

13

Pragmatics

Brian Clancy and Anne O’Keeffe

1 Introduction

Corpus pragmatics is a methodological framework that allows for the interpretation of spoken or written meaning, with an emphasis on providing empirical evidence for this interpretation (see O’Keeffe *et al.* 2011). It is a relatively recent development within the field of corpus linguistics and interest in this “subfield” has blossomed as spoken corpora have become more readily available. Meaning is an elusive concept to say the least but what is clear is that participants in interaction, especially those engaged in spoken discourse, negotiate meaning through a series of sometimes almost barely perceptible “clues” that are supplied by the participants themselves, their shared knowledge (both personal and cultural) and the situation in which the interaction takes place. Given that classical pragmatics has its roots in the philosophy of language, traditionally, the study of pragmatics has employed an interpretative methodology in order to account for this negotiation of meaning. Therefore, many of the illustrative examples are invented rather than “attested” or “in use.” Corpus linguistics has emerged as a sympathetic methodological companion for the study of pragmatics providing researchers with representative samples of real-life language in use, and an attendant empirical tradition.

Corpus pragmatics is distinct from other fields in corpus linguistics. However, in common with other fields, corpus pragmatics investigates the co-textual patterns of a linguistic item or items, which encompasses lexico-grammatical features such as collocation or semantic prosody. However, where corpus pragmatics’ “added value” lies is in its insistence that these patterns be considered in light of the *context* – the situational, interpersonal, and cultural knowledge that interactional participants share. Through an iterative process, corpus pragmatics therefore moves beyond important but surface observations of lexico-grammatical patterns to allow a more nuanced interpretation of these patterns taking into

consideration who uses them, where they were used, for what purposes, and how this use has changed over time. In this way, corpus pragmatics has retained in part its original interpretative nature but has endeavored to supply this interpretation with objective supporting evidence.

We contend that the studies critically examined here exemplify many of the strengths of corpus pragmatics. We examine some of the current concerns of the field in key concept areas such as speech acts, pragmatic markers, and pragmatics and power. Although the majority of the research concentrates on pragmatic features of spoken language, we also include studies that highlight the importance of corpus pragmatics to the written context.

2 The state of the art in corpus pragmatics

The blend of corpus linguistics and pragmatics, though a relatively recent development, is a mutually beneficial one. Therefore, the area is ripe with research opportunities. One of the latest (and most fruitful) synergies of pragmatics and corpus linguistics is the area of *historical corpus pragmatics* which is primarily concerned with the diachronic study of speech acts (see, for example, Jucker and Taavitsainen's 2008a edited volume). This research has proven especially beneficial in addressing some of the difficulties associated with using a corpus-based methodology. Primary among these is that speech act form and function do not directly correlate and, therefore, speech acts are not automatically retrievable through the use of corpus software (see Adolphs 2008; Rühlemann 2010). The pragmatic tagging of corpora would be a major advancement, and this is underway. Recent historical corpus pragmatic research has seen the development of sophisticated methodologies (Kohnen 2008) and pragmatic annotation schemes (Wichmann and Culpeper 2003) which allow for the examination of utterance-by-utterance interaction between conversational participants. Moreover, the release of corpora such as SPICE-Ireland (Kallen and Kirk 2008), a pragmatic and discourse-annotated version of the Irish component of the *International Corpus of English*, opens up a wealth of research paths given that, amongst other advantages, it is one of the first corpora to provide a searchable system where prosodic and pragmatic information sit side-by-side. The application of these annotation schemes in other corpora and other contexts, and also an examination of their robustness, is one area of potential research.

There is also a substantial body of research building up within corpus pragmatics around the area of pragmatic markers. Much of the seminal work in this area has been led by Karin Aijmer. Aijmer (2013) focuses on pragmatic marker variation with regard to social, cultural, and regional factors. Her rationale for exploring the variability of these markers in the ICE-GB corpus (the British component of the *International Corpus of English*) is

to contribute to the discussion surrounding the influence of context on the function and meaning of pragmatic markers. A possible focus for future research in this area is the continued development of the cross-cultural analysis of the functions of pragmatic markers. For example, Clancy and Vaughan (2012) highlight how corpus pragmatics can be employed to characterize a language variety's pragmatic system. They found that the marker *now* was markedly more frequent in Irish English than in British English due to the fact that it has additional pragmatic functions in Irish English – it is used as both a pragmatic marker and deictic presentative. It is also important to note that research on pragmatic markers is not limited to native-speaker usage. Aijmer and Simon-Vandenberg (2006) bring together a volume focusing cross-linguistically on pragmatic markers, while Fung and Carter (2007) explored the use of pragmatic markers in two pedagogical corpora – a corpus of learners of English in Hong Kong, and the pedagogic subcorpus from the *Cambridge and Nottingham Corpus of Discourse in English* (CANCODE). Fung and Carter found that markers, while present in the student corpus, are generally less frequent than in British English, especially in relation to those whose function is interpersonal, for example, *you know*, *well*, *sort of*, or *yeah*.

