

“I apologise for my poor blogging”: Searching for Apologies in the *Birmingham Blog Corpus*

Ursula Lutzky¹  · Andrew Kehoe² 

Received: 30 October 2016 / Accepted: 25 January 2017

© The Author(s) 2017. This article is published with open access at Springerlink.com

Abstract This study addresses a familiar challenge in corpus pragmatic research: the search for functional phenomena in large electronic corpora. Speech acts are one area of research that falls into this functional domain and the question of how to identify them in corpora has occupied researchers over the past 20 years. This study focuses on apologies as a speech act that is characterised by a standard set of routine expressions, making it easier to search for with corpus linguistic tools. Nevertheless, even for a comparatively formulaic speech act, such as apologies, the polysemous nature of forms (cf. e.g. *I am sorry* vs. *a sorry state*) impacts the precision of the search output so that previous studies of smaller data samples had to resort to manual microanalysis. In this study, we introduce an innovative methodological approach that demonstrates how the combination of different types of collocational analysis can facilitate the study of speech acts in larger corpora. By first establishing a collocational profile for each of the Illocutionary Force Indicating Devices associated with apologies and then scrutinising their shared and unique collocates, unwanted hits can be discarded and the amount of manual intervention reduced. Thus, this article introduces new possibilities in the field of corpus-based speech act analysis and encourages the study of pragmatic phenomena in large corpora.

Keywords Speech acts · Apologies · Collocation · Big data · Blogs

✉ Ursula Lutzky
ursula.lutzky@wu.ac.at

Andrew Kehoe
andrew.kehoe@bcu.ac.uk

¹ Vienna University of Economics and Business, Welthandelsplatz 1, 1020 Vienna, Austria

² Birmingham City University, 4 Cardigan Street, Birmingham B4 7BD, UK

Introduction

One of the main challenges when approaching the study of pragmatic phenomena, such as speech acts, in large corpora is that they cannot, for the most part, be identified automatically. This is due to the fact that they may, on the one hand, be expressed in a potentially infinite number of ways and, on the other hand, that the forms which are prototypically associated with a specific speech act (e.g. *sorry*) may also be attested with other functions (e.g. a *sorry* state). As a consequence, studies of speech acts using a corpus linguistic methodology have tended to be based on smaller (annotated) corpora, to resort to manual forms of analysis, or to adopt eclectic approaches, focusing for instance on specific speech act verbs. In this study, we show how collocational analysis can be used to facilitate the extraction of the speech act of apology by distinguishing forms with the illocutionary force of an apology from other uses of the same forms.

To this end, we use the *Birmingham Blog Corpus* (BBC 2010), a diachronically-structured corpus comprising 630 million words of both blog posts and comments from the period 2000 to 2010. This corpus is searchable through the WebCorp Linguist's Search Engine (WebCorpLSE) software built by the Research and Development Unit for English Studies (RDUES) at <http://www.webcorp.org.uk/blogs>. Given the nature of our data, the results, in addition to demonstrating how the speech act of apology can be identified in a corpus of this size, provide insights into the areas in which bloggers and their readers see a need for apologising as well as into the distribution and clustering of specific types of apologies in blog posts versus comments.

In the following, we will begin by discussing the nature of apologies and previous studies in the field (“[The Speech Act of Apology](#)” section). We review the challenges entailed by corpus-based analyses of speech acts in the “[Corpus Pragmatics and the Study of Speech Acts](#)” section, before presenting the data and methodology employed in the current study (“[Data](#)” and “[IFID Selection](#)” sections). Starting out from a list of prototypical Illocutionary Force Indicating Devices (IFIDs), we will devise a collocational profile of each of the forms studied (“[IFID Selection](#)” section) to then move to an investigation of their shared and unique collocates (“[Collocational Analysis](#)” section). This will allow us to determine to what extent apology IFIDs overlap in their function and tend to co-occur with a similar set of words (shared collocates). At the same time, we will consider ways in which these forms differ from each other in terms of the company they keep (unique collocates). Taken together, these different steps in our analysis will allow us to illustrate a methodological approach that can be used to narrow down attestations of linguistic forms to those performing a specific speech act, such as that of apologies, and therefore to open up new avenues for future speech act research in large corpora.

The Speech Act of Apology

The speech act of apology is generally regarded as being of considerable importance and it has recently been described as “perhaps one of the most ubiquitous and frequent ‘speech acts’ in public discourse and social interaction” in a Special Issue

on Apologies in Discourse (Drew et al. 2016: 1). It is the omnipresent nature of apologies in our lives, where “we are (almost) all the givers and receivers of apologies, on an almost daily basis” (Drew et al. 2016: 2), and their inherent link to politeness norms that makes them a very significant means of linguistic expression on a social and cultural level. This is because the utterance of an apology can be regarded as remedial action that is taken to acknowledge a breach of social or cultural norms and express regret (see e.g. Fraser 1981; Wierzbicka 1987).

As a consequence of their importance to human interaction, apologies have been the focus of many studies in linguistics approaching the topic from a variety of perspectives. Thus, studies have discussed apologies and language learning (e.g. Flores Salgado 2011; Mulamba 2011), they have specifically investigated the importance of cross-cultural awareness (e.g. Kondo 2010), and they have engaged in the analysis of differences in apologising between specific languages and politeness cultures (e.g. Tanaka et al. 2008; Ogiermann 2009). At the same time, studies have focused on different types of data, approaching apologies in spoken data, such as English telephone calls (e.g. Drew et al. 2016 and the papers appearing in this Special Issue), the London-Lund Corpus of Spoken English (LLC, Aijmer 1996), or the spoken component of the British National Corpus (BNC, Deutschmann 2003), but also in written data, for instance by exploring the use of apologies in online media, such as in emails (Harrison and Allton 2013), on Twitter (Page 2014), or in blogs, as in the current study.

As stated by Blum-Kulka and Olshtain (1984: 206), “apologies are generally *post-event acts*”, indicating that a certain event has already happened and something is presupposed. Coulmas (1981: 71) refers to apologies as a reactive speech act as “[t]hey are always preceded (or accompanied) by a certain intervention in the course of events calling for an acknowledgement”. That is to say that apologies tend to occur in second position and presuppose that the speaker assumes a negative or unwanted intervention to have occurred. While this may be described as the prototypical context in which apologies are attested, anticipatory or *ex-ante* apologies also occur when the speaker assumes that what they are going to say is undesired or disruptive. This, for example, includes dispreferred answers and requests such as asking for clarification or repetition, and assigns apologies a disarming, softening but also attention-getting function (see also Aijmer 1996: 98–100).

