

Jonathan Culpeper, Isolde van Dorst and Mathew Gillings

10 A corpus-based exploration of British English impoliteness formulae

Abstract: (Im)politeness is often said not to be inherent in linguistic forms (e.g., Van der Bom and Mills 2015). If that is true, impoliteness formulae should not exist. This study scrutinises the set of British English conventionalised impoliteness formulae described in Culpeper (2010, 2011). It does so by constructing powerful corpus queries to retrieve those formulae from the spoken component of the British National Corpus 2014. Those queries tap into the words, structures and semantics of the formulae. The success or otherwise of these queries is taken as a proxy for whether those formulae have clearly defined clusters of formal features. Such clusters are not, of course, evidence of impoliteness. Consequently, this study also establishes the degree to which instances of those formulae appear in impoliteness contexts. If they do, that would be evidence that they could become conventionalised for impoliteness. Given the subjective nature of impoliteness, the study uses inter-rater reliability techniques in order to secure robust judgements about whether the contexts of occurrence are really impolite. Five impoliteness formulae variants were established as having clearly defined formal features and strong associations with impoliteness contexts: personalised negative vocatives, one variant of personalised negative assertions, two variants of dismissals, and silencers. Taboo words did not seem to influence the association with impoliteness contexts, though our evidence was not strong. More abstract, more grammatical impoliteness formulae seemed to attract less impoliteness. The study also investigated banter and metalinguistic cases of impoliteness in relation to impoliteness formulae.

Keywords: impoliteness, impoliteness formulae, corpus linguistics, conventionalisation, banter, metalinguistic impoliteness

Jonathan Culpeper, Lancaster University, Bailrigg, LA1 4YW, United Kingdom,
e-mail: j.culpeper@lancaster.ac.uk

Isolde van Dorst, WU Vienna University of Economics and Business, Welthandelsplatz 1, 1020 Vienna, Austria, e-mail: isolde.van.dorst@wu.ac.at

Mathew Gillings, WU Vienna University of Economics and Business, Welthandelsplatz 1, 1020 Vienna, Austria, e-mail: mathew.gillings@wu.ac.at

1 Introduction

This chapter explores various impoliteness formulae in the Spoken British National Corpus 2014 (Spoken BNC2014; Love et al. 2017), an 11.5-million-word corpus of everyday spoken British English conversation. The formulae examined are the conventionalised impoliteness formulae in British English presented in Culpeper (2010, 2011), a set that includes insults, threats, dismissals and more. The very existence of such formulae is a matter of controversy, as the current orthodoxy is that (im)politeness is not inherent in linguistic forms (e.g., Van der Bom and Mills 2015). An implication of this, if it were true, is that a corpus-based approach to the exploration of impoliteness would fail, as search procedures, assuming that they are not reliant on the manual annotation of functional characteristics, interrogate linguistic form. Key research questions, therefore, are (a) whether those conventionalised impoliteness formulae exist (i.e., whether they are indeed conventionalised), and (b) whether they have the formal features they are purported to do. A corollary of our findings here will, of course, relate to the searchability of these formulae in a corpus. Note that our investigation of conventionalisation here focuses on British English, though we cite studies, including those in this very volume, of impoliteness formulae in a range of other languages (see Section 2), thereby suggesting the wider applicability of our work.

Scholars have, in fact, already attempted to explore (im)politeness via corpus methods. The edited volume, Ruhi and Aksan (2015), is the first book-length publication to be devoted to the exploration of both politeness and impoliteness in a range of corpora. However, most of its studies treat corpora as data repositories for subsequent qualitative analyses of impoliteness examples, rather than opportunities to develop and exploit sophisticated search techniques. Moreover, even when particular structures are searched for and studied, such work has largely focussed on single words, such as swear words or taboo words, as a starting point for retrieving these (im)politeness structures. Çelebi and Ruhi (2015), for example, discuss methods for investigating impoliteness in spoken corpora. The approach they develop has the merit of being comprehensive (it includes the more implicational aspects of impoliteness), but, to find impoliteness formulae, it does no more than search for single words like *shit*, and then engage in much manual screening of the results (for a more recent example, see also Jucker and Landert 2023).

Culpeper (2010, 2011) indicates in his notation that most of his impoliteness formulae are not simply linguistic forms of impoliteness but often comprise structural units beyond the single word, and briefly alludes to Pattern Grammar (Francis, Hunston, and Manning 1996, 1998) as a possible theoretical approach, but provides no detail on or evidence for this. For example, the first structure listed

and probably the most frequent of all of them is the *YOU+NP* structure (e.g., *you moron*), a structure recently examined in detail in Van Olmen, Andersson, and Culpeper (2023), who claim and demonstrate that it constitutes an impoliteness construction. Other formulae, however, such as *fuck/damn you*, seem much less productive, and are possibly restricted to a limited number of phrases or words. Additionally, Culpeper (2010, 2011) provides no frequency information. Frequency is a key driver of the conventionalisation of formulae (Terkourafi 2005a: 213); the more frequent a particular structure performs impoliteness, the more likely it is to become conventionalised impoliteness over time. The issue here is not simply how often the form of a particular impoliteness formula occurs but also how often each form is taken to be or experienced as impolite. A search string that retrieves instances of a form occurring in exclusively impoliteness contexts is likely to be highly conventionalised for impoliteness, whereas the opposite is also the case, as would be the various degrees in between. We will report the success of our search strings to reveal forms as partly a proxy for their degree of conventionalisation. A key issue in all this will be our ability to identify impoliteness contexts; search strings might reveal the regularity of forms and structures, but they do not reveal whether they co-occur with impoliteness contexts and effects. We need a way of identifying impoliteness contexts that is independent of the grammar and semantics that inform our searches (to do otherwise would result in circularity). Yet the concept of impoliteness, like its close relative politeness, is notoriously difficult to define and highly subjective. We will state our definition of impoliteness and the evidence we consider in establishing it. Moreover, we will submit each impoliteness context to the independent scrutiny of three coders and then assess their inter-rater reliability.

The following section elaborates on impoliteness, specifically, its definition and the existence or otherwise of impolite forms and structures. Section 3 turns to matters of data – specifically, corpora – and method. Regarding data, it airs the issue of the frequency of impoliteness formulae in relation to corpora, describes the corpus used in this study (the Spoken BNC2014), and raises the problem of ethics and the way that is likely to suppress impoliteness formulae frequencies. Regarding method, it reviews evidence for impoliteness, and takes a position on metalinguistic cases and banter. Section 4 examines our set of impoliteness formulae for formal features as part of a quest to devise corpus search strings, and reports the success or otherwise of those search strings. Section 5 focusses on the contexts of those retrieved formulae, examining whether they really do contain impoliteness, and also noting whether they involve metalinguistic cases or cases of banter.

2 Impoliteness

2.1 Impoliteness?

Impoliteness is notoriously difficult to define, having been defined and redefined multiple times over the decades. However, on a rather more positive note, there is reasonably substantial overlap amongst the definitions. For the purposes of this chapter, we follow the fairly broad definition given in Culpeper (2011: 23, our emphasis):

Impoliteness is a negative attitude towards specific behaviours occurring in specific contexts. It is sustained by expectations, desires and /or beliefs about social organisation, including, in particular, how one person's or a group's identities are mediated by others in interaction. *Situated behaviours are viewed negatively – considered “impolite” – when they conflict with how one expects them to be, how one wants them to be and/or how one thinks they ought to be.* Such behaviours always have or are presumed to have emotional consequences for at least one participant, that is, they cause or are presumed to cause offence.