Corpus pragmatics is also to the forefront in exploring the link between language, power, and ideology. Baker and McEnery (2005: 223) highlight the usefulness of corpus pragmatics for critical social research. They point out that corpora are beneficial to fields of study such as Critical Discourse Analysis as “by looking at the collocational strength of lexical items in a corpus of general language, we are given an objective sense of the themes and associations that are embedded in words due to their continual pairing with other words.” This research yielded, amongst other findings, the co-occurrence of *refugee* with metaphors associated with the movement of water such as *flood*, *stream*, or *swell* – words that frequently have negative connotations in normative patterns of language use. In relation to power and the workplace, Holmes and Stubbe (2003) explore the tension associated with the need to get things done at work while maintaining good collegial relationships. Using the Language in the Workplace Corpus, they examine a number of speech acts typical of the context, such as advice and instruction, in addition to aspects of workplace discourse such as humor and small talk. On humor, Vaughan and Clancy (2011: 51) suggest it is a “powerful, polyvalent pragmatic resource” and that it requires much more focused research in the realm of corpus pragmatics and power relations.

Corpus pragmatics has also advocated the synergy of corpus linguistics and conversation analysis in order to investigate the organizational level of pragmatics, that which explores turn-taking phenomena such as pauses, overlaps, interruptions, or backchannels (Schneider 2012). Clancy and McCarthy (2014) have focused on the benefits of corpus linguistics in the study of the co-construction of speaker turns. Co-construction is an inherently pragmatic activity given that the concept ‘hinges on the notion that

meaning is created through the interaction itself, in the specific context of that interaction' (Kereckes 2007: 1943). Three high-frequency items identified as traditionally involved in the co-constructions process, *if*, *when*, and *which*, were identified and their patterns explored through an examination of concordance of these in turn-initial position based on their reported tendency to occur at points of co-construction. Corpus researchers habitually "enter" a corpus using items identified as high frequency on the corpus frequency list. Wichmann (2004) presents a potential avenue for future research when she suggests that researchers need to focus on low-frequency phenomena in addition to high-frequency items.

Finally, in relation to deixis, personal pronouns feature prominently in corpus frequency lists, especially spoken ones. Personal pronouns are strongly associated with deictic reference: a system of reference that facilitates contextual orientation. The use of personal pronouns to negotiate identity has received some attention in corpus pragmatics. O'Keeffe (2006) looks at deictic centering and othering in corpora of media discourse, particularly through the pronouns *we* and *they*. Vaughan and Clancy (2013) examine the complexity of reference of *we* in establishing community identity in both workplace and family discourse. Hyland (2002a, b) explores the use of first-person pronouns in two academic written corpora – an "expert" corpus and a "novice" corpus. He finds that academic writing is not impersonal or faceless, as it has often been portrayed in textbooks and style guides, and that expert writers use first-person pronouns three times more frequently than novice writers. He cautions against novice writers avoiding these pronouns as it may result in them not establishing an effective authorial identity. However, in saying this, deixis remains, as noted by Levinson (2004: 97), "one of the most empirically understudied core areas of pragmatics" and, although there remains much work to be done, researchers in corpus pragmatics appear to be at the forefront of the efforts to redress this discrepancy.

With the state of the art in corpus pragmatic research established, we now turn to a more fine-grained discussion of exemplar studies in the areas outlined above.

3 A critical discussion of previous research at the forefront of corpus pragmatics

3.1 Speech acts

The study of speech acts outside the field of corpus pragmatics has predominantly been based on elicited data generated from discourse completion tasks (DCTs) or role plays (O'Keeffe *et al.* 2011). As already stated, corpus pragmatics is characterized by the analysis of real-life language in use. The use of corpus pragmatics approaches has led to a reconsideration

of the results produced by elicited data in light of corpus-based findings. For example, in a study of DCTs and corpus data, Schauer and Adolphs (2006) compare and contrast the results yielded by corpus data, in this case CANCODE, and DCTs in relation to speech acts of gratitude in order to determine if they can be used in conjunction with one another to inform teaching materials. At an “actional level”¹ (Schneider 2012: 1027), in both the DCT and corpus data, expressions of gratitude involving the *thank* stem and the item *cheers* were the most frequent. Where the results diverge, however, is at an interactional level, i.e. “how speech acts combine into larger units of discourse ranging from adjacency pairs to conversational phrases” (ibid.). At this level, the corpus data demonstrates how *cheers* is not employed primarily as an expression of gratitude but as a response to an expression of gratitude, “used in this way it is difficult to determine whether it actually marks gratitude or whether it functions as a discourse marker that signals the end of the encounter or discourse episode” (Schauer and Adolphs 2006: 125). This possible discourse-marking function is absent in the DCT data due to the non-interactive nature of the data they produce. Another difference noted between corpus and DCT data is the length of speaker turn in which the expression of gratitude was produced. The corpus data highlighted, in addition to expressions of gratitude most frequently used, that these expressions tend to cluster across extended speaker turns as part of a process of collaborative negotiation, and that this pragmatic use of *thanks*, for example, is often found in gate-keeping encounters such as service encounters. Finally, the corpus data highlighted the importance of the ability to produce an expression of gratitude twinned with a polite refusal, a *thanks but no thanks*, which was very infrequent in the DCT data but flagged by Schauer and Adolphs as “one of the main skills that students may need to possess in a native speaker context” (2006: 129). Although we are not suggesting here that DCTs are unsuitable for the study of pragmatics, we are once again highlighting the suitability of corpus linguistics as a complementary methodology due to the nature of the data contained in corpora.