Apologies are an expressive speech act, similar to thanking or praising (Searle 1976: 12–13, 1979: 15–16), in that they pertain to the level of the speaker’s feelings. Contrary to many other (expressive) speech acts, such as complimenting, apologies are characterised by a fairly routinized set of prototypical forms that are generally used when carrying out the speech act of apologising in Present Day English. That is to say that a comparatively small number of constructions is used to signal the illocutionary force of an apology, contributing to the formulaic nature of the speech act and accounting for “the great majority of the explicit apologies” (see e.g. Holmes 1990: 175).

Of the routine expressions used to apologise, *sorry* is “the overwhelming favorite” (Meier 1998: 216; see also Owen 1983: 65; Blum-Kulka and Olshtain 1984: 206–207; Wierzbicka 1987: 217). Thus, Holmes (1990: 175) found *sorry* to

be used in 79% of all apology attestations in her New Zealand data. In Aijmer's study of the LLC (1996: 84), *sorry* accounted for almost 84% of attestations and is therefore described as an unmarked routine form. On the other hand, the explicit and unambiguous form *apologize* has been found to be infrequent and to mainly occur in "formal, written, and professional interactions" (Meier 1998: 217, see also Owen 1983: 63). In addition to *apologize* and *sorry*, studies on apologies have also investigated standard routine constructions comprising variants of the forms *afraid*, *apology*, *excuse*, *forgive*, *pardon* and *regret* (see e.g. Blum-Kulka and Olshtain 1984; Holmes 1990; Meier 1998; Deutschmann 2003).

Due to this set of routine expressions that are prototypically associated with the speech act of apology and the fact that in English speakers only rarely apologise without using one of these IFIDs (see e.g. Holmes 1990: 175; Ogiemann 2009: 93–95), it has been claimed that identifying apologies is "relatively easy" (Deutschmann 2003: 36). Deutschmann (2003), therefore, based his study of apologies in a 5 million word sub-section of the spoken component of the BNC on a list of eight IFIDs and their variants (*afraid*, *apologise*, *apology*, *excuse*, *forgive*, *pardon*, *regret* and *sorry*). He searched for these forms in the BNC and then examined his search output manually to determine which of them functioned as explicit apologies, resulting in a total of 3070 examples of apologies. Thus, Deutschmann used corpus linguistic tools but had to resort to manual microanalysis to filter out forms not serving an apologetic function. Harrison and Allton (2013) in their study of apologies in a 1.8 million word corpus of emails and Page (2014) in a 1.6 million word corpus of tweets adopted a similar approach. In our study, we introduce ways of reducing the rather time consuming step of manually checking the extracted examples for their pragmatic function by referring to shared and unique collocates, as we will discuss further below.

Corpus Pragmatics and the Study of Speech Acts

The marriage between the theoretical framework of pragmatics and the methodology of corpus linguistics can still be regarded as a more recent union in the study of linguistics. This is reflected in handbooks on corpus linguistics published in the last 10 years. Thus, pragmatics does not, for instance, feature at all in Teubert and Krishnamurthy (2007), whereas Rühlemann inquires what a corpus can tell us about pragmatics in *The Routledge Handbook of Corpus Linguistics* (O'Keeffe and McCarthy 2010) and both pragmatics and historical pragmatics are discussed as two areas of linguistic study subject to corpus analysis in *The Cambridge Handbook of English Corpus Linguistics* (Biber and Reppen 2015).

Indeed, corpus pragmatics is a field of study that has gained increased attention over the last decade (see e.g. Romero-Trillo 2008; Jucker et al. 2009; Jucker 2013; Aijmer and Rühlemann 2015) and scholars have investigated ways in which the two areas of study could be combined. This pertains in particular to addressing the question of how specific pragmatic phenomena can be searched for with corpus linguistic software given that they often represent linguistic functions rather than forms, as is the case with speech acts, which can "not be searched for directly in

large computerised corpora” (Jucker and Taavitsainen 2014b: 257). In fact, due to the nature of speech acts in general and apologies in particular, previous studies tended to revert to other methodological means to obtain data, such as role plays or discourse completion tests, or they implemented more ethnographic approaches (see Meier 1998: 225; Deutschmann 2003: 15–17; Félix-Brasdefer 2010; Drew et al. 2016: 3, and sources cited therein). As a consequence, the data were constructed rather than naturally occurring and therefore limited to a certain extent.

However, previous corpus-based studies of speech acts were also faced with certain limitations: for instance, analyses were based on smaller corpora, approached speech acts in an eclectic way, investigating speech act verbs only (e.g. Taavitsainen and Jucker 2007) or scrutinising typical forms or patterns of a speech act (e.g. Aijmer 1996; Deutschmann 2003; Adolphs 2008), or they focused on metacommunicative expressions, studying both performative and discursive uses of speech acts (Jucker et al. 2012; Jucker and Taavitsainen 2014b). That is largely because automated searches of electronic corpora which have not been pragmatically annotated rely on specific linguistic forms or phrases that can be input into search software. Taavitsainen and Jucker (2008: 10) even go as far as to state that “[c]omputerized searches for specific speech acts can only be undertaken if the speech act tends to occur in routinized forms, with recurrent phrases and or with standard Illocutionary Force Indicating Devices”, as is the case, for instance, for the speech act of apologising.

Nevertheless, even the study of speech acts that are characterised by a set of routinized forms is not without difficulty when approaching larger electronic corpora. First and foremost, one is met with the problem of polysemy that reduces the precision of automated searches for speech acts (cf. Jucker 2013; Jucker and Taavitsainen 2014b: 258), even when basing one’s study on a set of prototypical IFIDs, that is words and phrases that are generally associated with a specific speech act. For example, the form *sorry*, which has a very strong link with the speech act of apologising, may appear as part of a noun phrase, such as *my sorry self*, where it clearly does not serve the function of an apology. In the case of apologies, studies based on electronic data have therefore often resorted to some sort of manual microanalysis to separate pragmatic from non-pragmatic uses, as discussed in the “[Speech Act of Apology](#)” section.

In this study, we suggest an innovative combination of different types of collocational analysis to allow for the study of speech acts in large corpora, where manual interventions are generally unfeasible.¹ Similar to previous studies on the speech act of apology, we start out from a set of standard routine expressions that we want to investigate further in our corpus of blog posts and comments. We therefore adopt a form-to-function mapping approach, which Jucker (2013) lists as one of three main approaches to data analysis in the field of corpus pragmatics. At the same time, we are only interested in one specific function of the forms in question and that is their function as IFIDs signalling the speech act of apology,

¹ Note that Adolphs (2008) also referred to collocation to distinguish between the different functions of the speech act expression *why don't you* but she only focused on this one construction rather than comparing several forms, as is the case in the current study, and she did not study shared or unique collocates.

which aligns our study with a function-to-form mapping approach (see Jucker 2013). In order to combine the two and thereby facilitate the automatic analysis of corpora of a considerable size, we will demonstrate how collocation can be used to separate different functions of a form and thereby improve the precision of the search output.