2.2 Impoliteness in forms?

Many scholars (e.g., Fraser and Nolen 1981: 96; Locher and Watts 2008: 78; van der Bom and Mills 2015) argue that politeness, and by extension impoliteness, is not inherent in linguistic forms but is instead determined by a contextual judgement. This enduring position has become near enough the orthodoxy in politeness studies at the current time. However, there are reasons why an extreme contextual position on politeness is untenable. One important reason is that people are able to judge how polite or impolite a word or expression is out of context, and another is that the conditions for conventionalised (im)politeness are there. For example, Van Olmen, Andersson, and Culpeper (2023), as already noted, provide evidence for a highly regular association between the YOU+NP structure and impoliteness effects across three languages, English, Dutch and Polish. Yet another reason is that it is cognitively implausible that interactants could work out the (im)politeness of language or behaviours afresh on every single occasion. As Clark (1996) argues, we need conventions, not least linguistic conventions, to help interactants coordinate meanings. Having said all this, everybody would agree that even conventionalised politeness or impoliteness can be overruled by the context. If this were not so, one could not account, for example, for sarcasm, where the words of politeness, such as *thank you*, might be used yet the context would suggest the very opposite of thanks. In short, the idea of a simple dichotomy whereby (im)politeness is either inherent in linguistic forms or not is untenable.

Terkourafi (e.g., 2001, 2005a, 2005b) proposes an elegant account of politeness in which it may flow from conventionalised linguistic forms *or* from the context, and Culpeper (2011) broadly follows that account in his work on impoliteness. In Terkourafi's (e.g., 2001, 2005a, 2005b) frame-based account of politeness, expressions become conventionally associated with their stereotypical contexts. Thus, expressions like *thank you*, or indeed *fuck you*, are not actually independent of context; their typical contexts, including (im)politeness effects, are stored with the expressions in one's mind. Terkourafi elaborates that it is through that regularity of co-occurrence that we acquire "a knowledge of which expressions to use in which situations" (2002: 197), that is, "experientially acquired structures of anticipated 'default' behaviour" (2002: 197). She offers the following definition of conventionalisation:

a relationship holding between utterances and context, which is a correlate of the (statistical) frequency with which an expression is used in one's experience of a particular context. Conventionalisation is thus a matter of degree, and may well vary in different speakers, as well as for the same speaker over time. This does not preclude the possibility that a particular expression may be conventionalised in a particular context for virtually all speakers of a particular language, thereby appearing to be a convention of that language. (Terkourafi 2005a: 213; see also 2001: 130)

Such a definition usefully aligns with the typical frequency-based methods of corpus linguistics.

2.3 Impoliteness beyond the word?

The oldest line of linguistic research on impoliteness is philological in flavour, often concerns swearing or taboo words, and is solidly focussed on words and short expressions. The classic work is Montagu's *Anatomy of Swearing* (2001 [1967]). More recently, and building on this tradition, we have seen Hughes' extensive *An Encyclopaedia of Swearing* (2006). The advent of modern sociolinguistics saw studies broaden to include a more rigorous social perspective (e.g., McEnery 2006), and one that combines both social and cognitive issues (e.g., Jay 2000). Nevertheless, aside from the fact that there is more to being impolite than just swearing, the focus remains on words. The advent of corpus linguistics probably aided and abetted this tendency. For example, Jucker and Landert (2023: 129), a paper that aims to quantify impoliteness in a corpus, states "To compile a list of terms with impolite meanings is perhaps somewhat more difficult [than with polite meanings]. For practical reasons, we focus on swear words and ignore other possible expressions and constructions".

There is no denying that impoliteness can and often is expressed through single words, and, moreover, that those single words may be embedded in broader structures. Nevertheless, impoliteness is not expressed solely through single words. Consider, for example, that an insulting vocative might just consist of *moron*, but could equally be *you moron* or *you stupid moron*. The constructional status of YOU+NP is discussed in Van Olmen, Andersson, and Culpeper (2023: Section 2). Discussing the Polish example *Ty językoznawco* (*you linguist*), they observe that

a typically non-evaluative noun such as ‘linguist’, if not modified by an adjective like ‘stupid’, seems rather incompatible with YOU+ NP’s overall function, at first glance. At the same time, if the construction nonetheless featured such a noun, it would coerce an evaluative reading. (Van Olmen, Andersson, and Culpeper 2023: 27)

Moreover, they note scholars arguing that this evaluative reading is likely to be negative in English (Culpeper 2011: 135; Jain 2022: 371), German (Finkbeiner, Meibauer, and Wiese 2016: 4), and Scandinavian languages (Julien 2016: 91). Furthermore, Van Olmen, Andersson, and Culpeper (2023) themselves find further evidence for this negative bias. Not only do they find a highly significant correlation between the YOU+NP construction and impoliteness effects, but they note that

the construction tends to be interpreted as an insult (without any context too!) even if the NP is (i) evaluatively neutral (like English *you theoretician* [. . .]), (ii) contains pseudowords (such as Dutch *jij blug*; see also Jain, 2022: 389), or (iii) is not actually spelt out, as in the Polish example in (28) [*Ty. . .!* (*You. . .!*)]. Put differently, YOU+NP tends to coerce not just an interpretation as addressee evaluation but one as negative addressee evaluation in particular. (Van Olmen, Andersson, and Culpeper 2023: 37)

From a methodological point of view, note that examples like *you linguist*, *you theoretician*, or simply *you!* would not be retrieved by a search string comprised of simple taboo words.

Of course, the YOU+NP construction is far from the only impoliteness formula. This chapter examines the full list of conventionalised impoliteness formulae given in Culpeper (2010: 3242–3243), which was slightly revised in Culpeper (2011: 135–136). Candidates for this list, given below, were manually derived from an extensive collection of impoliteness data (e.g., army training discourse, car parking arguments, diaries narrating offensive interactions), and then checked for validity in the Oxford English Corpus (OEC), a corpus of 2 billion words. Square brackets indicate some of the “slots” that comprise formulae. Slashes indicate alternatives, whether alternative slots or alternative illustrative examples, drawn from the data collection.

Insults

1. Personalized negative vocatives
 - [you] [fucking/rotten/dirty/fat/little/etc.][moron/fuck/plonker/dickhead/berk/pig/shit/bastard/loser/liar/minx/brat/slut/squirt/sod/bugger etc.] [you]
2. Personalized negative assertions
 - [you] [are] [so/such a] [shit/stink/thick/stupid/bitchy/bitch/hypocrite/disappointment/gay/nuts/nuttier than a fruit cake/hopeless/pathetic/fussy/terrible/fat/ugly/etc.]
 - [you] [can't do] [anything right/basic arithmetic/etc.]
 - [you] [disgust me] / [make me] [sick/etc.]
3. Personalized negative references
 - [your] [stinking/little] [mouth/act/arse/body/corpse/hands/guts/trap/breath/etc.]
4. Personalized third-person negative references (in the hearing of the target)
 - [the] [daft] [bimbo]
 - [she] ['s] [nutzo]

Pointed criticisms/complaints

- [that/this/it] [is/was] [absolutely/extraordinarily/unspeakably/etc.][bad/rubbish/crap/horrible/terrible/etc.]

Challenging or unpalatable questions and/or presuppositions

- why do you make my life impossible?
- which lie are you telling me?
- what's gone wrong now?
- you want to argue with me or you want to go to jail?

Condescensions (see also the use of “little” in Personalized negative references)

- [that] ['s/ is being] [babyish/childish/etc.]

