CrossMark

Politeness at Work in the Clinton Email Corpus: A First Look at the Effects of Status and Gender

Rachele De Felice[1] · **Gregory Garretson**[2]

Abstract   This article introduces the Clinton Email Corpus, comprising 33,000 recently released email messages sent to and from Hillary Clinton during her tenure as United States Secretary of State, and presents the results of a first investigation into the effect of status and gender on politeness-related linguistic choices within the corpus, based on a sample of 500 emails. We describe the composition of the corpus and mention the technical challenges inherent in its creation, and then present the 500-email subset, in which all messages are categorized according to sender and recipient gender, position in the workplace hierarchy, and personal closeness to Clinton. The analysis looks at the most frequent bigrams in each of these subsets as a starting point for the identification of linguistic differences. We find that the main differences relate to the content and function of the messages rather than their tone. Individuals lower in the hierarchy but not in Clinton's inner circle are more often engaged in practical tasks, while members of the inner circle primarily discuss issues and use email to arrange in-person conversations. Clinton herself is generally found to engage neither in extensive politeness nor in overt displays of power. These findings present further evidence of how corpus linguistics can be used to advance our understanding of workplace pragmatics.

✉ Rachele De Felice
  r.defelice@ucl.ac.uk

1 University College London, London, UK

2 Uppsala University, Uppsala, Sweden

🍏 Springer

Introduction

Since May 2015, the United States Department of State (henceforth USDS) has released into the public domain over 33,000 email messages that had been stored on former Secretary of State Hillary Clinton's personal email server. These emails had been sent or received by Clinton (who authored approximately 25% of the messages) or a select group of State Department employees. Clinton used this private server for her email communication throughout her tenure as Secretary of State (January 2009 to February 2013), contrary to normal practice, leading to serious concerns that classified information had been at risk of discovery. The emails were released to the public following a Freedom of Information Act (FoIA) request, and were made available on the US Government's FoIA website.[1] Mirror collections were then created by Wikileaks and the *Wall Street Journal*.[2]

This dataset represents a tantalising resource for linguists in general, and for those working at the intersection of pragmatics and corpus linguistics in particular. Email collections of this size very rarely become accessible to researchers, and this one is of great interest for a variety of reasons. In some ways it resembles the largest and most commonly studied collection of professional email, the Enron database (Klimt and Yang 2004), but is over 10 years more recent. Further, it contains data from over 500 correspondents, both male and female and of varying ages; these individuals also represent varying levels of seniority, both within and without the organisation. The nature of the organisation itself—the USDS—leads to discussion in the data of events of both critical and minor importance, giving us a window into decision-making at various levels. And last but not least, Clinton herself is a public figure who has attracted extraordinary attention for her role as a powerful female politician, leading to many types of discussion about gender, politics and power. These emails present her from a different perspective, through her communication with colleagues and intimate friends.

This data is fertile ground for a wide range of potential linguistic analyses, from the use of speech acts in the workplace, to direct and indirect reported speech and other aspects of politeness, to questions of language and gender. It is also likely to be of interest to researchers in other fields, such as contemporary historians, political scientists, sociologists, gender scholars, and researchers in management and organisational studies. Certainly the USDS is deserving of study, and it might be valuable to compare its hierarchies and communication patterns to those of large corporations.

Unfortunately, the Clinton email data was released in the form of redacted PDF documents, a form that is not particularly amenable to corpus methods. For this reason, we have begun processing the data, cleaning it, and creating an XML corpus with all 33,000 email messages, including rich metadata about each message. While this work is underway, we have been using a hand-crafted pilot corpus of 500 messages, described below, to perform the study reported in this article.

---

[1] https://foia.state.gov/Search/Results.aspx?collection=Clinton_Email.
[2] https://wikileaks.org/clinton-emails/ and http://graphics.wsj.com/hillary-clinton-email-documents/.

    This study focuses on three related questions: (1) Do messages traveling 'up' and 'down' the workplace hierarchy display linguistic differences? (2) Are there any linguistic differences between messages exchanged within Clinton's 'inner' circle and those involving 'outer' circle members? (3) Do the gender of the author and the recipient of the messages seem to have any effect on the language used? Bargiela-Chiappini and Harris (1996: 637) have noted, a propos business letters:

> [A] factor which affects interpersonal communication, whether written or spoken, is the status of the communicators; the language used in conveying the (potentially) face-threatening act of a request reflects, among other things, the addresser's perception of his/her own status and that of the addressee.

This study sets out to assess this claim in the specialised workplace context of the USDS, by looking at hierarchy, familiarity and gender as potential factors conditioning the status of the interlocutors.

    The article is structured as follows: In the next section, we review previous work on the pragmatics of the workplace and related areas. Then we describe the larger Clinton Email Corpus, which is currently under construction, and the smaller pilot corpus that was used in the present study. The results of this study are then presented, organised according to the three questions posed above. Finally, we sum up our observations and look ahead to future work.

Background: Workplace Discourse and Corpus Pragmatics

Politeness and Power in Professional Contexts

As Coulmas (2013: 102) notes, "[p]oliteness is inextricably linked with social differentiation, with making the appropriate choices which are not the same for all interlocutors and all situations". Traditional politeness theory postulates that speakers will vary their communication styles depending on whether their interlocutor is an equal or not, as well as whether they are familiar with each other, with a greater power differential and a lack of familiarity leading to greater use of linguistic politeness strategies (Brown and Levinson 1987). Workplaces, then, are fertile ground for the study of politeness, as they are often hierarchically organised. Indeed, there is a rich body of corpus-based research into English-language workplace discourse (e.g. Harris 2003, Vine 2004, Mullany 2007, Handford 2010, Koester 2010, Holmes and Stubbe 2015). This focuses mainly on spoken interaction and has covered topics such as the structure of meetings; how directives are issued; the role of small talk, humour, and personal interaction; and the nature of leadership. There is also a growing body of work on workplace emails, examining both structural and pragmatic aspects, such as politeness and power relations (e.g. Gimenez 2006, Waldvogel 2007, Gilbert 2012, De Felice 2013, Prabhakaran and Rambow 2013, Leopold 2015, McKeown and Zhang 2015, Kim and Lee 2017, Murphy and De Felice forthcoming). The work presented in this paper is a contribution to this field. It follows in the tradition of earlier studies, yet benefits from the combination of a relatively large amount of data and the availability of metadata about the participants

in the corpus, enabling a broad investigation into the interpersonal dynamics within a large organisation.