In particular, we will adopt a three stage process in the study of apologies in blogs. First we will define the collocational profile of each of the IFIDs studied. In a second step, we will compare the individual collocational profiles with each other to arrive at a list of collocates that they share. Finally, we will use these insights to identify those collocates that are unique to specific forms. It is through this study of shared and unique collocates that attestations of apologies can be singled out from an initial pool of polysemous findings and the search output can be narrowed down to one with a higher incidence of relevant pragmatic functions.² Consequently, this study bridges the gap between manual, close reading analyses of smaller data samples and automated quantitative searches of large corpora by discussing means of corpus linguistic analysis that increase the accuracy of the results for larger data samples while keeping the level of manual intervention to a minimum. In the next two sections, we describe the data that formed the basis of our analysis and explain which IFIDs we used as a starting point of our study of apologies in blogs.

Data

This study is based on the BBC (2010), a diachronically-structured collection covering the period 2000–2010 and freely searchable through the WebCorpLSE at <http://www.webcorp.org.uk/blogs>. In our analysis of apologies, we focus on a 181 million word sub-corpus of the BBC downloaded from the WordPress and Blogger hosting sites³ (see Kehoe and Gee 2012 for a detailed description of how the BBC was constructed).

Our sub-corpus includes both blog posts (95 million words) and reader comments on these posts (86 million words), opening up new possibilities for pragmatic analysis. While the comment function of blogs allows readers to react or respond directly to posts, they do not have to do so immediately as the medium allows for a theoretically infinite time lag between post and comment, thereby enabling a certain degree of modulated interactivity or “interaction-at-one-remove” (Nardi et al. 2004: 46). At the same time, there may be no comment or response at all, which is to say that blogs allow for interactivity while not inherently requesting it, distinguishing this asynchronous means of computer mediated communication from other types of online exchanges, such as chat-room conversations.

Given their potential for interactivity and the fact that blogs exhibit features associated with spoken language and communicative immediacy (see Koch 1999),

² Given that this study is based on a finite set of search forms and uses automated corpus-searches, it cannot claim to provide a complete picture of the use of apologies in blogs, as it will naturally miss any attestations not using one of the IFIDs in question or not lexically signalled at all. In Lutzky and Kehoe (2017), we show, however, that shared and unique collocates can also be used to uncover further uses of apologies in online data, such as the form *oops*.

³ WordPress: <http://www.wordpress.com/>; Blogger: <http://www.blogspot.com/>.

they lend themselves in particular to studies in pragmatics, which traditionally engaged in the analysis of “contemporary, everyday spoken language” before branching out to “all forms of communicative language use” (Jucker and Taavitsainen 2014a: 7). By taking a closer look at language use in this online medium we aim to gain further insights into identifying the speech act of apology in online data.

IFID Selection

In constructing our initial word list, we followed Deutschmann (2003) and included the eight core apology IFIDs: *afraid*, *apologise*, *apology*, *excuse*, *forgive*, *pardon*, *regret* and *sorry*. These were lemmatised to include all relevant inflections (see also Deutschmann 2003: 18), as indicated below:

- **sorry**
- **pardon/pardons/pardoned/pardoning**
- **excuse/excuses/excused/excusing**
- **afraid**
- **apologise/apologises/apologised/apologising** (with both *s* and *z*)
- **forgive/forgives/forgave/forgiving**
- **regret/regrets/regretted/regretting**
- **apology/apologies**

In the following, when using the form in bold type we refer to all variants of the IFIDs listed above. Table 1 gives the raw and scaled frequencies of these forms in the BBC, and a comparison of our figures with those extracted by Deutschmann from his BNC sub-corpus. It should be pointed out that while Deutschmann’s figures refer to manually extracted attestations of apology IFIDs, the BBC figures comprise all automatically extracted attestations of the forms, which is to say that we do not yet distinguish between apology and non-apology uses at this stage. All searches in Table 1 and the remainder of the paper are case-insensitive.

Although cross-corpus comparisons are far from an exact science, it is worth noting that the overall rate of attestation is strikingly similar in both corpora, at around 59 per 1000 words. At 55.6% (or 33 per 1000 words), the proportion of the total occupied by *sorry* in the BBC sub-corpus is more in line with Deutschmann’s study (59.3%, 35.4 per 1000 words) than, for instance, with Holmes (1990: 174) at 75% or Aijmer (1996: 84) at 84%. The proportion of the total occupied by each of the other forms in our corpus is also similar to Deutschmann’s, with a few notable exceptions: the much higher frequency of *pardon* in the BNC sub-corpus, and the higher frequencies of *afraid* and *forgive* in the BBC sub-corpus. We would suggest that the first of these discrepancies is down to differences in corpus content. As already noted, Deutschmann’s BNC sub-corpus contains transcribed speech. In his manual analysis of the data (2003: 78) he classifies 78.3% of all instances of *pardon* as ‘hearing offences’, of which there are three types: “Not hearing, not understanding, not believing one’s ears” (2003: 64). Whilst the second and third of these are to be found in our corpus, the main ‘not hearing’ category is not possible in

Table 1 Corpus frequencies of (potential) apology IFIDs in BBC and BNC sub-corpora

Corpus size	BBC sub-corpus			BNC sub-corpus		
	181,405,634			5,139,082		
IFID	Freq.	Freq. per 1000	%	Freq.	Freq. per 1000	%
sorry	59,864	33.0	55.6	1820	35.4	59.3
afraid	14,270	7.9	13.3	63	1.2	2.1
excuse	10,759	5.9	10.0	320	6.2	10.4
forgive	6596	3.6	6.1	15	0.3	0.5
apologise	5591	3.1	5.2	36	0.7	1.2
regret	5536	3.1	5.1	1	0.0	0.0
apology ^a	3859	2.1	3.6	0	0.0	0.0
pardon	1127	0.6	1.0	815	15.9	26.5
Total	107,602	59.3		3070	59.7	

^a Although Deutschmann (2003) includes ‘apology’ in his list of IFIDs on p. 18, when he repeats the list on p. 50 this IFID is omitted. It is also omitted from all frequency counts elsewhere in the publication. We include ‘apology’ and its variants in our analysis nonetheless

written data and this is the likely reason for the difference in frequencies between corpora.

The other example, *afraid*, is one which highlights both the limitations of the methodology described thus far and the potential benefits of collocational analysis. Perhaps more so than any other IFID in the list, *afraid* is a highly polysemous word, with uses beyond the speech act of apology. Therefore, it is particularly challenging to identify apology uses of this form in a large corpus such as the BBC, where the total number of occurrences is so high that it does not allow for careful manual analysis to the same extent as adopted by Deutschmann (2003) for the BNC. While we could restrict our study to a random sample of attestations in the corpus, one of the aims of this research is to determine whether it would be possible to use collocation to filter out non-apologies automatically and to demonstrate how this could be achieved by using innovative approaches to collocational analysis. In the next section we describe the different steps of analysis that have been combined in order to move closer to the aim of automated speech act searches.