3.2 Pragmatic markers

As outlined in Section 2 above, pragmatic markers have received a large amount of recent attention. The study we have chosen here illustrates the complementary use of large and small corpora to investigate the pragmatic markers *I would say* and *I'd say*. Farr and O’Keeffe’s (2002) first step was to examine the occurrences of *I would say* and *I'd say* at a dialectal level using three large spoken corpora: the *Limerick Corpus of Irish English* (LCIE), CANCODE, and a corpus of American spoken data from the *Cambridge International Corpus* (CIC). Using purely quantitative data in the form of

¹ Schneider’s (2012) “actional level” represents the level of speech act analysis.

frequency lists, they found that, at this level, Irish speakers seem to be twice as “hedgy” as their American counterparts. However, an intervarietal, quantitative study such as this, is, they maintain, “restrictive in its insightfulness” (p. 29), in that it does not further any understanding we have of how, where, or why *would* is used as a hedge.

Farr and O’Keeffe recommend that the use of *would* as a hedge is investigated at the level of register in order to more fully appreciate its use. Therefore, its occurrence in two small corpora – a 55,000-word corpus of radio phone-in data and a 52,000-word corpus of post-observation teacher trainee interaction – was examined. They discovered, through an exploration of colligational and collocational patterns, that *would* is used in these institutional domains to redress asymmetry, to mitigate face-threatening acts or to “transpose” the focus of the talk to a hypothetical “safe band” (2002: 41). This treatment of *would* in small corpora at the level of register also demonstrated the different levels of context that have an influence on the use of a marker in that the results for *would* in the institutional context of the two small corpora are rooted in their sociocultural context – in this case within Irish society. Farr and O’Keeffe suggest that forwardness is not valued in Irish culture and that Irish society, in general, “does not place a high value on powerful or direct speech” (2002: 42), something that has been borne out in subsequent research into the Irish system of pragmatic marking (see Vaughan and Clancy 2011; Clancy and Vaughan 2012). Of particular note here are the contextual insights provided by the use of small corpora in corpus pragmatic research. These insights could not have been retrievable using a larger, more general corpus where the user usually has little contextual information of any depth. The effectiveness of small corpora in corpus pragmatics will be echoed in our empirical study presented below.

3.3 Language, power, and ideology

Critical discourse analysis (CDA) has successfully utilized corpus linguistics as a complementary methodology. The purpose of CDA is to show how language in use is often constructed and shaped by various social forces (see Fairclough 1989). Although it can also be applied to the study of spoken language, it is particularly associated with written language, especially media discourse. CDA has employed corpus linguistics to engage with texts such as newspapers (as per the example presented here) in order to bring to the fore how power is enacted through “inferred and indirect linguistic devices” (Wodak 2007: 204), such as the pragmatic concepts of *presupposition* and *implicature* (see Archer *et al.* 2012: Unit A12). Both implicature and presupposition rely on the determination of meaning through inference, i.e. beyond linguistic form. This is a process exemplified by Baker *et al.*’s (2008) analysis of the semantic prosody of *POSE* as presented here.

Baker *et al.* (2008) analyzed a 140-million-word corpus of UK tabloid and broadsheet newspapers, composed of daily and Sunday papers, both local and national, to investigate linguistic representations of and attitudes toward refugees, asylum seekers, immigrants, and migrants (RASIM). Corpus linguistics enabled them to map these data, “pinpointing areas of interest for a subsequent close analysis” (2008: 284). CDA has traditionally focused on grammatical features such as passivization or metaphor, however, one of the benefits of corpus linguistics for CDA is that it facilitates a focus on lexical patterns also. Once these lexical patterns had been identified quantitatively by, for example, keyword lists or cluster analysis, concordance lines were generated to allow subsequent qualitative examination. Take, for example, their corpus analysis of the semantic prosody of the multi-word item *POSE as*. Semantic prosody refers to a process of implicature through the tendency of words to “appear in particular environments in such a way that their meaning, especially their connotative and evaluative meaning, is spread over several words ... words might tend to occur in overwhelmingly positive or negative environments” (O’Keeffe *et al.* 2007: 14). A detailed analysis of the concordance lines in relation to *POSE as* revealed eight frames of use of the type “Actor(s) pose as X to achieve Y,” which revealed that the item can have a negative, positive, or neutral semantic prosody according to the actor and the context. For example, “RASIM posing as doctors/sports fans/tourists in order to gain entry/find work/receive benefits,” represents a negative prosody whereas “reporters posing as RASIM in order to investigate their plight in an asylum system” has a positive prosody. Baker *et al.* found that despite the tabloids using *POSE as* almost eight times more frequently in relation to RASIM, the traditional stereotype that tabloid newspapers are more negative in their portrayal of RASIM than broadsheets did not hold true. Generally, what they found was that tabloids do adopt a negative stance in their reporting on issues related to RASIM and broadsheets tend to adopt a more balanced perspective, which does, however, combine both positive and negative reporting. In this way, corpus linguistics assists CDA researchers in quantifying notions like bias. Therefore, it is the synergy of the methodologies that lends this study its analytic strength.