Message enforcers

- listen here (preface)
- you got [it/that]? (tag)
- do you understand [me]? (tag)

Dismissals

- [go] [away]
- [get] [lost/out]
- [fuck/piss/shove] [off]

Silencers

- [shut] [it] / [your] [stinking/fucking/etc.] [mouth/face/trap/etc.]
- shut [the fuck] up

Threats

- [I'll/I'm/we're] [gonna] [smash your face in/beat the shit out of you/box your ears/bust your fucking head off/straighten you out/etc.] [if you don't] [X]
- [X] [before] [I] [hit/strangle] [you]

Negative expressives (e.g. curses, ill-wishes)

- [go] [to hell/hang yourself/fuck yourself]
- [damn/fuck] [you]

Importantly, although derived from British English, these impoliteness formulae seem to be present in many and diverse languages and cultures, for example: German (Kleinke and Bös 2015), Chinese (Lai 2019), Arabic (Rabab'ah and Alali 2020), Swedish (Andersson 2022), Serbo-Croatian (Šarić 2019), the languages of Congo-Brazzaville (Tsoumou 2023) and Lithuanian (Ruzaitė 2023). Indeed, other chapters in this volume connect with the specific formulae listed above in their discussions of other languages, including negative expressives (specifically curses) in three Nakh-Daghestanian languages (Avar-Ando-Tsezic, Lak and Rutul Lezgic) and in Italian (see, respectively, Dobrushina and Paternoster, this volume), threats in German (see Finkbeiner, this volume), insults as personalized negative assertions in German (see Queisser and Pleyer, this volume) and insults as personalized negative vocatives in Polish (see Van Olmen and Andersson, this volume).

3 Impoliteness formulae: corpora and method

3.1 Corpora: The frequency of impoliteness formulae

Unlike politeness formulae, impoliteness formulae are generally rare. Leech (1983: 105) notes that “conflictive illocutions tend, thankfully, to be rather marginal to human linguistic behaviour in normal circumstances”. Intuitively, this makes sense: society would hardly function if every interaction suffered the negative emotional effects of prototypical impoliteness. Of course, this is not to deny that there are some specific discourses in which impoliteness is central, including army training discourse (e.g., Bousfield 2008), confrontational TV (Culpeper 2005), and some online communities, particularly those that are politically charged in some way (Teneketzi 2022; Tsoumou 2022). Possible evidence of the relative general infrequency of impoliteness formulae is given in Culpeper (2011: 130):

the icons of English politeness *please* and *thank you* occur so much more frequently than possible icons of impoliteness such as *cunt* and *motherfucker* (the two British English lexical items considered most offensive in the year 2000, according to Millwood-Hargrave 2000). In the two-billion word Oxford English Corpus the frequencies are: *please* (14,627), *thank you* (5,533), *cunt* (157) and *motherfucker* (88).

One might argue that the infrequency of *cunt* and *motherfucker* is simply an artefact of what was collected to create the Oxford English Corpus (OEC). Noam Chomsky (Chomsky 1962: 159, a conference paper delivered in 1958) expressed doubts about the ability of a corpus to represent naturally-occurring language, and specifically impolite language:

Any natural corpus will be skewed. Some sentences won't occur because they are obvious, others because they are false, still are those because they are impolite. The corpus, if natural, will be so wildly skewed that the description would be no more than a mere list.

Chomsky eschewed naturally-occurring language data altogether and opted for constructed data, which, of course, has its own well-known validity issues. However, importantly, Chomsky was writing decades ago when corpora hardly existed, and, if they did, were likely to be very small and biased towards high-brow scholarly or literary works – not the obvious repositories of impolite language. The OEC is very large (2 billion words), and contains a wide variety of texts, including relatively informal, unregulated and unedited texts.

The corpus we examine in this paper is the Spoken BNC2014 (Love et al. 2017). Containing approximately 11.5 million words, it is undeniably relatively small when considered against many other corpora. However, as a collection of spoken British English conversation it is one of the biggest, if not *the* biggest, and that is thus one reason why we opted to use it. Conversations within that corpus are constituted by everyday spontaneous interactions across a reasonably varied number of situations. Such interactions should promise more instances of impoliteness than the relatively formal and monitored language that is typical of genres (e.g., newspapers, academic writing) that are often present in written corpora. Of course, we are not saying that writing cannot be informal and unmonitored. Indeed, as noted above, some online written genres seem to attract impoliteness, and this is partly why the relatively recent super-sized corpora, such as the TenTen series on Sketch Engine, which are scraped from the internet, are a reasonable prospect for some impoliteness research. One further reason we use the Spoken BNC2014 is that it can be accessed using CQPweb (Hardie 2012), a browser-based search program that allows corpus search queries of considerable complexity. This ability is crucial to our work. Whilst not relevant to this chapter, our choice is also motivated by the fact that this corpus contains a large amount of metadata for a range of social cat-

egories, such as gender, age, socio-economic group and region, thereby allowing us to study social variation in future research.

One point of note here is that corpora constructed in recent times are generally mindful of ethical issues. When the BNC2014 was constructed, the participants in the spoken data were asked for their permission before they were recorded. Given that impolite language is proscribed and seen as bad and debased, one might suppose that participants would avoid using it knowing they are being recorded. However, this potential problem is mitigated to an extent by banter and metalinguistic impoliteness, as will be discussed in the next section.

3.2 Method: Cases of genuine impoliteness

The corpus approach to politeness or impoliteness can never rely solely on the retrieval of linguistic forms. As already mentioned, the form-impoliteness experience pairing is variable. Our solution, as adopted for politeness in Culpeper and Gillings (2018) and van Dorst, Gillings, and Culpeper (2024), is to both retrieve linguistic forms and check that those forms are genuinely doing impoliteness by manually scrutinising the context. With respect to *politeness*, Terkourafi argues that “[i]t is the regular co-occurrence of particular types of context and particular linguistic expressions as the *unchallenged* realisations of particular acts that create the perception of politeness” (2005b: 248, our emphasis; see also 2005a: 213). For impoliteness, then, it is often the *challenged* realisations of acts that creates such a perception. Example (1) provides an illustration.

- (1) S0690: >>I'll have another one if you're making that
 S0687: **get lost**
 S0690: no
 S0687: go and make your own
 S0688: **you are a cheeky beggar**
 S0690: I'm a cheeky beggar?

BNC2014, S4HW

The two impoliteness formulae are in bold: *get lost* is a dismissal and *you are a cheeky beggar* is a variant of personalized negative assertions (Culpeper 2010: 3242). Both are immediately challenged by the next speaker.

A following challenge is not the only source of evidence. Culpeper (2011: 11) mentions retrospective comments as a source of evidence. The broader category here is metalinguistic comments. Occasionally, impoliteness formulae are labelled

negatively as “rude”, “inappropriate”, “annoying”, and so forth. Additionally, in our data, quite frequently impolite conversations, very often focussing on impoliteness formulae, are re-presented or presented (some scenarios are hypothetical) to others *because they are considered impolite*. In Example (2), the speaker describes the discourse of engineering companies:

- (2) what I didn't like about it was it was so male dominated it was so much male talk it was so much like I'm an alpha male you're not you're be you're a whatever so you're gonna have to follow me because basically I've got the power **shut up fuck off** and you'll do what I say yeah? and that male dominated world I'm sorry if you don't fit into it it's shit it's crap

BNC2014, S4ZT

The two impoliteness formulae are in bold: *shut up* is a silencer and *fuck off* is a dismissal. Such impoliteness formulae, mentions rather than uses, are important in shaping impoliteness formulae. Terkourafi's (e.g. 2001, 2005b) discussions of politeness are geared towards the use of particular expressions in specific contexts. Culpeper (2011: 131) argues that such occasions involving direct experience are fine for politeness, but that impoliteness is rather different:

Impoliteness [. . .] casts a much larger shadow than its frequency of usage would suggest. Behaviours and expressions considered impolite are more noticed and discussed than politeness (cf. Watts, 2003: 5). Impoliteness formulae are far from marginal in terms of their psychological salience, because their very abnormality (relative to their general frequency of use) attracts attention – they are foregrounded against the generally expected state for conversation, namely, politeness (Fraser 1990: 233). Not surprisingly, then, they are commented on and debated in all types of media, in official documents and in everyday chat, and so on. However, psychological salience is only part of what is going on here. Metadiscourse plays a role in the group dynamic that gives rise to a behaviour being evaluated as impolite in the first place.