Power relations at work can be expected to play out both in speech and in writing. Gilbert (2012: 1037), for example, makes the claim that "[a]t work, email is the performance of power and hierarchy captured in text." Prabhakaran and Rambow (2013) describe the four types of power discerned in a subset of the Enron emails, namely hierarchical power (as determined by position within the company), situational power (which is independent of the organisational hierarchy, but rather task- or situation-dependent), power over communication (held by those who drive the communication by asking questions or issuing requests, rather than responding to such utterances), and influence (held by a person who has credibility or wants to convince others). They find that people with hierarchical power are less active in email threads—that is, they do not write as much–and note that their findings "suggest that bosses don't always display their power overtly when they interact" (ibid: 2013: 221). In fact, most findings on politeness and power in the workplace converge on the fact that even in situations of power asymmetry, more powerful speakers retain the use of politeness strategies. Kim and Lee (2017: 210), for example, found that "[a]lthough superiors may have legitimate power of control and regulation, encouraging subordinates to be autonomous and self-regulating individuals was valued, which led superiors to mitigate their requests".

If the same applies to our dataset, we would expect to find differences in who produces more questions and requests; shorter or fewer messages from more powerful individuals (e.g. Clinton); and few explicit linguistic displays of power.

Gender and Language in the Workplace

The role of gender in workplace interaction (mostly spoken) has been the subject of numerous studies (e.g. Holmes 2006, Holmes and Schnurr 2006, Mullany 2007, Baxter 2010). These works observe that, while individual workplace contexts each tend to follow their own set of communicative practices, overall, societies have a set of expectations regarding how men and women 'should' or 'will' behave in a professional context (though, as Marra et al. 2006 note, they are often found to be using the same linguistic strategies). For Western cultures, Mullany describes these expectations as of men being assertive, competitive, and aggressive, and of women being co-operative, supportive, and indirect (Mullany 2012: 513). Furthermore, with regard to Clinton's language specifically, Jones (2016: 635) notes that "Clinton's [spoken] linguistic style was most masculine during the years she served in the Senate and Department of State". Our analysis assesses the validity of these claims with respect to Clinton's email style and that of the other participants, looking at whether their linguistic behaviour is in line with stereotypical expectations.

Corpus Linguistics and Pragmatics

Although corpus pragmatics—the combination of methods from corpus linguistics and pragmatics—is a "relative newcomer" (Aijmer and Rühlemann 2015: 1) to the discipline of linguistics, there is already considerable evidence showing how they

can mutually benefit each other. Recent studies have explored pragmatic phenomena of language-in-use and contextual meaning from a corpus-based perspective, taking advantage of the quantitative analyses afforded by corpus linguistics (e.g. Aijmer 2002 on discourse markers, Adolphs 2008 on suggestions, Jucker et al. 2009, De Felice 2013 on commitments). In return, pragmatic theories can help us to better interpret the quantitative results of corpus research.

In this study, we examine the effects that hierarchy, social distance and gender appear to have on the use of politeness in the Clinton Email Corpus, which is described in the next section. This line of inquiry is an example of the "enormous potential of the combination of the two disciplines [corpus linguistics and pragmatics]" (Romero-Trillo 2017: 1) afforded by research in corpus pragmatics by providing a case study of how different sources of data and corpus analytic tools can be brought together to further our understanding of communicative practices in particular settings.

The Data: The Clinton Email Corpus

The current study was carried out on a set of 500 emails culled from the 33,000 messages that are presently being compiled by the authors into the Clinton Email Corpus (CEC; see Garretson and De Felice 2017). Here we comment on the nature of the emails released into the public domain, briefly describe the corpus compilation process, and then present the pilot corpus in more detail.

The Nature of the Emails

Despite being commonly referred to as the 'Clinton emails', in fact only a quarter of the messages released by the State Department (approximately 7500) are authored by Clinton herself; more frequently, she is the recipient of the message. Roughly 500 other individuals are represented in the corpus, many of whom work for the State Department, though there are also many people outside the USDS who had access to Clinton's private email address. About half of the messages represent communication within a relatively small group of individuals in the USDS including Clinton; Cheryl Mills, Counselor and Chief of Staff; Huma Abedin, Deputy Chief of Staff; and Jake Sullivan, Director of Policy Planning.

One might suppose that such messages would reveal global power relations and political intrigue. The messages do include reactions to critical global events and friendly emails with world leaders, but much of the daily work of these individuals, as in any organisation, is deeply mundane. Scheduling trips and meetings, planning phone conversations, asking to have documents printed, etc. make up a large proportion of these messages. While such interaction might prove to be of little interest to historians and others, it provides very useful data for linguists studying how workplace communication unfolds on a regular basis.

However, it must be noted that this mundane quality of the emails is due in part to the fact that the documents were redacted before release. Before the USDS

released these emails to the FoIA website, they were scoured by "a team of intelligence experts" (CNN 2015) who manually redacted material that was highly classified or that could endanger the privacy of non-public individuals. This redaction typically takes the form of a solid white box, covering up the text that was considered sensitive. The extent of the redaction ranges from removing individual email addresses and names to blotting out entire paragraphs in the body of the emails. In some emails, the entire text is redacted, leaving only the names of the sender and recipient, and other metadata. We estimate that two-thirds of the emails have been subjected to some degree of redaction. For researchers, there are both positive and negative sides to this redaction process. From a historical, political or institutional perspective, the fact that certain communication on certain topics between certain pairs of individuals has been redacted could be of interest; even if the actual content is obscured, much can be induced about the nature of the communication. From a linguistic perspective, it is clearly a hindrance to linguistic analysis when material is missing, especially from the perspective of co-occurrence analyses such as collocational studies. However, the data that remains is plausibly of equal interest from the perspective of pragmatics and workplace communication, as in the study presented below.

Figure 1 shows a typical example of a PDF from the FoIA website containing an email message from the corpus. Note the 'original message' included below the current message, as well as the multiple redactions, coupled with the various redaction codes at right.

Compiling the CEC: Technical Aspects

The work reported in this article is part of a larger project to produce a corpus in XML format containing all the Clinton emails, retaining both the text of the emails and metadata about the email's sender, recipient, subject line, date, timestamp, and ideally even the device on which the email was composed (computer, tablet or BlackBerry), as this can help to contextualise and explain the brevity of each message. This corpus, when complete, will be made freely available to researchers.

The process of compiling such a corpus is complicated by the fact that all 33,000 emails were submitted by Clinton to the USDS as paper printouts. These documents were scanned, supplemented with headers and footers, and then redacted by various agencies, yielding PDF files with partially degraded document text and various levels of additional text in overlays, plus the white redaction boxes—not an ideal corpus for linguistic analysis. Usefully, in 2015, the *Wall Street Journal* data team created a set of Python scripts for extracting data from the PDF files on the FoIA website.[3] Unfortunately, the quality of the resulting text is insufficient for corpus linguistic analysis (especially with regard to the effects of the redaction process), requiring the development of new tools for data extraction. The *Wall Street Journal* team also created an invaluable spreadsheet detailing all the names found in the email sender and recipient fields, and mapping onto one another the various names and addresses representing a single individual.