Collocational Analysis

We started the collocational analysis of the potential apology IFIDs in question by establishing a collocational profile for each of them individually. That is to say that we used WebCorpLSE to extract the top 100 collocates for each IFID at span 4 (i.e. a window of four words to the left of the IFID and four words to the right) to gain further insight into their general collocational behaviour. Table 2 lists the top 25 collocates of the IFID *apologise* as an example.

Table 2 Collocates of the IFID *apologise* (span 4)

Collocate	Collocate frequency	Co-occurrence frequency	z-score
profusely	348	105	101.70
advance	4308	131	100.89
e-mails	1027	72	66.49
lack	13,373	100	51.50
absence	3132	58	46.97
sincerely	1407	49	43.85
delay	1385	38	33.81
need	146,208	233	32.17
questions	24,276	78	28.92
inconvenience	400	28	26.28
usually	32,151	82	25.85
posting	21,150	55	21.37
first	19,764	52	20.90
quality	14,169	42	19.93
confusion	2733	24	19.21
behalf	2170	22	18.15
responding	1797	20	16.81
sooner	3488	21	15.88
late	30,091	49	14.97
behavior	6155	23	14.96
publicly	1863	18	14.95
again	130,378	113	13.90
offended	2132	17	13.80
humbly	531	15	13.47
length	5368	19	12.69

Table 2 is sorted by z-score, a measure of statistical significance which takes into account the frequency of the node (the IFID) and of each collocate in relation to corpus size. For instance, *profusely* is a relatively rare word (frequency 348) which, nevertheless, co-occurs with *apologise* 105 times. That is to say that over 30% of instances of *profusely* in our corpus appear within four words to the left or right of *apologise*. This collocational pair is given a high z-score as a result. In fact, all but three of the 105 co-occurrences of *apologise* and *profusely* are at span 1 (i.e. ‘profusely apologise’ or ‘apologise profusely’), and the only other words that collocate significantly with *profusely* are inflections of *thank*, *bleed* and *sweat*. We also see evidence in Table 2 of other semi-fixed phrases (‘apologise in advance’, ‘apologise for any inconvenience/confusion’, ‘apologise publicly’/‘publicly apologise’, etc.) thus adding further details to the collocational tendencies of the verb *apologise*. Building collocational profiles such as the one illustrated in Table 2 for

each of the eight potential apology IFIDs separately allowed us to engage in the next step of our collocational analysis, the study of shared collocates.

Shared Collocates

Shared collocates are, as the name implies, collocates which are shared by several lexemes. To arrive at a list of shared collocates, we used the collocational profiles produced for each of the eight potential IFIDs and taking each form in turn, we compared its top 100 collocates with the top 100 collocates of all the other IFIDs combined. As a result, we were able to uncover those collocates which are shared by the eight forms and gain further insight into their overlapping functions and meanings.

Figure 1 presents the results of our shared collocate analysis. Each column in Fig. 1 represents one of the eight (potential) apology IFIDs, each row represents a collocate, and shaded boxes indicate where IFIDs share collocates. For example, the first collocate, the first-person pronoun *I*, is shared by seven of the eight IFIDs (all except *apology*).

It is worth noting here that the fact we use a span of four words in our collocational analysis lessens the impact of grammatical restrictions on the results. Thus, *pardon* is allowed to collocate with *I*, largely in the phrase ‘I beg your pardon’. Given that the IFID *apology* also includes the plural, there is no grammatical reason why *I* does not collocate significantly with this IFID at span 4 (‘I offer my (sincere) **apologies**’ would be counted at span 4 for example).

This second step in our collocational analysis, the study of shared collocates as represented in Fig. 1, allows us to make several preliminary observations. Thus, Fig. 1 shows that some of the eight potential apology IFIDs exhibit greater similarities on a collocational level than others, which is visible by looking down the columns to determine how many of the collocates each IFID shares (see for instance the gaps for *afraid* and *regret* towards the top). This provides a good first indication as to their function as, when appearing with one of the shared collocates listed, the form in question is likely to serve an apology function and to act as an apology IFID. This is in particular the case for those collocates offering a clear indication of the reasons people apologise in the BBC sub-corpus, with a *lack*, *delay* or *absence* of something being particularly prominent (each a shared collocate of five or six of the apology IFIDs). Taking the IFID *apologise* as an example, there are 196 instances in our corpus of the words *lack*, *delay* or *absence* appearing within four words of this IFID. Table 3 gives ten examples extracted randomly using the ‘filter’ option in WebCorpLSE.

In the vast majority of examples in Table 3, the writer is apologising for his or her poor blogging etiquette, be it a lack of updates, delay in posting or general absence. Most of these examples are performative speech acts, as indicated by the use of the first person pronouns (cf. Taavitsainen and Jucker 2008: 22). The only exceptions can be found in concordance lines 1 and 8, which refer to or report on other people apologising.

Moreover, we see a set of shared collocates in Fig. 1 where writers in the blog corpus appear to be apologising for the poor quality of their written expression:

	sorry	pardon	excuse	afraid	apologise	forgive	regret	apology	
I									7
to									6
not									6
my									6
lack									6
for									6
delay									6
but									6
absence									6
please									5
if									5
being									5
you'll									4
will									4
spelling									4
sorry									4
so									4
say									4
poor									4
no									4
never									4
english									4
any									4
you									3
typos									3
that									3
she									3
saying									3
reply									3
quality									3
missed									3
late									3
ignorance									3

Fig. 1 Shared collocates (span 4) of the eight potential apology IFIDs

spelling, *english*, *typos* (and both *poor* and *quality* are themselves also shared collocates of three or four IFIDs each). This partly reflects the mixed native language background of blog writers and commenters, who are not necessarily native speakers of English due to the diverse demographics of online users, as, for

Table 3 Examples of *lack*, *delay* and *absence* as span 4 collocates of the IFID *apologise*