3.4 The organization of discourse

Corpus pragmatics has also successfully combined the methodological field of conversation analysis (CA) with that of corpus linguistics in order to provide a much more fine-grained analysis of spoken language than would be possible if each were used in isolation. This synergy of methodologies allows linguistic items to be examined at both a structural (syntactic) and interpersonal (pragmatic) level and enables us to understand how “words, utterances and text combine in the co-construction of meaning” (Walsh, 2013: 37). In one study of this type, McCarthy (2003) analyzes the

occurrence of a number of non-minimal response tokens in two corpora; 3.5 million words of the CANCODE corpus and a similarly sized sample of the CIC corpus. The corpus results provide a profile of the most frequent tokens used, for example, *right*, *wow*, *true*, *exactly*, or *sure*, thereby pointing toward these as salient items for further analysis. These results also indicate a shared set of non-minimal response tokens that occur within the core 2,000-word frequency lists for both British and American English. The more fine-grained, up-close view provided by the more qualitative CA approach means that we can determine their role in discourse and McCarthy demonstrates how examining non-minimal response tokens in their local context of use reveals how this role is essentially pragmatic in nature. On the structural level, he presents the accepted view that response tokens function structurally to design and organize discourse, however, he also reveals their role as “indexes of engaged listenership” (2003: 59). This is particularly the case when tokens occur as doublets or triplets (for example, *Lovely. Terrific. or Good. Good. Good.*). These patterns do not simply signal discourse boundaries but “inject a strong relational element of response to the situation (one of satisfaction, agreement and positive social bonding)” (2003: 54). In this way, listeners can be seen to be attending to pragmatic concerns – both social and affective – in their responses perhaps before attending to structural concerns such as taking the turn for themselves.

3.5 Deixis

The final corpus pragmatic work that we examine here is Rühlemann's (2007) unique study of the conversational subcorpus of the BNC. This study is significant because it analyzes frequent conversational features that previously had not been systematically researched using corpus linguistics. Furthermore, the study positions deixis as one of the cornerstones of the analysis. Rühlemann argues that conversation is characterized by a much greater wealth of shared context than most written language situations. In order to quantitatively and qualitatively prove this, Rühlemann examines, amongst other features, person, place, and time deixis – shared-context phenomena that are manifest in spoken corpora in general. One of the most obvious and frequent manifestations of person deixis is personal pronouns. Regarding the distribution of these pronouns, Biber *et al.* (1999: 333) have shown that *I* and *you* are far more common in casual conversation than in other registers such as academic prose. Rühlemann (2007: 66–69) posits four reasons for the preferred use of *I* and *you* in casual conversation; (i) *I* is prone to repetition (*I* is repeated at a frequency of about 200 times per million words in conversation (see Biber *et al.*, 1999: 334)); (ii) *I* and *you* have a high frequency of collocation, especially with cognitive verbs, for example, *I think* and *you know*; (iii) speakers in conversation show a clear tendency to prefer a direct mode than an indirect mode

and, importantly for studies in corpus pragmatics, (iv) conversation is co-constructed, with speakers taking turns and each new turn requiring the reconstruction of the new speaker's deictic system. Based on the evidence provided by personal pronouns, but also on the analysis of the role of deixis in the tense system, speech reporting, and vocatives, Rühlemann concludes that our deictic system demonstrates a "remarkable flexibility" (2007: 221), with participants seamlessly projecting, oscillating and varying their deictic centers.

In addition to exploring verbal shared context, Rühlemann also uses corpus techniques to examine non-verbal pragmatic items. In modern spoken corpora in general, a wealth of paralinguistic information is tagged, such as coughing or door slamming, much of which is of little importance, however, some of these features, such as laughter, are of significance to corpus pragmatics. Rühlemann uses frequency counts to demonstrate the importance of laughter to conversation, for example, if "between-speech laughter" is considered as a linguistic item in and of itself, it would be placed in 29th position on the BNC conversational subcorpus frequency list. A closer inspection of the contexts in which laughter occurs demonstrates that it fulfils core pragmatic functions. According to Rühlemann, it functions as a backchannel, and, as such, is an indicator of engaged listenership akin to McCarthy's (2003) non-minimal response tokens. It also has a discourse deictic function as a discourse marker which signals that word choices should not be understood too literally. Laughter is not always associated with humorous contexts but is concomitant with incongruity. The frequency of laughter in the BNC, Rühlemann argues, demonstrates the intrinsic importance of this incongruity to human interaction.