---

[3] https://github.com/wsjdata/clinton-email-cruncher.

UNCLASSIFIED U.S. Department of State Case No. F-2014-20439 Doc No. C05758884 Date: 10/30/2015

Classified by DAS, A/GIS, DoS on 06/30/2015 ~ Class: CONFIDENTIAL
~ Reason: 1.4(B), 1.4(D) ~ Declassify on: 07/25/2024

RELEASE IN PART
B1,1.4(B),1.4(D),B6

From:	H <hrod17@clintonemail.com>
Sent:	Sunday, July 5, 2009 5:44 PM
To:	'wburns66
Subject:	Re: Note from Bill Burns

B6

Bill--

Thanks for the update and for all the work you did and led to get us to this point. I hope you'll have time to send me periodic updates, especially your impressions of the Medvedev-Putin relationship.

I am considering a trip to the Caucasus and Central Asia in the Fall, perhaps attached to the Russia bilat, if we decide that makes sense.

As you know,

1.4(D)
B1

Travel safely. All the best, Hillary

----- Original Message -----
From: wburns66   <wburns66
To: H; H2
Sent: Sun Jul 05 15:46:20 2009
Subject: Note from Bill Burns

B6

Madam Secretary,

1.4(B)
1.4(D)
B1

Just a quick note before leaving for Moscow, Rome and Central Asia.

Moscow preparations are in reasonable shape. Afghan lethal transit agreement is done, and will be signed on Monday. There's also a broader statement on cooperation on counter-narcotics and reconstruction in Afghanistan. (Bakiyev has now signed the Manas agreement, which is a further boost on Afghanistan.)

The Presidents will announce the bilateral commission, and your trip to Russia in the fall to launch the process with Lavrov. All in all, a good, measured, realistic step in our effort to rebuild relations --

1.4(D)
B1

I'll be in Rome just for the first, "foreign policy" portion of the G8. We're working on a Leaders' statement that is likely to look a lot like last week's G8 Foreign Ministers statement on most issues, with an update on North Korea, and essentially the same, firm line on Iran.

I leave Rome on July 9 with an interagency delegation for stops in Kazakhstan, Uzbekistan, Turkmenistan and Kyrgyzstan, back in DC on July 14.

Hope you had a good Fourth, and all the best,

Bill

Fig. 1 Example of a PDF from the FoIA website

Because the nature of the relationships between the individuals represented in these emails is key in interpreting their communication, we have gathered as much and as accurate data as possible about these individuals. This includes each person's name, email address, gender, workplace, job title, and other relevant information (e.g. 'best man at Clinton's wedding'), according to publicly available sources such as the USDS website, Wikipedia, LinkedIn, and other professional websites. Of the 488 different individuals and organisations involved as senders and recipients, only 137 (28%) are within the USDS, with the remainder being in the White House and

other government departments, as well as other institutions. Of the entities or individuals represented, 262 (54%) are men, 129 (26%) are women, 69 (14%) are organisations, and 28 (6%) are entities for which further information could not be acquired. The variety of writers and the amount of available information about them allows us to explore the effect of several different variables on their communication.

Naturally, the corpus compilation process needs to deal with challenges common to all email corpora. One of these is that data is frequently duplicated in email chains, because a response to a text frequently includes the original text, either at the end or intercalated with the response. While this is useful in providing context for the message and rendering the communication easier to interpret, it is highly problematic for word counts, including word or n-gram frequency lists and collocational analyses.[4] Our solution is to use XML markup for quoted text and algorithms for excluding such material in word counts and other analyses. Unfortunately, while the markup process can be automated to some degree, it still involves some manual work.

Without question, the most salient and unusual challenge in the process of compiling the CEC is dealing with the many redactions in the corpus. These cause the obvious problem that much of the text is obscured; having large chunks of missing information impairs our understanding of the text, and its appropriate interpretation. However, the redactions also present less obvious problems; for example, many tools such as POS taggers, parsers, and even concordancers rely on complete sentences, or at least uninterrupted sequences of words.

Nevertheless, as mentioned above, the redactions themselves can be seen as an interesting feature. Their presence provides important information; it is useful for historians and political analysts to know that in a given message, between these two particular people, there was something considered sensitive by the intelligence services. So it behoves us to preserve information about the presence of redactions. Therefore, we are recording details about their position in the text, and a rough estimate of the extent of the redaction (e.g. one or two words vs. an entire paragraph).

The Pilot Corpus: Creation and Composition

For this study, we restricted our data to a sample of 500 emails to explore the viability of the corpus for pragmatic research while still fine-tuning technical details of the data extraction. The 500-email sample was hand-crafted to ensure coverage of both male and female writers, and of both Clinton and her staffers, in both sender and recipient roles. Further, we sought to include both individuals working for the USDS and outside individuals, to gain insight into the potential differences in communication within and without that workplace. Using an organogram from 2011 to understand hierarchical relations at the USDS,[5] and other contextual data for those not employed there, we labelled each message as traveling upward, downward, or across

---

[4] An n-gram is a sequence of *n* contiguous words: a 2-gram (or bigram) is a sequence of two words, a 3-gram (or trigram) is a sequence of three words, and so on.

[5] https://www.state.gov/s/d/rm/rls/perfrpt/2011/html/178728.htm.

Table 1 Breakdown of the messages in the pilot corpus by hierarchical direction

	No. of messages	No. of words	Mean words/ message	Example
Upward	249 (50%)	12937 (63%)	52	From: Special Advisor To: Director of Policy Planning
Downward	209 (42%)	4318 (21%)	21	From: Hillary Clinton To: Special Assistant
Neutral	42 (8%)	3377 (16%)	80	From: Cherie Blair To: Hillary Clinton
Total	500 (100%)	20632 (100%)		

the hierarchy to the best of our understanding, to enable research into the effect of hierarchical asymmetry (seniority) on communicative styles.

However, research into the profiles of the individuals in the corpus revealed that there is a further dimension of interpersonal relations to consider: inner versus outer circle, where 'inner' is defined as being a member of 'Hillaryland' (a self-described group of people who have worked with Clinton for a long time) or having an otherwise close connection to Clinton, regardless of USDS employment. For example, Huma Abedin, who was Deputy Chief of Staff at the time, is both a USDS employee and part of the inner circle, while Neera Tanden, President of the Center for American Progress (a progressive policy institute), is part of the inner circle while not being a USDS employee.