-
- 1: Sometimes I am annoyed when blog people **apologize** for their **absence** ... and I'm not so much as apologizing as just letting you know why there is radio silence on my end because I have M's family visiting [*post*]
 - 2: And finally, I just want to **apologize** for my **lack** of commenting on blogs this week [*post*]
 - 3: First off, I would truly like to **apologize** for the **lack** of posts since about August when school started [*post*]
 - 4: And I have to **apologize** for the **delay** in getting this post up [*post*]
 - 5: I've been cooking and I **apologize** for my **lack** of updates about that but I did make a pancetta and pasta dish as well as crab cakes, a yummy dessert, a vanilla cake and cupcakes [*post*]
 - 6: First, I must **apologize** for my appalling **lack** of presence in the blogosphere of late [*post*]
 - 7: First of all, I'd like to **apologize** for my **absence** in the last several days [*post*]
 - 8: 'We **apologise** for the **delay** to our arrival into Clapham Junction,' the man said over the loudspeaker system, 'this is due to very slight pause—'no particular reason' [*comment*]
 - 9: I **apologize** for my long **absence** and thank all of you who have kept in touch [*post*]
 - 10: I **apologize** for the **delay** in posting these images of the Solemn Requiem Mass according to the Dominican Rite celebrated at the Church of the Blessed Sacrament in Seattle WA on All Souls last [*post*]
-

instance, example 7 in Table 4 illustrates explicitly. Table 4 contains a random selection of the 70 occasions in the BBC sub-corpus where *spelling*, *English* or *typos* appear as a span 4 collocate of the IFID *excuse*. Note that only one of the examples is a nominal use of *excuse* (line 9).

We have indicated at the end of each concordance line in Table 4 whether the example appears in a blog post or in a reader comment on a post (information which appears in the WebCorpLSE interface when the user clicks on a particular concordance line). It is clear that the majority of examples here—all but three—appear in comments. In fact, 77% of the total matches for this query appear in comments, in

Table 4 Examples of *spelling*, *english* and *typos* as span 4 collocates of the IFID *excuse*

-
- 1: PS I slapped this puppy up fast, so, as per usual, please **excuse** the **typos** and nonsensical nature of this post [*post*]
 - 2: **Excuse** any **spelling** mistakes no time to check! [*post*]
 - 3: I am really tired, and just finished work and I'm sitting in a pretty loud cafe right now, so you'll have to **excuse** any **spelling** or sentence structure mistakes ... I'll probably re-read this in a few days and hate it and re write parts of it ... [*post*]
 - 4: **Excuse** my **English** is not good too [*comment*]
 - 5: To add a little—and **excuse** any **typos**—am in huge hurry [*comment*]
 - 6: and i think you need a kindle its a cool digital book that you you can buy books on(please **excuse** my miss **spelling** [*comment*]
 - 7: **excuse** the poor **english** its not my native language [*comment*]
 - 8: **excuse** me my bad **English** [*comment*]
 - 9: In fact there are a couple of glaring **typos**. My **excuse** is lack of in-browser spellcheck on the antique computer at the library [*comment*]
 - 10: **excuse** me for my **english!** [*comment*]
-

contrast to the *apologise + (lack, delay, absence)* query in Table 3 where only 17% of matches appeared in comments. Thus, the analysis of shared collocates has also revealed a medium-specific distribution of different types of apologies in blog posts and comments. While bloggers tend to apologise for infrequent updates or delays in posts, in comments their readers refer to the (often poor) standard of their language skills when apologising. This was also reflected in our previous analysis of the form *oops* (Lutzky and Kehoe 2017), where we found that commenters often apologise in a second comment and explicitly refer back to some infelicity on the level of content or language use that occurred in their first comment.

The examples given in this section have thus demonstrated that the analysis of shared collocates provides us with an overview of apologies in blogs, which would not have been possible if we had focussed on a single form only, and uncovers certain medium-specific types of apologies. Furthermore, Fig. 1 shows that there are gaps in the columns for some of the potential IFIDs, such as *afraid* and *regret*. While *apologise*, *apology*, *excuse* and *forgive* share many collocates with each other and with the other apology IFIDs, *afraid* and *regret* demonstrate less overlap. We therefore turn to the study of their unique collocates in the next section in order to examine the specific meanings of potential apology IFIDs in more depth and, after discussing some of the collocates they share, to find out what distinguishes them from each other.

Unique Collocates

While the study of shared collocates has pointed us in the direction of attestations showing the illocutionary force of an apology (such as the collocates indicating the reasons for an apology), it is the study of unique collocates that either suggests where a potential apology IFID may have non-apology uses or provides additional details as to the apologetic uses of a form, i.e. allowing us to see how potential apology IFIDs differ from each other. Figure 2 shows the top 20 unique collocates of each potential apology IFID at span 4, sorted by strength of collocation with that form (z-score). These are words which collocate strongly with a particular IFID but do not appear amongst the top 100 collocates of any of the other eight IFIDs studied.

Of the eight forms given above, four have a strong link with the speech act of apologising and are prototypically associated with it: *apology*, *apologise*, *excuse* and *sorry*. As Fig. 2 shows, the noun *apology* and the verb *apologise* behave very similarly in our data. The unique collocates of these IFIDs include several adjectives and adverbs, many of which express that the illocutionary force is genuine: *heartfelt*, *deepest*, *grovelling*, *profuse(ly)*, *sincere(-ly, -st)*, *humbl(ely)*, *unreservedly* and *publicly*. The unique collocates of *apologise* also offer a further indication of some of the common reasons for apologising in our corpus: (asking) *questions*, *mistake*, *fault*, *tone*. The following is an example of the last of these, taken from a comment in which the author is apologising for a previous comment he made two hours earlier (before the author of the original post has had the chance to respond):

I just reread my own comment. That was some rant, wasn't it! I do **apologize** for the **tone**, if not the content.



Fig. 2 Unique collocates (span 4) of the eight potential apology IFIDs (top 20 for each)

There is also a set of unique collocates relating to a specific kind of apology—*issued, public, formal*—as the below examples illustrate:

What is really interesting about Lazare's book is how he examines successful and unsuccessful **public apologies** (often those of politicians or celebrities).

I know it is very difficult for my son to do this, and his dad doesn't expect a **formal apology**, it is enough that we can forget about it and get on with life.

An **apology** was **issued** about the quality of the beverages, but not about the effects of too much caffeine and being stuck in a TV studio.

In contrast with *apology/apologise* and its strong collocates *sincere(ly)* and *heartfelt*, several of the unique collocates of *excuse* are words indicating the opposite: *lame, pathetic, flimsy, weak* (though it does also collocate with *good, valid* and *legitimate*). This IFID functions as both a noun and a verb, although the nominal uses dominate the unique collocates.

The unique collocates of *sorry* are perhaps the most striking of all in Fig. 2 as they clearly point to the more colloquial uses of this form compared to the ones discussed above. There are several terms of endearment, contractions and other informal, speech-like features which seem to be expressing sympathy, often used by a reader leaving a comment and referring back to something mentioned in the post. These include two spellings each of *oops* and *hon* (short for ‘honey’), as well as *aww*, *hugs*, *(a)bout* and *sucks*. In fact, when studying *oops* and its various spelling variations in more detail, we found that it is attested with apologetic uses too and can be regarded as an apology IFID in blogs, both when co-occurring with other IFIDs and when attested on its own (Lutzky and Kehoe 2017).

