. "Enregisterment: How Linguistic Items Become Linked with Ways for Speaking." *Lang Linguist Compass* 10: 632–643. https://doi.org/10.1111/lnc3.12210

Kirner-Ludwig, Monika. 2018. "Great Pretenders: The Phenomenon of Impersonating (Pseudo-)historical Personae in Medieval Blogs, or: Blogging for Someone Else's Fame?" In *The Discursive Construction of Identities On- and Offline. Personal – group – collective*, ed. by Birte Bös, Sonja Kleinke, Sandra Mollin and Nuria Hernández, 15–55. Amsterdam: John Benjamins.

Klemm, Michael, and Sascha Michel. 2014. "Big Data – Big Problems? Zur Kombination qualitativer und quantitativer Methoden bei der Erforschung politischer Social-Media-Kommunikation." In *Datenflut und Informationskanäle*, ed. by Heike Ortner, Daniel Pfurtscheller, Michaela Rizzolli and Andreas Wiesinger, 83–98. Innsbruck: Universitätsverlag.

Kuckartz, Udo. 2012. *Qualitative Inhaltsanalyse. Methoden, Praxis, Computerunterstützung*. Weinheim: Beltz.

Liebert, Wolf-Andreas. 2003. "Zu einem dynamischen Konzept von Schlüsselwörtern." *Zeitschrift für Angewandte Linguistik* 38: 57–83.

Mayring, Philipp. 2010. *Qualitative Inhaltsanalyse. Grundlagen und Techniken*. Weinheim: Beltz.

Minsky, Marvin. 1975. "A Framework for Representing Knowledge." In *The Psychology of Computer Vision*, ed. by Patrick Winston, 211–277. New York: McGraw-Hill.

Oakley, Todd, and Seana Coulson. 2000. "Blending Basics." *Cognitive Linguistics* 11 (3–4): 175–196.

Siever, Christina Margit. 2015. *Multimodale Kommunikation im Social Web: Forschungsansätze und Analysen zu Text-Bild-Relationen*. Frankfurt a.M.: Peter Lang. https://doi.org/10.3726/978-3-653-04420-1

Sperber, Dan, and Deirdre Wilson. 1981. "Irony and the Use-Mention Distinction." In *Radical Pragmatics*, ed. by Peter Cole, 295–318. New York: Academic Press.

Strauss, Anselm, and Juliet Corbin. 1990. "Grounded Theory Research: Procedures, Canons, and Evaluative Criteria." *Qualitative Sociology* 13 (1): 3–21. https://doi.org/10.1007/BF00988593

Strübing, Jörg. 2014. *Grounded Theory. Zur sozialtheoretischen und epistemologischen Fundierung eines pragmatistischen Forschungsstils*. Wiesbaden: Springer.

Tajfel, Henri. 1979. "Individuals and Groups in Social Psychology." *British Journal of Social and Clinical Psychology* 18 (2): 183–190. https://doi.org/10.1111/j.2044-8260.1979.tb00324.x

Thielemann, Nadine. 2016. "*Patriotyzm genetyczny, pólka kulturowa* and *Palikotyzacja X-a*: Blends as Catchwords in Polish Political Discourse." *Zeitschrift für Slawistik* 61 (1): 74–101. https://doi.org/10.1515/slaw-2016-0005

Zappavigna, Michele. 2011. *Discourse of Twitter and Social Media. How We Use Language to Create Affiliation on the Web*. London: Bloomsbury.

Zappavigna, Michele. 2018. *Searchable Talk. Hashtags and Social Media Metadiscourse*. London: Bloomsbury.

Ziem, Alexander. 2008a. *Frames und sprachliches Wissen. Kognitive Aspekte der semantischen Kompetenz*. Berlin, New York: de Gruyter. https://doi.org/10.1515/9783110209419

Ziem, Alexander. 2008b. "Frame-Semantik und Diskursanalyse – Skizze einer kognitionswissenschaftlich inspirierten Methode zur Analyse gesellschaftlichen Wissens." In *Methoden der Diskurslinguistik. Sprachwissenschaftliche Zugänge zur transtextuellen Ebene*, ed. by Jürgen Spitzmüller, and Ingo Warnke, 89–116. Berlin: De Gruyter. https://doi.org/10.1515/9783110209372.2.89

Index

M

N

P

R

S

This volume explores the concepts of reference and identity in public discourses. Its contributions study discourse-specific reference and labelling patterns, both from a historical and present-day perspective, and discuss their impact on self- and other-representation in the construction of identity. They combine multiple methodological approaches, including corpus-based quantitative as well as qualitative ones, and apply them to a range of text types that are or were (intended to be) public, such as letters, newspapers, parliamentary debates, and online communication in the form of reader comments, discussion pages, and tweets. In addition to English, the languages studied include Polish as well as European and Latin American Spanish. The volume is aimed at researchers from different research paradigms in linguistics and related disciplines, such as media communication or the social and cultural sciences, who are interested in the interplay of reference and identity.

ISBN 978 90 272 0420 2

9 789027 204202

John Benjamins Publishing Company