We believe that the pilot corpus represents a reasonable balance among these factors. The corpus includes messages from 26 female authors and 33 male authors, as well as two messages from authors whose identities have been redacted. These individuals also comprise 15 inner-circle members and 39 outer-circle members (not including the two anonymous messages). While the majority of the messages were sent either by Clinton or to her (218 and 270, respectively), Clinton's writing comprises only 23% of the words in the pilot corpus.

Tables 1, 2, 3 show the breakdown of the corpus by message type and word count for the three factors under study. Where the identity of a participant has been redacted, their gender and circle membership are unknown. Note also that there are no instances of Outer to Outer messages, because they would not have passed through Clinton's private server.

One caveat is in order regarding the length of the messages reported here: Some words were removed in the redaction process, but we have elected for this study to report the total number of unredacted words (those we can actually see) rather than attempting to estimate the original, higher, number of words.

Table 1 shows stark differences in the verbosity of writers at different levels of hierarchy, with messages traveling down (mostly from Clinton herself) being much shorter than those traveling upward, and with neutral ones being the longest of all. This suggests that when communicating downward in the hierarchy, fewer words are required to express one's thoughts and needs; however, it may also be indicative of

Table 2 Breakdown of the messages in the pilot corpus by inner/outer circle

	No. of messages	No. of words	Mean words/ message	Example
Inner→Inner	371 (74%)	13759 (67%)	37	From: Hillary Clinton To: Cheryl Mills (Counselor and Chief of Staff)
Inner→Outer	68 (14%)	2239 (11%)	33	From: Hillary Clinton To: Kris Balderston (Special Rep. for Global Partnerships, USDS)
Outer→Inner	59 (12%)	4485 (22%)	76	From: Miguel Rodriguez (Deputy Assistant Secretary) To: Huma Abedin (Deputy Chief of Staff)
Unknown→Inner	2 (0%)	149 (1%)	75	
Total	500 (100%)	20632 (100%)		

different types of communication taking place in the different directions. These possibilities will be explored in greater detail below.

Table 2, with the data broken down by circle, shows that the sample is heavily skewed towards emails exchanged within the inner circle. This is, however, roughly representative of the corpus as a whole. It is striking that messages emanating from the inner circle are much shorter than those coming into the circle, perhaps because individuals on the outside need to use more words to legitimise their communications, for example to provide longer introductions or explanations for their messages. It is also possible that the inner messages are more likely to be part of an ongoing conversation and therefore require fewer details, while the emails coming from outside are often more complete and self-contained texts. Again, these ideas will be explored below.

Table 3 also presents a skewed picture, with more, but on average shorter, messages written by women than by men. As will be discussed below, this is a consequence of several factors, including the preponderance of women in the sample, and their tendency to be members of the inner circle who are writing 'downward'.

The Study: Hierarchy, Group Membership, and Gender

In this section, we explore our pilot corpus from the perspectives of hierarchy, inner vs. outer circle, and gender, to determine to what extent these factors appear to have a discernible effect on the type of language used in this workplace. Our methodological approach is bottom-up, starting in each case from an n-gram analysis, which is used to find points of entry into the data that appear to reveal pragmatically relevant differences potentially motivated by politeness considerations. These have subsequently been explored using a concordancer.

Table 3 Breakdown of the messages in the pilot corpus by gender (where known/relevant)

	No. of messages	No. of words	Mean words/ message
Female→Female	294 (65%)	10794 (56%)	37
Male→Female	124 (28%)	7772 (40%)	63
Female→Male	30 (7%)	754 (4%)	25
Total	448 (100%)	19320 (100%)	

Language and Hierarchy

By referring to the metadata for the corpus, we were able to analyse messages between interlocutors at the same and different levels of hierarchy, to determine whether hierarchy appears to affect the linguistic forms in use. The messages in the corpus can be divided into those travelling upward or downward in the hierarchy, or across it (neutral). As seen in Table 1, the first two categories both include a large number of messages, but the first has a much greater number of words.

AntConc (Anthony 2014) was used to extract n-grams from each set of messages and create concordances. Table 4 shows the top 30 bigrams in each category; all bigrams tied for the final place are included. Note that the relative frequency (i.e. ranking) of the different n-grams is of more interest than their raw frequency, which is highly sensitive to the different sizes of the three datasets (see Table 1).

From these bigrams, we can observe substantial differences in the content of the three types of messages, and subtler differences in the tone of the messages. The upward messages appear to have three main functions. The first is to inform Clinton of her schedule, which is regularly sent to her by one of her assistants, and refers to the time and location of the Secretary of State's many engagements. These messages are highly impersonal and are responsible for the occurrence on the list of bigrams such as *secretarys office*, *en route*, and *am/pm secretarys* [*office*]. The second function is to thank the recipient—usually, but not always, Clinton, as denoted by bigrams such as *thank you* and *for your* [*time, efforts, response, insights*]. The third, and most substantial, function is to keep Clinton apprised of ongoing situations, and to assure her that further information is forthcoming: *I am* [*working to get you additional docs, checking in with Jim, putting together an action plan*]; *I have* [*not yet reviewed, been monitoring, done the points*]; *I will* [*get a version to you before the call, work on this and give you an update*]; *to discuss* [*remind me to discuss, will call you in the am to discuss, will be here to discuss when you arrive*]. There is no suggestion that Clinton is expected to do anything in response except receive the information and, at most, provide some steerage. In this sense, then, there is a clear awareness of the difference in status. However, we do not see excessive deference or facework, and apart from the use of *thank you*, there are no elaborate or formulaic expressions of any kind.

This contrasts with the content of the downward bigrams, which, we reiterate, come from a smaller pool of words, due to the brevity of Clinton's own messages. Here, the first five bigrams are not function words or schedule extracts, but represent very concrete actions and requests. *Pls print*, in second position, is very common, and stands out for being a very blunt, unvarnished request with no downtoners. It is usually directed at one of the assistants, and refers to attached memos, reports, and articles that Clinton needs to read. Since many of her email exchanges take place using her BlackBerry (FBI 2016), it is no surprise that she cannot read such documents on her device. While the direct nature of this request might appear too blunt, its routine, formulaic nature, combined with the low imposition of the task, licenses the unmitigated form. This is also in line with findings that *please* in American English tends to be used in imperative requests with low imposition, such as this one (Murphy & De Felice forthcoming). The most frequent bigram is *can you*, which is also a routine way of issuing a request: the majority of these refer to arranging a phone call—*can you* [*talk now/at a given time, call me to discuss*], or to retrieving information—*can you* [*find out, check it*]. The Secretary-centric structure of the organisation is also reflected by the third bigram on the list, *for me*, indicating the expected beneficiary of the actions discussed. It occurs most often in conjunction with *pls print*, but also with a small number of other requests such as *pls schedule time for me* and *do you have info for me*. This is not petulance, however—a bigram like *want to*, which might suggest directness, on closer examination is mainly used to solicit contact on the part of the interlocutor: *I'm up if you want to call*, *do you want to discuss now*, *when do you want to talk*.