The remaining four forms—*regret*, *pardon*, *forgive* and *afraid*—take up the peripheral positions in Fig. 2, as their use as apology IFIDs is also more peripheral compared to the four prototypical forms discussed above. While we cannot discuss each of them in detail, we will briefly mention *forgive*, *pardon* and *regret* below, before focussing on *afraid* in our demonstration of how unique collocates allow us to separate pragmatic from non-pragmatic uses of the form.

We noted in our initial analysis (Table 1) that the relative frequency of *forgive* is higher in the BBC sub-corpus than in Deutschmann’s BNC data, accounting for 6.1% of all eight forms taken together as opposed to 0.5% in the BNC sub-corpus. Figure 2 gives an indication of the reasons for this discrepancy, with the vast majority of the top 20 unique collocates of *forgive* in our corpus arising from its use in religious contexts, more specifically in the case of *trespasses* in the Lord’s Prayer. These are contexts far less likely to occur in Deutschmann’s spoken corpus. As we explained in the “IFID Selection” section, the relative frequency of *pardon* is much lower in our written corpus than in Deutschmann’s spoken one, likely as a result of the absence of ‘hearing offences’. Nevertheless, *pardon* does have specific uses in our corpus—both nominal and verbal—which are reflected in its unique collocates: *beg (your) pardon*, *pardon the expression*, *pardon my French*, *receive/given a Presidential pardon*.

Regret is an example where the unique collocates help to draw out multiple, often subtly different, uses of the same word. Although *regret* can be used in direct apologies (‘I regret to inform you’, ‘it is with regret that...’, to regret one’s actions towards another person), it is also possible to regret actions or decisions which affect only oneself. In addition, there is the wider use of *regret* to refer to *sadness* or *sorrow* (both unique collocates of this form).

Turning to *afraid*, we see evidence of non-apologetic uses in the form of things that people are commonly afraid—or frightened—of: *heights*, *spiders*, *losing*, (the) *dark*, *bugs*, *failure* and (perhaps less commonly) *clowns*. We also see the modifier *deathly*, which collocates significantly with this meaning of *afraid* but is unlikely to appear in an apology context. Our suggestion here is that the extraction of unique collocates such as these could be a useful means of automating the pragmatic analysis of large corpora. Taking our example of the potential apology IFID *afraid*, if we find one or more of the words *heights*, *spiders*, *losing*, *dark*, *bugs*, *failure*, *clowns* or *deathly* within four words to its left or right, this gives a strong indication that the particular instance we are considering is not actually part of an apology. Table 5 gives a random selection of the 439 examples of the form *afraid* that would be excluded were this filter to be applied.

Table 5 Examples of the form *afraid* excluded by collocate filtering

-
- 1: I am 32 and **afraid** of **clowns** [*post*]
 - 2: He is very **afraid** of **heights** [*post*]
 - 3: I imagine it is similar to being **afraid** of the **dark** [*post*]
 - 4: I'm **afraid** of **losing** faith in this story [*post*]
 - 5: Charles Laughton who played Captain Bligh was actually **deathly afraid** of the ocean and was seasick the whole time they were filming [*post*]
 - 6: Truth is, I'm not so great at change, quite shy socially when I don't know too many people, and **deathly afraid** of embarrassing myself [*post*]
 - 7: I had worked so hard at becoming "strong", and was so **afraid** of **losing** that and actually relying on another human being, that I pushed him away [*post*]
 - 8: I'm **afraid** that **spiders** are the one thing I'm terrified of though I, too, have a catch and release policy—mostly someone else has to do the catching and releasing though! [*comment*]
 - 9: It took me a long time to realize I was so **afraid** of **failure**, I was holding myself back [*comment*]
 - 10: Seventeen-year-old Vanessa Sands is **afraid** of everything—the **dark**, **heights**, the ocean—but her fearless older sister, Justine, has always been there to coach her through every challenge [*post*]
-

These examples demonstrate that the collocates selected are good indicators of non-apologetic uses of *afraid*, with *deathly* as perhaps the best indicator. Line 10 actually contains two of the collocates (with *heights* included in the span 4 window as a result of the missing space between *everything* and *the*). The only borderline case in Table 5 is line 8, where *afraid* collocates with *spiders* but it is actually the word *terrified* that is used to convey fear. The writer is, in effect, using the word *afraid* with an apologetic function to introduce the confession that they are scared of spiders.

In addition to these concepts that people are afraid of, the unique collocates also include the preposition *of* as a strong and unique collocate of *afraid*, illustrated by several of the examples in Table 5. In fact, the total number of occurrences of *afraid of* is 3587. Together with the constructions *afraid to* (3189) and *afraid for* (122), these amount to almost half of all attestations of *afraid* and can be separated as non-apologetic uses of *afraid* (see also Owen 1983: 88–89). Furthermore, according to the OED (s.v. *afraid*, adj. and n.), one can narrow down the (performative) apologetic uses of *afraid* to the constructions *I am/I'm afraid* + *dependent clause* or parenthetical attestations. The total number of *I am/I'm afraid* occurrences in our corpus is 4873. This count is case insensitive and includes the following spelling variations: *I'm afraid* (3670), *I am afraid* (969), *i'm afraid* (125), *i am afraid* (46), *Im afraid* (32), *im afraid* (17), *I'M AFRAID* (3), *I'm Afraid* (3), *I'M afraid* (2), *I AM AFRAID* (2), *I AM afraid* (2), *I am AFRAID* (1), and *I'm AFRAID* (1).

When combining the two filters, narrowing down the search to variations of *I am afraid* (4873) and excluding attestations collocating with one of the prepositions *of*, *to* or *for* in the first position to the right of *afraid* (870), we are left with a total of 4003 attestations. Here, compared to the initial output of 14,270 attestations, the salience of *afraid* as an apologetic IFID is much higher and comprises examples like those randomly extracted in Table 6.