Although these studies illustrate the state of the art in corpus pragmatics and showcase the field's methodological and analytical rigor, corpus pragmatics is, however, not without its thorns. As already mentioned, corpus linguistics has been criticized in relation to its suitability for the study of speech acts. Moreover, if corpus pragmatics is concerned with the interpretation of meaning in context, another disadvantage associated with the relationship between corpus linguistics and pragmatics is that many larger corpora are impoverished both textually and contextually (Rühlemann 2010). Koester (2010: 66–67) points out that in large corpora it is "very difficult, if not impossible" to connect what is said with its original context. In terms of spoken language, one of the main contributing factors is that in order to analyze conversation, corpus builders first "translate" spoken language into writing in the form of transcripts, many of which exclude important contextual features such as prosody. This results in what Rühlemann (2007: 13) refers to as a "written-speech paradox." One solution to these challenges has emerged from the increasing use of small corpora in corpus pragmatics. We strongly advocate small corpora as ideally suited for corpus pragmatics given their "constant interpretative

dialectic between features of texts and the contexts in which they are produced" (Vaughan and Clancy 2013: 70). Researchers working with these corpora are frequently both compiler and analyst (and, indeed, often participant), which gives them a unique insight into context. Small corpora also allow every instance of the item(s) under investigation to be mined, as illustrated by our case study here, resulting in refined and nuanced pragmatic profiles of the items under investigation.

4 Case study: a corpus pragmatic analysis of vocative function

This corpus-based investigation of vocative use focuses on two small corpora in an Irish English context. The results presented here are a combination of those from McCarthy and O'Keeffe (2003) and Clancy (forthcoming). The first of the two corpora utilized is a 55,000-word corpus of a daily Irish radio program *Liveline* which is broadcast on the national broadcaster RTÉ and has a listenership of approximately 10 percent of the Irish population. The second corpus, the 12,500-word *SettCorp*, represents the conversations of a six-member (father, mother, two boys, and two girls) middle-class Irish family. Interaction in radio phone-in takes place between people who very often do not know one another, and although the speaker relationship is hierarchical (the media persona has the power), a "pseudo-intimacy" (O'Keeffe 2006) is maintained in order that a greater level of disclosure might be achieved. In contrast, family discourse is characterized by an unequal intimacy (Blum Kulka 1997) – there exists in the family an in-built hierarchy where the parents have more conversational power than the children. Therefore, this case study is concerned with comparing and contrasting the use of vocatives among intimates in family discourse and those who simulate intimacy in radio phone-in. The goal of the study is to contrast how vocatives function within the two corpora, one in an intimate family context, and the other in a pseudo-intimate institutional (radio phone-in) context, in order to better understand how vocatives contribute to the creation and maintenance of intimacy and whether this is replicated in a pseudo-intimate context.

Vocatives, terms of address (Ervin-Tripp 1971), and forms of address (Brown and Gilman 1960) are, according to Leech (1999: 107), "closely related topics which are easily confused." He points out that a term of address "is a device used to refer to the addressee(s) of an utterance" whereas a vocative is a particular kind of address term: "a nominal constituent loosely integrated with the rest of the utterance." Formally, a vocative has a number of guises: endearments (*honey, baby, love*, etc.), kin titles (*Mammy, Daddy*, etc.), familiarizers (*mate, man, folks*, etc.), first names familiarized (*Brad, Jen*, etc.), full first names (*Bradley, Jennifer*, etc.), title and surname (*Mr. Holmes, Dr. Watson, Professor Moriarty*, etc.), honorifics (*sir*,

ma'am, etc.) and a group called "other" which includes nicknames in addition to some complex noun phrases (*Those of you who want to bring your pets along, please sit in the back of the space ship ...* (Leech 1999: 111)). Functionally, a vocative behaves peripherally in an utterance in a similar way to, for example, a discourse marker. Pragmatically, a vocative performs a range of functions and it is these functions that this empirical study is concerned with. Leech (1999) identifies three pragmatic functions of vocatives: (1) summoning attention, (2) addressee identification, and (3) establishing and maintaining social relationships between conversational participants. Subsequently, McCarthy and O'Keeffe (2003) developed Leech's (1999) functional categories and derived a more nuanced framework containing six functional categories: *relational, topic management, badinage, mitigation, turn management, and summons*. Relational vocatives establish and/or maintain social relations rather than transmit information or services. This category includes compliments and other positive face boosters, general evaluations, phatic exchanges, and ritualistic offers and thanks. Topic management refers to the occurrence of a vocative in an utterance or set of utterances that launch, expand, shift, change, or close a topic. Included here also is what McCarthy and O'Keeffe (2003: 162) have termed *topic validation*, whereby a speaker calls on another conversational participant by name to validate or confirm an assertion. Badinage refers to the use of a vocative in instances of humor, irony, and general banter among participants. Mitigators include those vocatives that help soften or downtone a potential threat to positive or negative face (Brown and Levinson 1987). Turn management vocatives select the next speaker or disambiguate possible recipients in multi-party talk. Summons concerns the use of a vocative to directly summon a conversational participant (these categories are illustrated with examples below). However, we stress that the vocatives themselves might not perform the function but they occur as a "signal of that utterance's intent" (Wilson and Zeitlyn 1995: 8).