In this paper, we will include and track such mentioned or metalinguistic instances of impoliteness formulae. Regarding the general paucity of impoliteness in corpora discussed earlier, one might suppose that participants are likely to feel less constrained in reporting the impoliteness of others compared with themselves.

The best known and most frequent cases of non-genuine or mock impoliteness fall under the heading of banter. Any quantitative study of impoliteness will need to take a position on banter. Broadly speaking, we follow the description of banter captured by Leech's (1983: 144) Banter Principle:

In order to show solidarity with *h*, say something which is (i) obviously untrue, and (ii) obviously impolite to *h*. [. . .] [This will give rise to an interpretation such that] what *s* says is impolite to *h* and is clearly untrue. Therefore what *s* really means is polite to *h* and true.

Note that sincerity is key here; or in Gricean (1975) terms, flouting the maxim of quality. Vergis (2017) provides empirical support for the role of sincerity. His results reveal that

breaching the Maxim of Quality in combination with obviously impolite remarks produces inferences in the predicted direction affecting all critical variables: When speakers were inferred to be less sincere, their ‘impolite’ remarks were perceived to be teasing and friendly, implicating a compliment. (Vergis 2017: 46)

The obviousness of the flout is often achieved by a contrast with the social context, notably, saying something impolite (e.g., *you bastard*) to somebody with whom you are close. Vergis (2017: 47; see also references therein) also found that “harsh” remarks in contexts involving close relations or solidarity “did not yield hostile or mean interpretations. In fact, the overall results of correlations showed a strong positive association among sincerity, teasing and friendliness”. Of course, not all banter involves neatly analysable mock insults like the one a few lines above. Exploiting sincerity, being “non-serious” can be done in the pursuit of humour or jocularity, and not to be “polite” or implicate “a compliment” (see Haugh and Bousfield 2012). It is also possible for participants to take teases as offensive, even in contexts where they are relatively frequent (Sinkeviciute 2014). One clue to banter is that its linguistic material can become conventionally associated with it. Vergis and Terkourafi (2015) show that the Greek collocation *re malaka* used in the context of young male Greeks in a close relationship is a case where a particular term of abuse has become conventionalised to show solidarity with the addressee rather than to insult. In our study, we take banter to be any case of mock impoliteness, that is insincere impoliteness, typically evidenced by the contrast with the discourse of close friendly relations of the participants, the playfulness or light-heartedness of the impoliteness (e.g., *you are a cheeky beggar*, as used in Example (1)) or the surrounding co-text (e.g., reciprocal banter). Importantly, for our purposes, banter still involves impoliteness at some level. Following Leech (1983: 144) above, a prerequisite for banter is that one is *saying* something that is obviously impolite, and an easy way of being obviously impolite is to produce a conventionalised impoliteness formula. We will include and track instances of impoliteness formulae performing banter.

4 Exploring the formal features of impoliteness through corpus search strings

4.1 Method

Our aim was to devise a search string for each of Culpeper's (2010) conventionalised impoliteness formulae listed in Section 2.3, and simultaneously identify at least some of their formal features. Given that the Spoken BNC2014 can be accessed within CQPweb, we had all of the affordances of that tool available to us, including the integration of two sophisticated tagging systems. In CQPweb, the Spoken BNC2014 has been tagged for part-of-speech according to the CLAWS6 tagset¹ and for semantic categories according to the USAS semantic tagger². What this means is that each word in the corpus has been assigned both a part-of-speech tag and a semantic tag. The word *dog*, for example, is tagged *_NN1*, a singular common noun, and is also tagged *L2: Living creatures generally*. Using CQP-syntax, the user can then combine a search for, say, a regular word form, with a search for a specific POS tag, a specific semantic tag, and also restrict it to appearing in a certain position within the utterance. As we will show, this is a powerful way of capturing the formal features of impoliteness formulae, if those formal features are present.

We examined each conventionalised impoliteness formula in turn, and, mindful of specific part-of-speech and semantic tags, broke it down into its specific formal components. We considered which elements in each structure were compulsory (i.e., had to be present in order for conventionalised impoliteness to be represented) and which were optional. This was an iterative process whereby we refined the search string over several rounds, testing searches within CQPweb and adjusting the search parameters to find the maximum number of impoliteness occurrences. The resultant strings are therefore a combination of the proposed impoliteness formulae, as per Culpeper (2010), and practical searches to improve both recall and precision within our particular corpus. Consider personalised negative vocatives. The structure listed in Culpeper (2010: 3142) and then our search syntax is as follows:

[you][fucking/rotten/dirty/fat/little/etc.][moron/fuck/plonker/dickhead/berk/pig/shit/bastard/loser/liar/minx/brat/slut/squirt/sod/bugger/etc.][you]

1 <https://ucrel.lancs.ac.uk/claws6tags.html>.

2 https://ucrel.lancs.ac.uk/usas/usas_guide.pdf.

```
("you" [pos="JJ.*" & fullsemtag="\|. *d\|"] [pos="N.*"] ("you")* </u>) | (<u>"you" [pos="JJ.*" & fullsemtag="\|. *d\|"] [pos="N.*"] ("you")* | ("you" [pos="JJ.*"]* [pos="N.*" & fullsemtag="\|. *d\|"] ("you")* </u>) | (<u>"you" [pos="JJ.*"]* [pos="N.*" & fullsemtag="\|. *d\|"] ("you")*)
```

The first element in the structure, the second person pronoun “you”, can simply be rendered in our search string as “you”. The next element (*fucking, rotten, dirty*, etc.) consists of an adjective, followed by a noun (*moron, fuck, plonker*, etc.). However, to maximize precision, we needed to be more precise than just including all adjectives and all nouns (*you darling*, for example, needs to be excluded). We wanted either the adjective or the noun to be negative in some way. Thus, we devised a search string which returned only those words that had been tagged by the POS tagger as an adjective and which were also tagged by the semantic tagger as negative (*[pos="JJ.*" & fullsemtag="\|. *d\|"]*), and were then followed by a noun (*[pos="N.*"]*).³ Or, words that had been tagged by the POS tagger as an adjective (as an optional element) which were followed by words tagged as a noun and by the semantic tagger as negative (*[pos="N.*" & fullsemtag="\|. *d\|"]*). At the end of the string, we needed to include “you” again as an optional element (*("you")**). Finally, we added an additional element to restrict the search to structures appearing only at the end of the utterance (*</u>*). We repeated these search strings two more times, separated by a pipe (“|” meaning “or” in CQP-syntax) to restrict the search to structures at the beginning of the utterance. The point of restricting this formula to the beginnings and ends of utterances helped exclude cases where the words happen to be embedded within another larger structure or were split across two or more structures. For example, *you sorry owner* follows the YOU+NP structure, but is in fact split across three utterances and three speakers: “S0520: bless you S0519: sorry S0521: owner will object” (BNC2014, S4CU). We repeated this process for each of the conventionalised impoliteness formulae.