Table 6 Examples of *afraid* remaining after collocate filtering

-
- 1: Sarah—yeah, **I'm afraid** that's true [comment]
 - 2: All an alien world to me, **I'm afraid**, and I can only observe that there are some beers which leave one's glass half-empty and some which leave it half-full [comment]
 - 3: Thanks so much for spreading the word, Lizzy! **I'm afraid** the hashtag has changed: it'll be bodytower, to avoid a flood of diet/fitness tweets. So sorry for the confusion [comment]
 - 4: Anyway, the last thing I wanted to hear about was Mum banging on about hitching her cart to Roger's wagon, so when she said: "I had hoped you might be pleased for me ..." I just put my ipod on. I needed to focus, and **I'm afraid** Mum just had to be a casualty of war at that point [post]
 - 5: Laura—yes, glutinous rice flour is entirely different from rice flour **I'm afraid**. [comment]
 - 6: awwwwwwwwww ... sorry babes ... **im afraid** i havent heard any of his songs(i probably have but i dunno. u make him sound like he was really great..may his soul RIP.. [comment]
 - 7: you have quite literally not grasped what is being said, and your replies have become increasingly incoherent and angry. **i am afraid** i shall have to ignore you henceforth [comment]
 - 8: I remember Ives Gammell saying that and I am sure it is a quote. But **I am afraid** I don't know whose [comment]
 - 9: I have contacted the hospital twice but with no success. **I am afraid** all records have been destroyed. All that is left are pictures like these, the stories and our imaginations. [comment]
 - 10: **Im afraid** there are still no pictures as i am having to wait until payday to buy a camera. But i can show you my newest listing in my Etsy Shop it called Lavender hills and it's genuine Swarovski Crystal and sterling silver. Pretty and very eye-catching [post]
-

This random sample reflects that *afraid*, when excluding the explicitly fear-based examples, tends to occur in comments on blog posts. Of the 4003 examples remaining after filtering, 2685 (67%) are in comments and 1318 (33%) are in posts. Thus, when appearing with an apologetic function, *afraid* shows a distribution similar to *excuse* discussed above (see “[Shared Collocates](#)” section), being used primarily in the comment section of blogs, where it forms part of the interaction between different commenters and the author of the blog post and is often accompanied by a specific form of address (e.g. *Sarah*, *Lizzy*, *babes*).

Overall, our discussion of unique collocates has shown that some of them provide us with further information about the specific apologetic uses of a form, such as the differences in the adjectives and adverbs collocating with the forms *apologise*, *apology* and *excuse*. Additionally, unique collocates allow insights into medium-specific uses of IFIDs, as in the case of *pardon* used in blogs. On the other hand, we have seen how some unique collocates provide a clear indication as to the non-apologetic uses of a form (see e.g. *forgive* or *afraid*), allowing us to distinguish apology from non-apology attestations. In combination with other pieces of information (e.g. *I am* as a signal of performative apologetic uses of *afraid*), they therefore facilitate the exclusion of unwanted hits and contribute to an improvement in the precision of search output.

Conclusion

To date, the study of speech acts has made use of a range of methodological approaches. Thus, many studies have been based on constructed examples obtained through discourse completion tests or role plays (see e.g. Meier 1998: 225; Félix-

Brasdefer 2010, and sources cited therein). On the other hand, the importance of naturally-occurring examples has been foregrounded over the last two decades and several studies have employed a corpus-based methodology. In doing so, they were, however, met by a number of limitations entailed by studying a functional phenomenon using corpus linguistic tools. As a consequence, scholars resorted to smaller data samples (e.g. Koester 2002; Garcia McAllister 2015), eclectic analyses of common forms or patterns associated with a speech act (see e.g. Aijmer 1996; Deutschmann 2003; Taavitsainen and Jucker 2007; Adolphs 2008), or they referred to metacommunicative expression analysis (see e.g. Jucker et al. 2012; Jucker and Taavitsainen 2014b), all of which generally demanded stages of manual microanalysis to separate unwanted hits from examples with specific pragmatic functions.

Our current study has opened a window on new methodological possibilities in corpus-based speech act analysis that is needed in corpus pragmatics as it allows us “to see the larger picture” and move “from the individual richly contextualized example to large numbers of similar examples” (Jucker 2013). By combining different types of collocational analysis, we have shown how the precision of the search output can be improved and the level of manual analysis reduced for studies of speech acts in large corpora. While smaller corpora may be amenable to manual intervention, this is not the case for larger corpora, where individual search terms may return several thousand attestations (cf. *afraid* in the BBC). In order to allow for the speech act analysis of these larger datasets, we have introduced the combined study of shared and unique collocates. By investigating similarities and differences in the collocational profile of several IFIDs, we have shown that functional overlaps but also divergences can be revealed, which can in turn be used to increase the incidence of relevant examples in the search output.

While we agree with Jucker (2013) that “corpus pragmatic research is never a case of automatic processing” only, our methodological approach allowed us to streamline the search for the fairly routinized speech act of apology in our blog data. Although initial steps at extending this approach to other expressive speech acts have already been taken (see e.g. Lutzky and Kehoe 2016), further studies are needed to see how transferrable this methodology is to other types of speech acts that may be less formulaic. As Clancy and O’Keeffe (2015: 251) state, “[w]hile the study of pragmatic items can be challenging in a corpus, it is eminently possible” and it is equally possible in large corpora. Instead of shying away from big data when it comes to the study of functional phenomena, we need to offer new means of analysis to overcome limitations and our current study is a crucial step in this direction.

Acknowledgements Open access funding provided by Vienna University of Economics and Business (WU).