4.1 Research question and methodology

This study sets out to answer the following question: what are the similarities and differences in pragmatic function between family discourse and radio phone-in, that is, between their use in real intimate relationships and pseudo-intimate relationships? This examination of the function of vocatives is important for a number of reasons. First, it vividly illustrates, through the application of a corpus-based, bottom-up methodology, the close interrelationship between pragmatic function and context. In doing so, it showcases the usefulness of small corpora for pragmatic research, particularly with regard to how their investigation can deepen our understanding of the descriptive framework for spoken genres. One of the criticisms of small corpus research is that a small corpus does not allow for generalization. However, the encouraging aspect of this research is

that the findings from a number of small corpora in similar contexts can be used to generalize if a number of features emerge as constant. For example, small corpora in institutional contexts, such as those used by McCarthy (2000), O'Keeffe (2006), and Vaughan (2007), have all demonstrated a “blurring of the lines” through the presence of relational talk in this context. Another fundamental benefit of small corpora in pragmatic research is that they allow the researcher to maintain close contact between authentic, naturally occurring language and its context of use (Vaughan and Clancy 2013). Moreover, Leech (1999: 107) notes that vocatives are a “surprisingly neglected” aspect of English grammar and, although it can be argued that there is a robust research tradition in relation to the broader *term of address*, it is still true now that studies utilizing a corpus-based approach to the study of vocatives are under-represented, relatively speaking, especially in comparison to say hedging research (notable exceptions include, however, Shiina 2005 and Busse 2006).

In terms of the methodology utilized, in the case of both corpora, the data were read manually and every vocative classified. Hence, this methodology, if transferred to a larger dataset, needs to employ a sampling strategy so as to arrive at a manageable amount of vocatives. This in itself poses a challenge as vocatives are not normally tagged. In these studies, the vocatives in the datasets were categorized according to vocative type (for example, endearments, kin titles, etc.) and function (relational, summons, etc.) and, although not discussed here, position (initial, medial, or final). Concordance lines were then generated using the categories as the search items so that the function and position of the vocative is placed in relief. Figure 13.1 illustrates a random sample of the concordance lines for the kin title *mam* (node word in bold) in SettCorp where Wordsmith Tools™ has been used to resort the concordance lines two words to the right (2 R) and three words to the right (3 R):

N	Concordance
1	with Quimper? It's in France. Yeah. . Where's Sacramento mam [kt] [turn] [final]? California. Now the warmest place shall
2	and now return to microphone range How far down is it mam [kt] [topic] [final]? You'd need to get it I'd say or you'll be
3	reflect yeah. They refeck. What about some Christmas music mam [kt] [topic] [final]? The reflection. What about? They
4	[final]? Ah we do cos then we can get a Dalmatian can't we Mam [kt] [topic] [final]? Mam's goin getting us. You can't have
5	yourself. It's hard work now. I wipe it off with this then do I mam [kt] [topic] [final]? Hm? I wipe off with this yoke? the dog.
6	over Christmas. Here put the glass ones near lights. Look mam [kt] [sum] [final]. Watch. speaker three laughs the
7	laughs mockingly What's he doin? Leave him alone now. Here mam [kt] [sum] [final]. Will you throw me up a couple of+
8	. It's great isn't it? Yeah yeah. Yeah it looks like it was+ Mam [kt] [sum] [initial] we need the Hoover first. That's okay.
9	size is your shoe? They're ten. Look at the size of him. Mam [kt] [sum] [initial] can I've my shoes back please my feet
10	. Had they lunch in Jury's? No himself and your man had. Mam [kt] [sum] [initial]. Aren't we already twi+= And who was
11	couple of+ Tryin to get the lights up. Look at the dirt of you. + mam [kt] [sum] [stand alone]. B [fn] [sum] [initial] will you
12	he didn't have seventy Americans with him in Jury's yesterday. Mam [kt] [sum] [stand alone]? They were just after goin back.
13	Sacramento. Really? Yeah. And how's he getting on with it? Mam [kt] [sum] [stand alone]? Fine. Well he didn't have
14	you eat the cards. You won't play. You're goin out. What? Mam [kt] [rel] [initial] will you leave me the car? I won't leave
15	five we were. Yeah that's+ Ninety-four ye were in Spain mam [kt] [mit] [final]. +ninety five was the good summer we
16	here. Get that+ I'll get the Chr= +get the shiny silver one mam [kt] [mit] [final]. I will yeah. We need Chris= D (J).
17	speaker four coughs pause of six seconds Do you give up mam [kt] [mit] [final]? Do you know I have loads of these? .
18	now cos Daddy can't take any time off like. Ah he can. Ah Mam [kt] [mit] [initial]. He's gettin his business off the ground
19	you find the start of that there? Come on. Yeah you had it mam [kt] [bad] [final]. . What? Down on your knees bubs
20	[initial] shut up. Just shut up. pause Cards anyone?. Yeah Mam [kt] [bad] [final]. Don't be sayin things like that. .
21	. shouting Good fack off. pointing to a decoration on the tree Mam [kt] [bad] [initial] take off that stupid thing. handing a