In other words, it is legitimate for Clinton and other people writing downward to directly ask for things, but not the other way round. The strongest modal verb found is *need*, though its primary use of requesting something is mitigated by use of the first person plural form, *we need* [*to do something in writing, to get a team dedicated, to monitor closely*]. The use of *we* reduces the weightiness of the imposition by appealing to the addressee's sense of belonging to a team (cf. McCarthy and Hanford 2004: 177; Vine 2004: 97–8) and increases the likelihood of compliance. Notably, there is only one instance of *you should*, typically a direct way of performing a directive, but in fact in this case it is used in a face-enhancing utterance: *you should feel proud and satisfied*.

In addition to arranging calls and making practical requests, there is also a strong consultative element to the downward emails, not unlike the upward ones: *What do you think can be done?; Let me know what they commit; Can you/pls find out; Do you want to discuss?* This openness to consultation and discussion has been described as characteristic of a female management style, as opposed to more stereotypically masculine traits such as assertiveness and directness (see e.g. Mills 2005: 273, and below). We suggest two possible interpretations for this penchant for consultation. A positive interpretation could present this as a sign of teamwork in action, of a Secretary of State willing to discuss events and plans with a team she evidently trusts. In contrast, a negative reading could see this as a sign of weakness, of someone who has difficulty doing her job on her own and needs the help and support of other people. Unfortunately, in the absence of the communication records of

Table 4 Most frequent bigrams in the pilot corpus, by hierarchical direction

Upward		Downward		Neutral	
N-gram	Count	N-gram	Count	N-gram	Count
of the	42	can you	21	on the	15
in the	38	pls print	14	of the	13
secretarys office	37	for me	12	in the	12
on the	36	to do	12	i will	11
state department	34	want to	12	memo on	8
en route	25	do you	11	for the	7
to the	23	i will	11	hague meeting	6
white house	23	to call	11	hrc memo	6
to be	22	and i	10	stays as	6
will be	19	as	10	that the	6
with the	19	before	10	the us	6
for the	18	have to	10	there is	6
if am	18	you talk	10	and the	5
thank you	18	if you	9	i am	5
am secretarys	17	let me	9	in dc	5
this is	18	me know	9	it is	5
have a	17	find out	8	says that	5
i have	17	for the	8	they are	5
i will	17	i am	8	at the	4
we have	17	to the	8	he says	4
you are	17	a few	7	i have	4
and i	16	me to	7	is an	4
pm secretarys	16	thanks for	7	is the	4
and the	15	to discuss	7	on bbc	4
health care	15	to me	7	on new	4
private residence	15	need to	7	relationship with	4
that the	15	we need	7	spoke with	4
we are	15	you want	7	they want	4
conference room	14	about the	6	to the	4
you have	13	are you	6	want to	4
for your	12	i can	6	what they	4
to discuss	12	i don't	6	will be	4
want to	12	i was	6	your meeting	4
		id like	6		
		update on	6		
		we have	6		
		who is	6		

other former Secretaries of State, we have no way of telling whether this is a behaviour specific to Clinton, or is in fact typical of anyone in this position.

This highlights a common obstacle encountered when working with this data: the lack of a directly comparable corpus. While we do have another collection of workplace email in the Enron database, it is from a very different environment, and over a decade earlier than the CEC. In determining whether Clinton's behaviour is remarkable, we would ideally want to compare her communicative and leadership style to that of other politicians in the same office, but no such data is currently available.

The third category of messages from the perspective of hierarchy are the neutral messages. Their relative scarcity in the pilot corpus limits the extent to which we can draw conclusions, as we cannot say with certainty whether what we observe does indeed reflect the relative parity of status between the interlocutors, or simply represents the stylistic traits of the individual writers. However, we can observe the general absence of phrases that suggest requests and commitments (except for a commitment to keep communication ongoing: *I will* [*forward soon, let you know*]), and the relative frequency of words and phrases reflecting exchange of information: *memo, meetings, says that, relationship with, spoke with*. Indeed, the individuals involved in these messages are either close associates who act as informal advisors to Clinton or senior members of the USDS or other comparable institutions, for whom sharing information is more appropriate than discussing tasks.

It is interesting to compare our findings to Gilbert's (2012) study based on the Enron data. He investigated whether it is possible to predict the direction of travel of any given email message in the workplace hierarchy (upward or downward) by looking at the phrases contained within the message. He developed a list of roughly 7000 n-grams (of one to three words) that were found to predict, to varying degrees of certainty, the direction of travel of an email message. Comparing the n-grams from the CEC pilot corpus to Gilbert's list, we found only limited overlap: Of Gilbert's upward phrases, 34% are found among the CEC upward phrases, but of his downward phrases, only 20% are found among the CEC downward phrases. Conversely, 23% of Gilbert's upward phrases are in fact found among the CEC *downward* phrases, and 38% of his downward phrases are found among the CEC *upward* ones. This suggests a low level of similarity between the two datasets.

N-grams that were found to be 'upward phrases' for both corpora include *think about*, while n-grams that were 'downward phrases' for both corpora include *have time*. Yet many phrases represent mismatches, such as *would you*, which in the Enron data is predictive of downward communication but in the CEC pilot corpus is more common in upward communication. Many of the differences can be attributed to domain-specific factors: while Gilbert attempted to remove n-grams that are clearly specific to the Enron dataset, several remain (e.g. *customer*, *heat rates*, and *kitchen*—this last being the surname of a key Enron employee). Similarly, the CEC also contains n-grams that are specific to the USDS environment, such as *detainee*, *POTUS*, and *Prime Minister*. Nevertheless, it is also possible that the two workplaces display different managerial and communicative styles. Our comparison of the two corpora underscores the importance of having more such large-scale datasets available, so that we may converge on key phrases that are indeed universally shared across workplaces.

Inner-circle and Outer-circle Messages

Shifting focus slightly, here we examine the messages in the pilot corpus from the perspective of social ties, dividing them into those sent by someone within Clinton's 'inner circle', and those sent by someone outside it, termed here 'outer circle'. We consider both the status of the sender and that of the recipient, focusing on those issues not already brought up in the discussion of hierarchy.

Table 5 presents the top 30 bigrams for three types of messages: inner-circle to inner-circle, inner-circle to outer-circle, and outer-circle to inner-circle. Note that no outer-to-outer messages are included due to the nature of the corpus data.