4.2 Results

The corpus queries and the related formal features of all the target conventionalised impoliteness formulae are listed in three tables below. The first two tables, Tables 1 and 2, contain descriptions of search strings and their formal features that proved inadequate for the retrieval of impoliteness formulae. There were two

³ Those familiar with the USAS semantic tagger may be wondering why we did not simply restrict the search to *S1.2.4- Impolite*. This tag is only assigned to words *about* impoliteness, rather than impolite words themselves. We thus opted to widen the search to *all* negative semantic tags (i.e., those listed on this page with a minus sign: https://ucrel.lancs.ac.uk/usas/semtags_subcategories.txt).

main reasons for this inadequacy: (1) precision, that is, the search string captured a bundle of formal features that were not conventionalised (enough) for impoliteness and retrieved a high proportion of non-impolite cases; or (2) recall, that is, the search string worked well in capturing the relevant bundle of formal features, but it hardly appeared in impolite contexts in this specific corpus. The challenge with precision was relevant for personalised negative references, personalised third-person negative references, pointed criticisms/complaints, and some types of silencers. These impoliteness formulae and their corpus queries can be found in Table 1. For these formulae, the search strings and their formal features were too generic and included very few impoliteness contexts. The high frequencies for some structures in this table also suggest that they were not conventionalised enough for impoliteness contexts.

Table 1: Culpeper (2010)'s impoliteness formulae, and their associated CQP syntax strings and formal elements: excluded due to lack of precision.

Culpeper (2010)		CQP syntax	Formal elements	Freq.
Personalised negative references	[your] [stinking/little] [mouth/act/arse/body/corpse/hands/guts/trap/breath/etc.]	("your" [pos="JJ.*" & fullsemtag="\ . *d\ "] [pos="N.*"]) ("your" [pos="JJ.*"]* [pos="N.*" & fullsemtag="\ . *d\ "])	(your + negative adjective + noun) OR (your + optional adjective + negative noun)	1,345
Personalised third-person negative references (in the hearing of the target) ⁴	[the] [daft] [bimbo] [she] ['s] [nutzo]	"the" [pos="JJ.*" & fullsemtag="\ . *d\ "] [pos="N.*" & fullsemtag="\ . *d\ "] [pos="PPHS1*"] [pos="VB.*"] [pos="JJ.*" & fullsemtag="\ . *d\ "]	the + negative adjective + negative noun third person pronoun + verb be + negative adjective	221 1,332
Pointed criticisms/complaints	[that/this/it] [is/was] [absolutely/extraordinarily/unspeakably/etc.] [bad/rubbish/crap/horrible/terrible/etc.]	[pos="(DD1 DD2 PPH1)"] [pos="VB.*"] [pos="R.*"]* [pos="JJ.*" & fullsemtag="\ . *d\ "]	that/this/these/those/it + verb be + optional adverb + negative adjective	19,538

⁴ Not only did we find that this was not conventionalised enough, we also have no way of knowing whether it is in the hearing of the target.

Table 1 (continued)

Culpeper (2010)		CQP syntax	Formal elements	Freq.
Challenging or unpalatable questions and/or presuppositions	why do you make my life impossible?	<u> [pos="RRQ.*"] [pos=".*"]* "\?" </u>	start of utterance + WH-question word + any word, as often as needed + question mark + end of utterance	7,697
	which lie are you telling me?			
	what's gone wrong now?	<u> [pos="PPY"] [pos=".*"]* "\?" </u>	start of utterance + you + any word, as often as needed + question mark + end of utterance	5,055
	you want to argue with me or you want to go to jail?			
Silencers	[shut] [it] / [your]	"shut" ([pos="DD1"] ("your" [pos="JJ.*"]* [pos="N.*"]))	shut + (it) OR (your + optional adjective + noun)	37
	[stinking/fucking/etc.]			
	[mouth/face/trap/etc.]			

The challenge with recall was relevant for condescensions, message enforcers, and some types of personalised negative assertions and negative expressives. These impoliteness formulae and their corpus queries can be found in Table 2. Here we found that the search strings we developed did precisely what we wanted them to do, yet only very few concordance lines were returned to us. The low frequencies for each search suggest to us that the formulae could be reasonably conventionalised for impoliteness, but the lack of impolite contexts in the data meant that nothing could be verified. Strictly speaking, the issue is not bad recall, but insufficient examples to be recalled in this corpus.

Table 2: Culpeper (2010)'s impoliteness formulae, and their associated CQP syntax strings and formal elements: excluded due to lack of recall.

Culpeper (2010)		CQP syntax	Formal elements	Freq.
Personalised negative assertions	[you] [can't do]	"you" [pos="VM.*"] [pos="XX"] [pos="VD.*"] [pos="PN1"] [pos!="I.*"]	you + modal auxiliary verb + negation + verb do + indefinite pronoun singular + NOT followed by a preposition	30
	[anything right/basic arithmetic/etc.]			

Table 2 (continued)

Culpeper (2010)		CQP syntax	Formal elements	Freq.
Condescensions (see also the use of 'little' in Insults)	[that] ['s/is being] [babyish/childish/ etc.]	"that" [pos="VB.*"] [pos="VBG"] [pos="JJ.*"]	that + verb be + being + adjective	11
Message enforcers	listen here (preface)	<u>"listen" "here"	start of utterance + listen + here	1
	you got [it/that]? (tag)	"you" "got" [pos="(DD1 PPH1)"] "\?"	you + got + this/ that/it + question mark	33
	do you understand [me]? (tag)	"do" "you" "understand" "me"* "\?"	do + you + understand + me + question mark	7
Threats	[I'll/I'm/we're] [gonna] [smash your face in/beat the shit out of you/box your ears/bust your fucking head off/ straighten you out/ etc.] [if you don't] [X]	[pos="(PPIS1 PPIS2)"] ((([pos="(VBR VBM)"] [pos="VVGK"] [pos="TO"]]) "will") [pos="VVI" & fullsemtag="\ . *d\ "] "you"	I/we + (am/are + going + to) OR (will) + negative verb + you	36
	[you'd better be ready Friday the 20th to meet with me/do it] [or] [else] [I'll] [X]	((("you" [pos="VHD"])* "better" [pos="V.*"]) ("or" "else" [pos="PPIS1"])))	(you + optional had + better + verb) OR (or + else + I)	94
	[X] [before I] [hit you/ strangle you]	"before" [pos="(PPIS1 PPIS2)"] [pos="V.*" & fullsemtag="\ . *d\ "] "you"	before + I/we + negative verb + you	3
Negative expressives (e.g., curses, ill-wishes)	[go] [to hell/hang yourself/fuck yourself]	"go" (([pos="II"] "hell") ([pos="V.*"] [pos="PPX.*"])))	go + (to + hell) OR (verb + reflexive pronoun)	8

Table 3 details those impoliteness structures whose formal features we were able to adequately convert into corpus syntax. "Adequate", here, means that more than 30% of the results on the first page of concordance lines (i.e., 15 out of 50) were pointing towards impoliteness usage at first glance (we scrutinised impoliteness contexts further in the second phase of our study; see Section 5). Ultimately, we were left with 5 impoliteness formulae types that, including their variants, could be searched for via 8 corpus search strings. These were: personalised negative voc-

atives, personalised negative assertions, dismissals, silencers, and negative expressions. In total, the 8 impoliteness formulae that we include amount to 1,502 raw occurrences in the Spoken BNC2014.

Table 3: Culpeper (2010)’s impoliteness formulae, and their associated CQP syntax strings and formal elements: inclusion list.