Figure 13.1 Concordance lines for *mam* in SettCorp (sorted 2R then 3R)

Table 13.1 Occurrences of vocatives in Liveline and SettCorp, normalized to 10,000 words (raw results in brackets)

	Liveline	SettCorp
Total vocatives	42 (232)	129 (161)

Figure 13.1 demonstrates that tagging and sorting the concordance lines in this manner demonstrates that *mam* functions primarily as a summons vocative [sum] and is, to a lesser extent, involved in topic management [topic], mitigation [mit], and badinage [bad]. In terms of vocative position, 12 of the 21 occurrences are in final position. All topic management functions are final, for example, as are the majority of mitigators, and this perhaps indicates the relationship between these functions and their position in an utterance. Therefore, despite there being only 21 concordance lines, there is evidence of the pragmatic behavior of this vocative (see below for further analysis).

4.2 Findings

Table 13.1 illustrates the initial frequency results for vocatives in the two datasets. For the purposes of comparability, these are normalized to occurrences per 10,000 words.

This finding illustrates that despite the pseudo-intimacy simulated in the radio phone-in context, vocatives are more frequent in intimate casual conversation. Interestingly, this runs contrary to Leech's (1999: 117) findings which note that vocatives do not occur among what he terms "close associates," his study includes mother-daughter and wife-husband interactions, where, he maintains, neither the addressee identification nor the social relationship function is necessary. One of the crucial analytical aspects of corpus pragmatics is to offer compelling quantitative evidence in support of automatically generated, frequency data. Not only this, automatically generated frequency data can be sorted and categorized in order to generate a more in-depth and nuanced quantitative interpretation of the results in Table 13.1, a functional analysis of all vocatives in the two corpora was conducted, as illustrated in Figure 13.2.

Figure 13.2, perhaps unsurprisingly, reveals a contrast in vocative function between the two datasets. The starkest difference in the contextual functional categories is between their use as mitigators in family discourse and call management in radio phone-in. While it is not surprising that vocatives occur frequently in the context of call management on a radio phone-in and that this has no occurrence in the family data, it is striking that their use in the context of mitigation is so low in radio phone-in. It is here that we move to the second step in the analytical process. We have

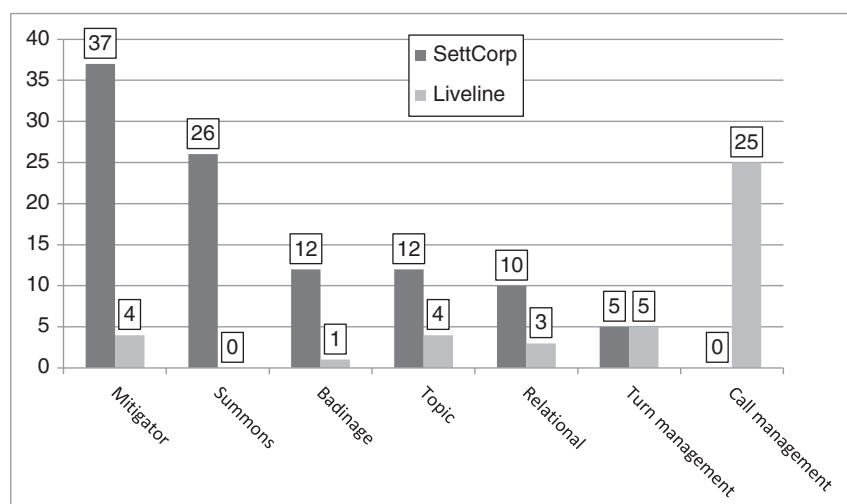


Figure 13.2 Functions of vocatives in *Liveline* and SettCorp (normalized to 10,000 words)

already extracted telling quantitative evidence and now we can synthesize these quantitative insights by taking a qualitative approach. Extract 2 is illustrative of a typical use of a vocative in the context of call management where the radio host identifies the next speaker and cues them to speak by means of a vocative, thus initiating the next caller and his tale:

Extract 2

[Introducing a caller whose son narrowly escaped death from meningitis]

- <\$1> Now to a couple that had very very difficult Christmas this year
however all's well that ends well ah Austin good afternoon to you.
<\$2> Good afternoon Marian.
<\$1> Your little boy went back to playschool yesterday?
<\$2> Yesterday that's right.

Mitigation, the most frequent context of use in family data, accounting for 36 percent of all vocatives in that dataset, is typified by the following example where the family is decorating the Christmas tree:

Extract 3

- <Son> Oh look the state of the one that mam hate mam hates that
because they're+
<Mother> It's awful.
<Daughter> It's rotten.
<Mother> Don't put it up.
<Daughter> It's rotten Jimmy.
<Mother> It's all dirty and everything.