The most general observation we can make about the inner-circle vs. outer-circle distinction is that the two groups seem broadly to be using email communication for different purposes. While the members of the outer circle are asked to perform practical tasks such as printing (cf. above) and doing other things 'for me' (*pls schedule time for me to see, I have to/want to/'d like to meet with*), they are not the recipients of more prototypical indirect requests of the 'can you' variety. Within the inner circle, in contrast, *can you* is a frequent bigram, occurring almost always in contexts such as *can you* [*talk/call me*]. Its absence from the other two sets of data suggests that talking over the phone is a central activity for members of the inner circle, but not for others, a finding strengthened by the inner-circle use of *want to* in phrases like *do you want to call* and *I'm available if you want to discuss*. In sum, Clinton seems to prefer to communicate orally with members of her inner circle, which suggests that such communication bears great significance.

Another major difference between the circles can be seen in the positioning of the writer as expressed by personal pronouns. Within the inner circle, we see a great deal of *I* and *you*. Interlocutors inform each other of their whereabouts and their plans (*I will be*), appraise Clinton's behavior or public appearance (*you are*), and, crucially, consult each other as to what to do. The centrality of *discuss* has already been noted, both over the phone and in phrases like *we have to discuss soon* and *remind me to discuss*. Other examples of consultation, giving centrality to *you* as well, come through in utterances like *let me know* [*what I should do, if I'm supposed to call, what you hear about today*] and *what do you think can be done now?* The use here of *let me know* is particularly interesting, because while this phrase has been found to be very common in business emails (e.g. Enron), it is typically used there just as a courtesy closing. Here, instead, it seems to carry real directive force and is not an empty instruction.

Messages moving outwards from the inner circle also have a focus on *I* and *you*, but now with different roles. As seen above, the *I* is mostly issuing requests, and when the *you* is not required to carry out a task, it is typically the recipient of courtesy messages such as *thank you*, *thanks for*, and good wishes for various festivities *to you and your family*. This last function, of explicitly performing politeness routines, is more or less absent within the inner circle. This distinction seems to be in line with the oft-repeated claim that we engage in more overt politeness routines with those we are more socially distant from than with those we know well (Brown and Levinson 1987; Clancy 2015). This is also mirrored in the presence of *thank*

Table 5 Most frequent bigrams in the pilot corpus, by inner/outer circle

Inner→Inner		Inner→Outer		Outer→Inner	
N-gram	Count	N-gram	Count	N-gram	Count
in the	40	pls print	17	on the	18
on the	37	for me	11	secretarys office	15
of the	34	to do	8	state department	15
i will	25	i will	7	to the	13
secretarys office	22	thanks for	7	with the	13
state department	21	to the	6	in the	11
can you	19	want to	6	we have	11
for the	19	all the	5	thank you	10
this is	19	for all	5	for the	9
white house	19	for the	5	i am	9
you are	19	have to	5	to be	9
and i	18	i asked	5	pm secretarys	8
i have	18	i was	5	we are	8
want to	18	in dc	5	and the	7
will be	18	print for	5	i will	7
health care	17	thank you	5	will be	7
i am	17	to meet	5	en route	6
that the	16	to see	5	for a	6
you have	16	w the	5	for your	6
to the	15	you and	5	have a	6
have a	14	a few	4	have been	6
let me	14	and I	4	private residence	6
to discuss	14	for I	4	vote on	6
am secretarys	13	at the	4	am depart	5
and the	13	in the	4	by the	5
at the	13	of the	4	conference room	5
it is	13	be sure	4	i have	5
to be	13	do you	4	on this	5
do you	12	for you	4	that the	5
if you	12	forward to	4	the u.s.	5
is a	12	happy new	4	the vote	5
meeting with	12	have a	4	we can	5
		i am	4	we will	5
		id like	4	what we	5
		let me	4	would be	5
		me know	4		
		meet w	4		
		so much	4		
		the best	4		
		the work	4		
		to me	4		
		you be	4		
		you have	4		
		you to	4		

you in the outer-to-inner messages, though here there may also be an element of deference at play, with many of the thanks being directed at Clinton.

The nature of messages coming from the outer circle is markedly different from that of the outgoing messages: There is much less individuality and much more 'we'. These 'wes' are very busy engaging in activities which are often of a more 'practical' nature than the consultation going on within the inner circle: *we have been thinking carefully*; *we are* [*still working*, *continuing to work*, *reviewing tonight*]; *we can update and send tomorrow*; *we will* [*stay engaged*, *provide you with recommendations*]. One can envision a division of labour between the inner and outer circles which only partly overlaps with typical hierarchical distinctions: The outer circle does the practical work of providing and circulating documents and information, and the inner circle does the intellectual work of discussing this information and making decisions (though again, much of this activity appears to take place in person or over the phone). We have also noted a greater frequency of question marks in messages going from the inner circle outwards than vice versa, further bolstering the view that the inner circle requests information from the outer circle, rather than vice-versa.

In sum, the main difference between inner-circle and outer-circle communication is not one of politeness, but rather one of function: the messages traveling in these two different directions appear to be trying to accomplish different types of communicative tasks.

Language and Gender

Several factors make it very tricky to perform a study of gender effects on language in this dataset. First, although the dataset was hand-crafted to include a balance of male and female participants (see above), this balance does not hold true for the total number of messages (or words) written by men and women, as can be seen in Table 3. Second, the USDS under Clinton was skewed in terms of gender, with women occupying many of the highest positions, and so, inevitably, a greater proportion of the messages are written by women. Third, since the corpus comprises data from Clinton's private server, the data consists predominantly of messages authored by her or sent to her, with the result that fully one-fourth of the 'female' data in the pilot corpus comes from a single individual, Clinton herself.

The fourth reason that we must proceed carefully is that there is considerable overlap between gender and the factors discussed above, hierarchy and social ties. Due in part to the staffing situation at the USDS, and in part to the fact that the data comes from Clinton's private server, all the female-to-male correspondence is downward in the hierarchy or neutral, and the male-to-female messages are correspondingly upward or neutral. Further, two-thirds of the messages from men and 95% of those from women are from the inner circle discussed in the previous section. This means that there are multiple confounds that we must bear in mind when analysing this data from a gender perspective, as it is very difficult to separate the effects of gender from those of hierarchy and inner/outer-circle status (and indeed it might not even be advisable to isolate different factors in this way). The final issue that derives from the source of the corpus is that all the emails of necessity involve

a female sender or recipient, with the result that we have no data for male-to-male communication. This imbalance means that we must treat any findings with caution, as we can only observe how the two genders differ in addressing female interlocutors, rather than in overall communication. For this reason, we will focus on the female-to-female and male-to-female categories in the analysis below.