Culpeper (2010)		CQP syntax	Formal elements	Freq.
Personalised negative vocatives	[you] [fucking/rotten/dirty/fat/little/etc.][moron/fuck/plonker/dickhead/berk/pig/shit/bastard/loser/liar/minx/brat/slut/squirt/sod/bugger/etc.] [you]	("you" [pos="JJ.*" & fullsemtag="\ . *d\ "] [pos="N.*"] ("you")* </u>) (<u>"you" [pos="JJ.*" & fullsemtag="\ . *d\ "] [pos="N.*"] ("you")* ("you" [pos="JJ.*" & fullsemtag="\ . *d\ "] ("you")* </u>) (<u>"you" [pos="JJ.*" & fullsemtag="\ . *d\ "] [pos="N.*" & fullsemtag="\ . *d\ "] ("you")*)	(you + negative adjective + noun + optional you + end of utterance) OR (start of utterance + you + negative adjective + noun + optional you) OR (you + optional adjective + negative noun + optional you + end of utterance) OR (start of utterance + you + optional adjective + negative noun + optional you)	97
Personalised negative assertions	[you] [are] [so/such a] [shit/stink/thick/stupid/bitchy/bitch/hypocrite/disappointment/gay/nuts/nuttier than a fruit cake/hopeless/pathetic/fussy/terrible/fat/ugly/etc.]	("you" [pos="VB.*"] "so" [pos="JJ.*" & fullsemtag="\ . *d\ "] ("you" [pos="VB.*"] [pos="DA.*"]* [pos="A.*"] [pos="JJ.*"]* [pos="N.*" & fullsemtag="\ . *d\ "] ("you" [pos="VB.*"] [pos="DA.*"]* [pos="A.*"] [pos="JJ.*" & fullsemtag="\ . *d\ "] [pos="N.*"])	(you + verb be + so + negative adjective) OR (you + verb be + optional such + determiner + optional adjective + negative noun) OR (you + verb be + optional such + determiner + negative adjective + noun)	262
	[you] [disgust me] / [make me] [sick/etc.]	("you" "disgust" [pos="P.*"] ("you" "make" [pos="P.*"] [pos="JJ.*" & fullsemtag="\ . *d\ "]	(you + disgust + pronoun) OR (you + make + pronoun + negative adjective)	4

Table 3 (continued)

Culpeper (2010)		CQP syntax	Formal elements	Freq.
Dismissals	[go] [away]	("go" "away" </u>) (<u> "go" "away")	(go + away + end of utterance) OR (start of utterance + go + away)	104
	[fuck/piss/shove] [off]	("fuck" "piss" "shove") "off"	fuck/piss/shove + off	314
	[get] [lost/out]	("get" "lost out" </u>) (<u> "get" "lost out")	(get + lost/out + end of utterance) OR (start of utterance + get + lost/out)	136
Silencers	shut [the fuck] up	"shut" [pos="A.*"]* [pos="N.*"]* "up"	shut + optional determiner + optional noun + up	482
Negative expressives	[damn/fuck] you	("damn" "fuck") "you"	damn/fuck + you	105

5 Exploring the impoliteness contexts of impoliteness formulae

5.1 Method

Whilst we expected many instances of retrieved impoliteness formulae to be performing impoliteness (especially due to the 30% threshold noted above), we could not assume that would be the case for all instances. Following the procedure outlined in Culpeper and Gillings (2018) and van Dorst, Gillings, and Culpeper (2024), we thus went through each concordance line in turn and made a judgement about whether each instance of the impoliteness formulae in the corpus was actually impolite or not. However, because such judgements are subjective, we used inter-rater reliability procedures to enable more robust judgements: three independent judgements were made and then compared for similarity.

The inter-rater reliability judgements proceeded in three stages, in tune with best practice (see Mackey and Gass 2021: Chapter 4). First, we (i.e., the three authors of this paper) had a joint training session, where we took 10 instances of each included conventionalised impoliteness formulae and had a discussion around what we considered to be impolite usage, grounding our discussion in the impoliteness definition given in Section 2.1 and the types of evidence of impoliteness given in Section 3.2. As noted in Section 3.2, we made the decision to include

banter (noting that for banter to work effectively, speakers must collectively agree that the construction used is impolite to at least some degree), and to also include metalinguistic usage (including reported speech, reported thought, and hypothetical examples). This meant, for example, that if speakers were reconstructing an earlier conversation where someone was rude to them, it was included as impolite; yet if someone was simply having a discussion about what a particular impolite phrase means, then it was excluded.

Second, after the initial joint training session, we then proceeded to independently code a sample set of concordance lines, that is, 15% of the 1,502 total lines. Since there were only 4 instances of personalised negative assertions (and these were already coded during the joint training session), the concordance lines were divided by the remaining 7 search strings, resulting in 33 concordances of each impoliteness formula (due to rounding). This resulted in 231 concordances being screened in total.

After independently screening the 231 lines, we met again, but all reported that we had difficulty in interpreting some instances. The origin of such difficulty was the formatting of our concordances. We had exported the concordance lines from CQPweb as an Excel spreadsheet, and when exporting them, lines lost their speaker ID tags, and consequently information on where one person's utterance ends and where the next person starts. This information is vital in much work using concordance analysis (Gillings and Mautner 2024; Baker, 2018), but perhaps even more so in corpus pragmatic work such as this where it is necessary to know who says what to whom, and how other interactants respond. Rather than discuss our independently-screened lines at this point, we opted to begin a third step and recode our training data for a second time, but this time within CQPweb. Performing the concordance analysis within CQPweb meant that we not only now knew who said what to whom, but we could also access the expanded concordance line to gain further co-text, including the turn-taking structure, to help aid our interpretation. Given that we were reading the wider contexts (wider than a concordance line) of each of the 231 instances, this represented a tough but manageable task.

As part of this third step, we also developed a secondary coding scheme in order to make sure that we were not simply identifying genuinely impolite cases but also tracking the occurrence of banter and metalinguistic usages. In addition, we tracked a particular usage of impoliteness formulae when they perform incredulity. Incredulity markers, as we termed them, are rather more than simple exclamations: they display incredulity at what the speaker has just said, implying that it could not possibly be true, by using an impoliteness formula, specifically a silencer or dismissal, that literally tells the target in an attitudinally marked way to stop what they are doing or come away from their expressed

opinion (compare with the grammaticalized expression *come off it*). Example (3) is an illustration:

- (3) S0331: they have erm penises that are shaped like a like a well apparently
 S0328 [??]: yeah?
 S0331: yeah
 S0330 [??]: I don't understand like what
 S0326: but but they have they have erm
 S0326: er and like the and like river dolphins in the Amazon you get you get like
 river dolphins in like
 S0331: >>river dolphins they sound weird
 S0326: put them in a brothel like dead
 S0330 [??]: what? **shut up**
 S0326: and dudes go there and have sex with a dolphin

BNC2014, S3M9

5.2 Results

After all three raters had again independently re-coded the 231 concordance lines according to whether they were impolite or not, we compared responses and finalised our judgements. A Fleiss' kappa inter-rater reliability test (Fleiss 1971) was conducted to assess how similar those judgements were. Fleiss' kappa provides a value between -1 and +1, where a positive number indicates agreement beyond chance. Overall, the inter-rater reliability (IRR) kappa was 0.792 across the 231 concordances, suggesting substantial agreement (Hartling et al. 2012; McHugh 2012)⁵. What these judgements revealed was a fairly high impoliteness rate, that is, a high proportion of searched-for impoliteness formulae occurring in impolite contexts – more precisely, 63.2% of all concordances screened. There was, however, some variation across the formulae. In Table 4, we break down the impoliteness rate and the inter-rater reliability agreement score for each of the 8 conventionalised impoliteness formulae (this includes variants within the 5 types) listed in Table 3. The results given in these Tables include cases of banter, for reasons discussed in Section 3.2. Later in this section, we conducted further analyses for cases of banter and meta-linguistic use.