Open Access This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

References

- Adolphs, S. (2008). *Corpus and context. Investigating pragmatic functions in spoken discourse*. Amsterdam: John Benjamins.
- afraid, adj. & n. OED online. December 2016. Oxford University Press. <http://www.oed.com/view/Entry/3628?redirectedFrom=afraid>. Accessed 6 Feb 2017.
- Aijmer, K. (1996). *Conversational routines in English: Convention and creativity*. London: Longman.
- Aijmer, K., & Rühlemann, C. (2015). *Corpus pragmatics. A handbook*. Cambridge: Cambridge University Press.
- Biber, D., & Reppen, R. (2015). *The Cambridge handbook of English corpus linguistics*. Cambridge: Cambridge University Press.
- Birmingham Blog Corpus (BBC). (2010). Compiled by the Research and Development Unit for English Studies at Birmingham City University. <http://www.webcorp.org.uk/blogs>. Last accessed July 1, 2016.
- Blum-Kulka, S., & Olshtain, E. (1984). Requests and apologies: A cross-cultural study of speech act realization patterns (CCSARP). *Applied Linguistics*, 5(3), 196–213.
- Clancy, B., & O’Keeffe, A. (2015). Pragmatics. In D. Biber & R. Reppen (Eds.), *The Cambridge handbook of English corpus linguistics* (pp. 235–251). Cambridge: Cambridge University Press.
- Coulmas, F. (1981). ‘Poison to your soul’. Thanks and apologies contrastively viewed. In F. Coulmas (Ed.), *Conversational routine. Explorations in standardized communication situations and prepatterned speech* (pp. 69–91). The Hague: Mouton.
- Deutschmann, M. (2003). *Apologising in British English*. Umeå: Umeå University.
- Drew, P., Hepburn, A., Margutti, P., & Galatolo, R. (2016). Introduction to the special issue on apologies in discourse. *Discourse Processes*, 53(1–2), 1–4.
- Félix-Brasdefer, J. C. (2010). Data collection methods in speech act performance. DCTs, role plays, and verbal reports. In A. Martínez-Flor & E. Usó-Juan (Eds.), *Speech act performance: Theoretical, empirical and methodological issues* (pp. 41–56). Amsterdam: John Benjamins.
- Flores Salgado, E. (2011). *The pragmatics of requests and apologies: Developmental patterns of Mexican students*. Amsterdam: John Benjamins.
- Fraser, B. (1981). On apologizing. In F. Coulmas (Ed.), *Conversational routine. Explorations in standardized communication situations and prepatterned speech* (pp. 259–271). The Hague: Mouton.
- Garcia McAllister, P. (2015). Speech acts: A synchronic perspective. In K. Aijmer & C. Rühlemann (Eds.), *Corpus pragmatics. A handbook* (pp. 29–51). Cambridge: Cambridge University Press.
- Harrison, S., & Allton, D. (2013). Apologies in email discussions. In S. C. Herring, D. Stein, & T. Virtanen (Eds.), *Pragmatics of computer-mediated communication* (pp. 315–337). Berlin: Walter de Gruyter.
- Holmes, J. (1990). Apologies in New Zealand English. *Language in Society*, 19(2), 155–199.
- Jucker, A. H. (2013). Corpus pragmatics. In J.-O. Östman & J. Verschueren (Eds.), *Handbook of pragmatics online*. Amsterdam: John Benjamins.
- Jucker, A. H., Schreier, D., & Hundt, M. (Eds.). (2009). *Corpora: Pragmatics and discourse. Papers from the 29th international conference on English language research on computerized corpora (ICAME 29)*. Amsterdam: Rodopi.
- Jucker, A. H., & Taavitsainen, I. (2014a). Diachronic corpus pragmatics. Intersections and interactions. In I. Taavitsainen, A. H. Jucker, & J. Tuominen (Eds.), *Diachronic corpus pragmatics* (pp. 3–26). Amsterdam: John Benjamins.
- Jucker, A. H., & Taavitsainen, I. (2014b). Complimenting in the history of American English. In I. Taavitsainen, A. H. Jucker, & J. Tuominen (Eds.), *Diachronic corpus pragmatics* (pp. 257–276). Amsterdam: John Benjamins.
- Jucker, A. H., Taavitsainen, I., & Schneider, G. (2012). Semantic corpus trawling: Expressions of ‘courtesy’ and ‘politeness’ in the Helsinki Corpus. In C. Suhr & I. Taavitsainen (Eds.), *Developing corpus methodology for historical pragmatics (Studies in variation, contacts and change in English 11)*. Helsinki: VARIENG. http://www.helsinki.fi/varieng/series/volumes/11/jucker_taaitsainen_schneider/.
- Kehee, A., & Gee, M. (2012). Reader comments as an aboutness indicator in online texts: Introducing the Birmingham Blog Corpus. In S. Oksefjell Ebeling, J. Ebeling, & H. Hasselgård (Eds.), *Aspects of*

- corpus linguistics: Compilation, annotation, analysis (Studies in variation, contacts and change in English 12)*. Helsinki: VARIENG. http://www.helsinki.fi/varieng/series/volumes/12/kehoe_gee.
- Koch, P. (1999). Court records and cartoons. Reflections of spontaneous dialogue in early Romance texts. In A. H. Jucker, G. Fritz, & F. Lebsanft (Eds.), *Historical dialogue analysis* (pp. 399–429). Amsterdam: John Benjamins.
- Koester, A. J. (2002). The performance of speech acts in workplace conversations and the teaching of communicative functions. *System*, 30, 167–184.
- Kondo, S. (2010). Apologies: Raising learners' cross-cultural awareness. In A. Martínez-Flor & E. Usó-Juan (Eds.), *Speech act performance: Theoretical, empirical and methodological issues* (pp. 145–162). Amsterdam: John Benjamins.
- Lutzky, U., & Kehoe, A. (2016). *Your blog is (the) shit*. A corpus linguistic approach to the identification of swearing in computer mediated communication. *International Journal of Corpus Linguistics*, 21(2), 165–191.
- Lutzky, U., & Kehoe, A. (2017). “Oops, I didn’t mean to be so flippant”. A corpus pragmatic analysis of apologies in blog data. *Journal of Pragmatics*. doi:10.1016/j.pragma.2016.12.007.
- Meier, A. J. (1998). Apologies: What do we know? *International Journal of Applied Linguistics*, 8(2), 215–231.
- Mulamba, K. (2011). The use of illocutionary force indicating devices in the performance of the speech act of apology by learners of English as a foreign language. *Journal of the Midwest Modern Language Association*, 44(1), 83–104.
- Nardi, B. A., Schiano, D. J., Gumbrecht, M., & Swartz, L. (2004). Why we blog. *Communications of the ACM*, 47(12), 41–46.
- O’Keefe, A., & McCarthy, M. (2010). *The Routledge handbook of corpus linguistics*. London: Routledge.
- Ogiermann, E. (2009). *On apologising in negative and positive politeness cultures*. Amsterdam: John Benjamins.
- Owen, M. (1983). *Apologies and remedial interchanges. A study of language use in social interaction*. Berlin: Mouton.
- Page, R. (2014). Saying ‘sorry’: Corporate apologies posted to Twitter. *Journal of Pragmatics*, 62, 30–45.
- Romero-Trillo, J. (Ed.). (2008). *Pragmatics and corpus linguistics. A mutualistic entente*. Berlin: Mouton de Gruyter.
- Searle, J. (1976). A classification of illocutionary acts. *Language in Society*, 5(1), 1–23.
- Searle, J. (1979). *Expression and meaning. Studies in the theory of speech acts*. Cambridge: Cambridge University Press.
- Taavitsainen, I., & Jucker, A. H. (2007). Speech act verbs and speech acts in the history of English. In S. Fitzmaurice & I. Taavitsainen (Eds.), *Methods in historical pragmatics* (pp. 107–137). Berlin: Mouton de Gruyter.
- Taavitsainen, I., & Jucker, A. H. (2008). Speech acts now and then. Towards a pragmatic history of English. In A. H. Jucker & I. Taavitsainen (Eds.), *Speech acts in the history of English* (pp. 1–23). Amsterdam: John Benjamins.
- Tanaka, N., Spencer-Oatey, H., & Cray, E. (2008). Apologies in Japanese and English. In H. Spencer-Oatey (Ed.), *Culturally speaking. Culture, communication and politeness theory* (pp. 73–94). London: Continuum.
- Teubert, W., & Krishnamurthy, R. (2007). *Corpus linguistics: Critical concepts in linguistics*. London: Routledge.
- Wierzbicka, A. (1987). *English speech act verbs. A semantic dictionary*. Sydney: Academic Press.