Table 6 presents the 30 most frequent n-grams for the three categories of messages (including n-grams tied for last place). Because the female-to-male dataset is so small, with no n-grams occurring more than four times, we have only presented those with a frequency of 3 or greater. As before, the analysis will focus only on those n-grams which point to differences not already discussed in the previous sections.

In general, the lists do not contain phraseology not already encountered in the previous discussion. What this comparison does, however, is highlight how the gender variable interacts with the other ways in which we have been categorising participants, and how viewing the data through this prism sheds light on the 'gendered work' distinctions in place at the USDS. In particular, our limited data analysis suggests that the men in this group are more often involved in practical work, and are—as already indicated by what we know about Hillaryland—often on the margins of the circle.

We begin by analysing the cluster of 'I'-based bigrams. *I will, I'll, I am* and *I'm* are mostly used to indicate concrete future actions, but in many cases the actions seem to differ between genders. The men undertake actions that require some practical work, while the women signal their future engagements or whereabouts (cf. the distinction between 'action commitments' and 'information commitments' in De Felice 2013). Men write: *I will* [*work on this, check information*]; *I'll* [*call him this week, tell you what Obama said*]; *I am* [*putting together an action plan, planning a visit to El Salvador*]; *I'm* [*laboriously filling out forms, doing some more recon, at my desk and available*]. Women write: *I will* [*see her, be at home*]; *I am* [*heading for the airport, going to make a push*]; *I'm* [*free/up*]—but only when writing to other women.

This difference is in all likelihood due not to gender, but to roles: the male emails are from junior members of the group, who are more likely to be expected to carry out the practical work that underpins USDS activities. Similarly, they are less likely to need to be informed about the whereabouts of the more senior women (particularly Clinton) and to be required to announce their own. We previously noted that the inner circle emails are frequently used to arrange phone calls, and the prominence of utterances indicating one's location reflects this. The presence of *can you* [*talk*]*, to call,* and *to discuss* only in the female-to-female messages reinforces the view of a female-dominated inner circle. This closeness is also apparent in the different uses of *you*, which occurs in female-authored messages in personal, emotive phrases such as *hope you're* [*well/feeling ok*], *you are the strongest person on the planet*, *the public thinks you are doing a fantastic job*. However, in the absence of comparable messages by men in an equal relationship to their recipient, it is difficult to say whether these kinds of exchanges are attributable more to gender or more to intimacy.

Table 6 Most frequent bigrams in the pilot corpus, by gender

Female→Female		Male→Female		Female→Male	
N-gram	Count	N-gram	Count	N-gram	Count
in the	26	on the	35	i m	4
of the	22	of the	34	let me	4
i have	20	in the	26	me know	4
i m	20	i will	17	all the	4
on the	20	that the	16	follow up	4
for the	18	to the	15	for all	3
want to	18	with the	14	for the	3
can you	17	i am	13	i will	3
health care	17	for the	12	if you	3
don t	16	will be	12	in dc	3
it s	16	and i	11	of the	3
have a	15	and the	10	thanks for	3
to do	15	i m	10	to the	3
you are	15	it s	10	w you	3
hope you	14	we have	10	we need	3
is a	14	and in	9		
pls print	14	but i	9		
thank you	14	this is	9		
this is	14	for a	9		
to be	14	i ll	8		
and i	13	is the	8		
i will	13	it is	8		
re update	13	on this	8		
to call	13	to be	8		
to the	13	want to	8		
to discuss	12	we are	8		
will be	12	have a	7		
you have	12	i think	7		
you want	12	meeting with	7		
at the	11	there is	7		
do you	11	they are	7		
i am	11	to you	7		
in a	11	we should	7		
to talk	11	we will	7		
we have	11				

Some differences can also be observed in the use of 'I' in directives. The bigram *I think* is a canonical downtoner for requests and suggestions (Holmes 1984). The use of downtoners is usually attributed more to women than to men (e.g. Coates 2013: 31–49), but in this dataset the bigram is mainly used by men to hedge suggestions: *I*

think [*Jake should be on the call, calling now would be appropriate, it's also worth talking to him*]. We hypothesise that this, too, is a hierarchy-dependent difference, with hedging required because junior members are giving advice to their senior colleagues. In contrast, only female-to-female messages contain phrases like *I don't* [*know, think, understand*]. The question of whether willingness to admit uncertainty or ignorance depends on gender will require closer analysis in future work.

As for *we*, the data shows that men use it only to report information (about completed or planned work), while women also use it to formulate instructions; both facts are in line with the hierarchical direction of travel of the emails. In the male-authored emails, we find *we have* [*that paper, no formal consultative responsibility*], *we will [provide you with recommendations]*, *we are [reviewing]*. Female-authored emails include directives such as *we have to* [*go over it, think through*]. When men wish to convey a request or suggestion, as already noted with regard to *I think*, they use forms such as *we should*: *We should* [*discuss your speech*, *publicly and forcefully build on that*], which express necessity less explicitly than the *we need* and *we have to* found in downward messages, as discussed above.

The overall impression produced by a gender-based analysis of the data is that the gender differences reflect the different types of roles inhabited by the participants in the corpus. For this reason, this first foray into gender-based language patterns in the CEC does not lead to conclusive findings about the role of gender, but it does bring into clearer relief the types of roles and relationships of people within and without the USDS, and how gender overlaps with these.

Conclusions, Caveats, and Future Directions

Our study of this early sample from the CEC has shown that, in this workplace, politeness is represented less by linguistic differences than by functional ones. Hierarchy is performed not through language, but through actions: Clinton and other senior members of the group ask others to do things, and these junior members report to them. Members of the inner circle have privileged access to one another, in the form of face-to-face and telephone conversations, where we assume the bulk of the USDS's consultative work actually occurs. In line with previous research on workplace language, Clinton was found to prefer a concise and direct style generally devoid of both overt politeness markers and explicit displays of power.

The effect of gender is less clear: the men and women in our sample occupy different roles in the hierarchy, and the linguistic patterns observed are more likely ascribable to these factors than to gender. As regards Jones' (2016) claim that Clinton displayed a more masculine style during her tenure as Secretary of State, what we can say, absent comparable email datasets from other periods of Clinton's career, is that her language use appears to be conditioned primarily by her status as the head of a governmental department. While some may expect only men to occupy such roles, we see no indications of so-called 'masculine' language here.

We must also acknowledge that the lack of directly comparable datasets limits the generalisability of our conclusions. The data in the CEC are unique in being publicly available electronic communication within a US government department.

The ideal comparable corpus would be a collection of the email of Secretary of State Colin Powell, Clinton's predecessor, but no such collection is available. The most closely analogous collection of text that we are aware of are the emails of Governor Jeb Bush of Florida, who released these texts in a bid for transparency when he himself ran for president. However, these are still not entirely analogous to the communication in the CEC.