⁵ Interpretation of Fleiss' kappa values: ≤ 0 as indicating poor agreement, 0.01–0.20 as slight agreement, 0.21–0.40 as fair, 0.41–0.60 as moderate agreement, 0.61–0.80 as substantial, and 0.81–1.00 as almost perfect agreement (Hartling et al. 2012)

Table 4: Screened frequencies and impoliteness rates for 8 conventionalised impoliteness expressions.

Impoliteness formula (Culpeper, 2010)		Screened freq.	Average impoliteness rate (%)	IRR (Fleiss' kappa)
Personalised negative vocatives	[you] [fucking/rotten/dirty/fat/little/etc.] [moron/fuck/plonker/dickhead/berk/pig/shit/bastard/loser/liar/minx/brat/slut/squirt/sod/bugger/etc.] [you]	21	63.6	0.915
Personalised negative assertions	[you] [are] [so/such a] [shit/stink/thick/stupid/bitchy/bitch/hypocrite/disappointment/gay/nuts/nuttier than a fruit cake/hopeless/pathetic/fussy/terrible/fat/ugly/etc.] [you] [disgust me] / [make me] [sick/etc.] ⁶	13 3	39.4 75	0.667 1 ⁷
Dismissals	[go] [away]	24	72.7	0.844
	[fuck/piss/shove] [off]	29	87.9	0.734
	[get] [lost/out]	14	42.4	0.698
Silencers	shut [the fuck] up	30	90.9	0.182
Negative expressives	[damn/fuck] you	15	45.5	0.877

As can be seen from the individual impoliteness rates in Table 4, 5 of the 8 formulae and their variants have an impoliteness rate that is well above 50%. But clearly there is variation: some of the formulae appear to have a stronger association with impoliteness than others. In particular, the dismissal *[fuck/piss/shove][off]* and the silencer *shut [the fuck] up* score most highly, with 87.9% and 90.9% respectively. In contrast, another dismissal, *[get][lost/out]*, is far less conventionalised for impoliteness at 42.4%. This suggests that we cannot generalise from one variant of an impoliteness formula across all the variants.

One might have expected that formulae with more overtly taboo words in them would be associated more strongly with impoliteness contexts. After all, Culpeper (2011: 139) states:

⁶ Note that the numbers for this formula are taken from the training data, as only 4 concordances existed in the first place.

⁷ As this formula was only screened as part of the training data, this was done collectively and agreed upon collectively. This formula is not included in the overall inter-rater reliability test.

Using various means – the addition of modifiers, taboo words, particular prosodies, non-verbal features, and so on – to exacerbate the offensiveness of an impoliteness formula is not simply an optional extra that calibrates where exactly on a scale of impoliteness the item falls. A formula [. . .] might not be considered impolite at all, were it not for the fact that it is intensified.

This does not appear to be borne out by our results. For example, negative expressives, [*damn/fuck*] *you*, occur 105 times, and all but eight involved the word *fuck*. Out of the screened sample, all but one involved the word *fuck*. Yet the impoliteness rate of negative expressives is only 45.5. However, this, we think, has more to do with the contemporary weakening of *fuck* (see Love and Stenstrom 2023); it is not the taboo word it once was.

The two categories of insults listed, personalised negative vocatives and personalised negative assertions (specifically the first variant), are clearly more productive, more abstract and more dependent on grammar than those impoliteness formulae comprised of a limited number of specific lexical items. Their impoliteness rates are 63.6% and 39.4% respectively. 39.4% is the lowest score of any impoliteness formula, and 63.6% is lower than all but two of the remaining formulae. This finding is probably a consequence of the fact that the slots within the structures that comprise these formulae may be filled by a range of different words with a range of different impoliteness values. Intuitively, for example, saying *you are fussy* seems very different from saying *you are ugly*. This broad range of values probably has had a dilution effect on the overall impoliteness rates.

Regarding inter-rater reliability, overall, the kappa scores given for each impoliteness formula are similar, with all but one having substantial or higher agreement. The impoliteness formula scoring significantly lower was the silencer *shut [the fuck] up*, with a Fleiss' kappa of 0.182. However, this does not match its high impoliteness rate of 90.9%. It is exactly due to this extremely high impoliteness rate that the Fleiss' kappa did not perform as expected, as there was not enough variation in the ratings (i.e., not enough concordances were assessed as “not impolite”) for the statistic to measure chance agreement effectively. Most notably, the silencer was considered to be impolite in 100% of cases by one of the raters and only in 72.7% of cases by another (thus, a 27.3% difference in impoliteness rate between raters). Note that none of the examples of *shut [the fuck] up* included the infix *the fuck*; generally, this feature is quite rare. This uncertainty regarding *shut up* probably reflects the fact that it is rather weak as an impoliteness formula, not clearly securing impoliteness perlocutionary effects. More generally, such differences, despite multiple training sessions and the use of conventionalised impoliteness formulae, remind us that impoliteness remains to an extent complex and personal.

The secondary coding of banter and metalinguistic usage shows generally lower inter-rater reliability kappa scores of 0.407 for banter and 0.531 for incre-

dulity markers, along with 0.756 for metalinguistic use. However, those numbers are somewhat misleading. Whereas the impoliteness coding was a binary choice between “impolite” and “not impolite”, the banter, metalinguistic and incredulity codes were optional. Thus, although 231 instances of impoliteness formulae were scrutinised, the values given above do not represent agreement across 231 codes. Table 5 provides a more complete picture.

Table 5: Inter-rater reliability and distributions of banter, metalinguistic and incredulity codes amongst the 231 instances of impoliteness formulae.

	IRR (Fleiss' kappa)	No. of cases with at least 1 rater code	No. of cases with 3 rater agreement codes
Banter	0.407	38	6
Metalinguistic	0.756	63	35
Incredulity marker	0.531	13	3

Regarding banter, 38 cases were identified by at least one rater as banter. We do not have comparative data by which we can judge whether that proportion of the 231 cases is large or small, but certainly it is not overwhelmingly dominant. Only 6 times did all three raters identify an instance as banter (thus, full agreement was only reached in 16% of cases). This is not entirely surprising. Identifying the key feature (see Section 3.2) of sincerity is not easy, all the less so when our data is lacking visual cues (e.g., gestures) and prosodic cues, as well as some contextual knowledge (e.g., the discourse habits of the people involved in the conversation). Given that our banter analysis is based on sample of 33 concordances for each formula variant, we cannot make claims that banter correlates with some formulae more than others with any confidence. There seems to be a preference for personalised negative vocatives, but that needs to be verified by future research.

Regarding metalinguistic cases, these were slightly more frequent than banter, but much more consistently identified across raters. Of the 231 impoliteness formulae instances, 63 were identified by at least one rater as metalinguistic. That 27% of impoliteness formulae instances are metalinguistic does seem to provide some support for Culpeper's (2011) contention that metalinguistic discussions are a key feature of impoliteness. However, we do not of course have comparative data, so we cannot say for sure that this is more than it would be for politeness formulae, though that seems unlikely. 35 cases were identified by all three raters as a metalinguistic usage (thus, full agreement was reached in 56% of cases). This relatively high rate is not surprising, because the metalinguistic nature of the example was often explicitly marked by, for example, the quotative *like*.

Finally, regarding incredulity, in 13 cases at least one rater identified an incredulity marker, and 3 times all three raters agreed on its occurrence (thus, in 23% of cases). Whilst the numbers are slight, proportionally, it was more reliably identified than banter.