The son locates one particular decoration and it is reasonable to surmise, judging by speaker reactions, that the mother and daughter dislike it. He is

then instructed, in the form of an unmitigated parental directive, not to put the decoration on the tree. The daughter, in the consecutive utterance, mitigates the directive by giving a reason why and softening using a mitigating vocative in the form of a first name familiarized. McCarthy and O’Keeffe (2003) also noted the high frequency of vocatives in the context of mitigation in CANCODE casual conversation data, saying that they are neither syntactically nor semantically necessary and they function solely as pragmatic downtoners of challenges, adversative comments, and disagreements. However, adversarial situations are relatively uncommon in the *Liveline* radio phone-in data. The interaction is usually between the radio host and the caller. However, there are a few instances where two adversaries give their opposing view on a conflictual situation and in these contexts, we can find some examples of vocatives used when mitigating the force of a challenge, etc., as in extract 4:

Extract 4

[Two callers, Colm and Máirtín, are in a heated dispute]

<\$3> Oh Colm come on . . . pull the other one.

The other difference of note in the two datasets is in relation to the use of vocatives in summons. Again this is very context specific. It would suggest that families frequently use vocatives to summon other family members. However, we note that McCarthy and O’Keeffe (2003) found only four percent of vocatives in the context of summons in the casual conversation data between family and friends. This is in contrast to 25 percent of vocatives being used to summon in the Irish family discourse in this study. There are a variety of reasons why this may be the case – for example, there may be a differing age profile in the datasets and this can have an impact as younger children generally demand more attention from their parents than older children. The sampling process could also have a role to play in that if the conversations were recorded in situations where everyone was already present, such as dinner time, then there would be less need to summon. In order to ascertain this, further cross-cultural analysis using data from CANCODE family-only contexts would be needed.

Identified as the largest functional category by McCarthy and O’Keeffe (2003: 160), the relational function in the family data is surprisingly low, accounting for 10 percent of all uses:

Extract 5

<Son 1> Yeah. That’s exactly what I did. Cards anyone?

<Son 2> No thank you Jimmy.

However, the fact that the discourse takes place in this intimate context, coupled with fixed and pre-established speaker relationships and the “politeness license” (Blum-Kulka 1997; Clancy 2005) granted to families,

serves to render the relational function almost redundant. Given the exogenously defined roles and task-focus of radio phone-in the low occurrence of vocatives in this context is possibly not surprising.

Conclusion

All in all, this brief look at two small datasets shows us the usefulness of having data from different contexts which can be used for the empirical analysis of a pragmatic item. Reflecting on the process, we can see that the corpus software aided our analysis but much of it had to be done manually. This is manageable with small datasets. We can also see that by isolating family-only data from an Irish context, we have set up at least one further research question, namely, whether there are cross-cultural dimensions to the use of vocatives (do they function differently in different families in other cultures?). The antecedent work in McCarthy and O'Keeffe (2003) provides data on family and friends and the next step would be to isolate family-only data from other language varieties and cultural contexts to see if there is a different pattern of vocative use. By contrasting the Irish family data with Irish radio phone-in data, we are able to audit contrasting variables such as the differing modes of communication in the two datasets, family discourse being face-to-face and radio phone-in being aural only (via telephone/radio), and see how this impacted on real versus pseudo-intimacy. We could also potentially examine the impact of the contrasts between the, usually, multiparty nature of the family data and the, usually, dyadic nature of the radio phone-in data. The important point to glean from this is the richness of possibility that real corpus data offer in comparison to intuitive data. Corpus data provide a defined context of use and the rigor of metadata as a standard component of corpus design. This kind of research could not be conducted without such real data and their metadata.

Although, relatively speaking, corpus pragmatics is in its early stages, clearly there is already a healthy picture in terms of the application of corpus linguistics to the study of pragmatics. As corpora become more sophisticated in areas such as their pragmatic tagging, it is predicted that they cannot be ignored as research tools by those researching pragmatics. Indeed, we speculate that in years to come, much, if not all, pragmatics research will involve corpus linguistics. However, in the meantime, corpus pragmatics can do more to show the real worth of methodology to the wider field of pragmatics. We need to better articulate our methodology and its rationale and make corpora more user-friendly for those researchers unfamiliar with them. It needs to be explained that pragmatic research which truly focuses on language in use has to be corpus-based because there is no convenient way of making speakers say what you want them to say where language is unpredictable, messy, and extended over many

turns. While the study of pragmatic items can be challenging in a corpus, it is eminently possible. It requires an iterative approach, whereby the research goals need to be refocused with reference to the data and where the research question is continually refined by going deeper and deeper into the data. It sometimes requires the researcher to manually read and code speech acts or speech events such as pragmatic markers but there are many rewards for this (see, for example, Clancy and Vaughan 2012). In addition to the richness of the results, and the surprises they hold, the research creates a really useful and reusable resource through tagging and coding. Pragmatic tagging, while time-consuming and challenging, is the holy grail for this field.