Processing the 500-email CEC pilot corpus has given us important insights into what is required to create the full CEC, and ongoing work will build on this as we develop the corpus. As we expand the range of participants, we expect a richer picture of the interplay of status and gender to emerge and strengthen our findings. Furthermore, we plan to carry out semi-automated speech act annotation of the data to ensure a more comprehensive understanding of the pragmatics of this workplace.

This work offers a first insight into how the complex relationships within Clinton's State Department are instantiated through language, and how one's role(s) in this community can determine one's linguistic choices. We believe that work in this vein clearly demonstrates the value of corpus resources like the CEC in pragmatic research.

Compliance with Ethical Standards

References

Adolphs, S. (2008). *Corpus and context: Investigating pragmatic functions in spoken discourse*. Amsterdam: John Benjamins.

Aijmer, K. (2002). *English discourse particles*. Amsterdam: John Benjamins.

Aijmer, K., & Rühlemann, C. (Eds.). (2015). *Corpus Pragmatics: a handbook*. Cambridge: Cambridge University Press.

Anthony, L. (2014). AntConc (Version 3.4.4) [Computer Software]. Tokyo, Japan: Waseda University. Available from http://www.laurenceanthony.net/.

Bargiela-Chiappini, F., & Harris, S. (1996). Requests and status in business correspondence. *Journal of Pragmatics, 26*(5), 635–662.

Baxter, J. (2010). *The language of female leadership*. Basingstoke: Palgrave Macmillan.

Brown, P., & Levinson, S. (1987). *Politeness: some universals in language usage*. Cambridge: Cambridge University Press.

Clancy, B. (2015). *Investigating Intimate Discourse: Exploring the spoken interaction of families, couples and friends*. Abingdon: Routledge.

CNN. (2015). Emails released from Clinton's private server heavily redacted. http://edition.cnn.com/2015/07/31/politics/hillary-clinton-private-emails-classified/. Accessed Nov 23, 2017.

Coates, J. (2013). *Women, men, and everyday talk*. Basingstoke: Palgrave Macmillan.

Coulmas, F. (2013). *Sociolinguistics: the study of speakers' choices*. Cambridge: Cambridge University Press.

De Felice, R. (2013). A corpus-based classification of commitments in Business English. In J. Romero-Trillo (Ed.), *Yearbook of Corpus Linguistics and Pragmatics 2013* (pp. 153–171). Dordrecht: Springer.

FBI. (2016). *Clinton e-mail investigation—Mishandling of classified–Unknown Subject or Country*. Retrieved from https://vault.fbi.gov/hillary-r.-clinton/Hillary%20R.%20Clinton%20Part%2001%20of%2013/view. Accessed Nov 23, 2017.

Garretson, G. & De Felice, R. (2017). The Hillary Rodham Clinton emails: Examining world leaders through corpora. *Paper presented at the 38th ICAME Conference*.

Gilbert, E. (2012). Phrases That Signal Workplace Hierarchy. In *Proceedings of the 2012 ACM Conference on Computer Supported Cooperative Work,* 1037–1046.

Gimenez, J. (2006). Embedded business emails: meeting new demands in international business communication. *English for Specific Purposes, 25*(2), 154–172.

Handford, M. (2010). *The language of business meetings*. Cambridge: Cambridge University Press.

Harris, S. (2003). Politeness and power: Making and responding to requests in institutional settings. *Text, 23*(1), 27–52.

Holmes, J. (1984). Modifying illocutionary force. *Journal of Pragmatics, 8*(3), 345–365.

Holmes, J. (2006). *Gendered talk at work*. Oxford: Blackwell.

Holmes, J., & Schnurr, S. (2006). 'Doing femininity' at work: more than just relational practice. *Journal of Sociolinguistics, 10*(1), 31–51.

Holmes, J., & Stubbe, M. (2015). *Power and politeness in the workplace: A sociolinguistic analysis of talk at work* (2nd ed.). Abingdon/New York: Routledge.

Jones, J. (2016). Talk "like a man": The linguistic styles of Hillary Clinton, 1992–2013. *Perspectives on Politics, 14*(3), 625–642.

Jucker, A., Schreier, D., & Hundt, M. (Eds.). (2009). *Corpora: Pragmatics and discourse*. Amsterdam: Rodopi.

Kim, S. H., & Lee, H. (2017). Politeness in power-asymmetrical e-mail requests of Korean and American corporate employees. *Intercultural Pragmatics, 14*(2), 207–238.

Klimt, B. & Yang, Y. (2004). Introducing the Enron corpus. *Proceedings of the 1st Conference on Email and Anti-Spam (CEAS).*

Koester, A. (2010). *Workplace discourse*. London: Continuum.

Leopold, L. (2015). Request strategies in professional e-mail correspondence: Insights from the United States workplace. *TESL Canada Journal, 32*(5), 1–29.

Marra, M., Schnurr, S., & Holmes, J. (2006). Effective leadership in New Zealand workplaces. In J. Baxter (Ed.), *Speaking out: The female voice in public contexts* (pp. 240–260). Basingstoke: Palgrave Macmillan.

McCarthy, M., & Handford, M. (2004). "Invisible to us": A preliminary corpus-based study of spoken Business English. In U. Connor & T. Upton (Eds.), *Discourse from the professions* (pp. 167–201). Amsterdam: John Benjamins.

McKeown, J., & Zhang, Q. (2015). Socio-pragmatic influence on opening salutation and closing valediction of British workplace email. *Journal of Pragmatics, 85,* 92–107.

Mills, S. (2005). Gender and impoliteness. *Journal of Politeness Research, 1,* 263–280.

Mullany, L. (2007). *Gendered discourse in the professional workplace*. Basingstoke: Palgrave Macmillan.

Mullany, L. (2012). Discourse, gender, and professional communication. In P. Gee & M. Handford (Eds.), *The Routledge Handbook of Discourse Analysis* (pp. 509–522). Abingdon: Routledge.

Murphy, M.L. & De Felice, R. (forthcoming). Routine politeness in American and British English requests: use and non-use of please. *Journal of Politeness Research*.

Prabhakaran, V. & Rambow, O. (2013). Written dialog and social power: manifestations of different types of power in dialog behavior. In *Proceedings of the Sixth International Joint Conference on Natural Language Processing*, 216–224.

Romero-Trillo, J. (2017). *Corpus Pragmatics. Corpus Pragmatics, 1*(1), 1–2.

Vine, B. (2004). *Getting things done at work*. Amsterdam: John Benjamins.

Waldvogel, J. (2007). Greetings and closings in workplace email. *Journal of Computer-mediated Communication, 12*(2), 456–477.