6 Conclusion

The very existence of impoliteness formulae is challenged by the idea that impoliteness cannot be inherent in linguistic form. Our study shows not only that some impoliteness formulae can be successfully searched for in the corpus data on the basis of formal features, but also that it can be done for impoliteness formulae that are not restricted to single words or relatively fixed phrases. These impoliteness formulae were: personalised negative vocatives (e.g., *you moron*), some types of personalised negative assertions (e.g., *you are shit*), dismissals (e.g., *fuck off*), silencers (e.g., *shut [the fuck] up*) and some types of negative expressives (e.g., *fuck you*). Together they amounted to 1,502 raw occurrences in the Spoken BNC2014.

However, mere regularity of certain clusters of forms does not mean they are impoliteness formulae. What we needed to establish is that they regularly occur in impoliteness contexts, establishing the kind of regularity that results in anticipated, default behaviours and understandings and thus leads to conventionalisation (Terkourafi 2001). The second part of our study, therefore, set about quantifying the proportion of the instances of impoliteness formulae that appeared in impoliteness contexts. The identification of these contexts is difficult because of the subjectivity of impoliteness. We were clear about our notion of impoliteness and the features that might evidence it, but in addition we deployed inter-rater techniques so that we could formulate more robust judgements. Of the 8 variants of formulae types, 5 – personalised negative vocatives, one variant of personalised negative assertions, two variants of dismissals, and silencers – emerged with strong associations with impoliteness contexts. The dismissal [*fuck/piss/shove*][*off*] and the silencer *shut [the fuck] up* scored most highly, with 87.9% and 90.9% respectively. However, we noted that particular variants *within* an impoliteness formula type could vary considerably (another dismissal, [*get*] [*lost/out*] occurred in impoliteness contexts only 42.4% of the time). The presence or otherwise of taboo words such as *fuck* seems to make little difference to whether or not there is a correlation with impoliteness contexts. If this is the case, studies that rely solely on taboo words for impoliteness diagnostics would be problematic. Having said that, caution is needed here, as *fuck*, the most frequent taboo word in our data, has considerably weakened over time.

We noted that the more productive, more abstract formulae that are more dependent on grammar as opposed to specific words (e.g., personalised negative vocatives and personalised negative assertions (specifically the first variant)) seem to attract slightly lower impoliteness scores, possibly because they accept a wide range of words with a wide range of impoliteness values in their grammatical slots, with the result that impoliteness scores are averaged downwards. We also observed that inter-rater agreement tended to be weaker for impoliteness formulae that are less strongly associated with impoliteness contexts, an example being personalised negative assertions like *you are shit*, which has a low impoliteness rate of 39.4% and a low IRR of 0.667, presumably because, being more weakly associated with such contexts, they are less able to clearly secure impoliteness effects.

We argued that both banter and metalinguistic uses should be counted as cases of impoliteness, because impoliteness is involved at some level in each. Banter accounted for a small but significant proportion of the data, 38 of the 231 instances (16.4%). But banter is certainly not easy to identify, achieving only 0.407 IRR agreement amongst the three coders. If sincerity is a core feature, it is easy to imagine why banter is difficult to identify. Metalinguistic cases of impoliteness account for a larger proportion, 27%, suggesting that cases of impoliteness do have a salience that stimulates talk about them, talk that helps conventionalise them as impolite. Such cases were also much easier to identify, achieving 0.756 IRR agreement. In addition, we suggested the category of “incredulity marker”. However, although it could be reliably identified, numbers were relatively small, preventing further exploration.

This study set out to explore the set of impoliteness formula presented in Culpeper (2010). That fact is also a limitation, as it does not explore items beyond that set, even when within those items the fact that the set might be expanded is indicated by “etc.”. Nevertheless, with nine different impoliteness formulae types and their variants, the set is large enough and diverse enough for the research objectives of this study. A key part of our method was to construct corpus search strings for each impoliteness formula and variant, using their success as a proxy for whether the impoliteness formulae had sufficient formal features and what they were. This was also a limitation, because a corpus search does not necessarily capture every kind of formal feature, and grammatical and semantic tagging can contain errors. Having said that, we were able to develop powerful search queries, incorporating words, parts of speech, semantic fields, positioning in the structure and a degree of optionality. We used the Spoken component of the BNC2014 as our corpus. A limitation here is that this is a relatively small corpus, because it contains transcripts of naturally occurring conversation, and impoliteness formulae do not occur frequently, even in informal unmonitored interactions. Nevertheless, it generally sufficed for our relatively intensive study methods. The exceptions con-

cerned condescensions (e.g., *that's being babyish*), message enforcers (e.g., *listen here*), some types of personalised negative assertions (e.g., *you can't do anything right*), and some types of negative expressives (e.g., *go to hell*). With these impoliteness formulae, we were able to retrieve them, but too few appeared in the corpus for us to be able to study them further.

Overall, our findings must be taken as indicative rather than conclusive. Nevertheless, we hope that we have demonstrated clearly that impoliteness has a life in conventionalised formal structures (i.e., in grammar). We also hope that our work stimulates further research. On that point, note that all our CQPweb search strings, as displayed in our Tables, can be copied and pasted into the search box of CPQweb (or other programs using that search syntax) in order to replicate the searches we conducted and retrieve the associated examples from the BNC2014. Of course, they could also be used to interrogate other corpora of English, or as inspiration for constructing searches in corpora of other languages and thereby facilitating cross-linguistic comparisons.

References

- Andersson, Marta. 2022. "So many "virologists" in this thread!" Impoliteness in Facebook discussions of the management of the pandemic of Covid-19 in Sweden – the tension between conformity and distinction. *Pragmatics* 32(4). 489–517.
- Baker, Paul. 2018. Conclusion: Reflecting on reflective research. In Charlotte Taylor & Anna Marchi (eds.), *Corpus Approaches to Discourse: A Critical Review*, 281–292. London: Routledge.
- Bousfield, Derek. 2008. *Impoliteness in Interaction*. Amsterdam: John Benjamins.
- Çelebi, Hatice & Şükriye Ruhi. 2015. Identifying impoliteness in spoken corpora: A methodological perspective. In Şükriye Ruhi & Yeşim Aksan (eds.), *Exploring (im)politeness in specialized and general corpora: Converging methodologies and analytic procedures*, 216–255. Cambridge: Cambridge Scholars Publishing.
- Chomsky, Noam. 1962. A transformational approach to syntax. In Archibald A. Hill (ed.), *Proceedings of the Third Texas Conference on Problems of Linguistic Analysis in English on May 9–12, 1958*, 124–158. University of Texas.
- Clark, Herbert H. 1996. *Using Language*. Cambridge: Cambridge University Press.
- Culpeper, Jonathan. 2005. Impoliteness and entertainment in the television quiz show: the Weakest Link. *Journal of Politeness Research* 1(1). 35–72.
- Culpeper, Jonathan. 2010. Conventionalized impoliteness formulae. *Journal of Pragmatics* 42(12). 3232–3245.
- Culpeper, Jonathan. 2011. *Impoliteness: Using Language to Cause Offence*. Cambridge: Cambridge University Press.
- Culpeper, Jonathan & Mathew Gillings. 2018. Politeness variation in England: A North-South divide? In Vaclav Brezina, Robbie Love & Karin Aijmer (eds.), *Corpus Approaches to Contemporary British Speech: Sociolinguistic studies of the Spoken BNC2014*, 33–59. London: Routledge.

- Finkbeiner, Rita, Jörg Meibauer & Heike Wiese. 2016. What is pejoration, and how can it be expressed in language? In Rita Finkbeiner, Jörg Meibauer & Heike Wiese (eds.), *Pejoration*, 1–18. Amsterdam: John Benjamins.
- Fleiss, Joseph L. 1971. Measuring nominal scale agreement among many raters. *Psychological Bulletin* 76(5), 378–382.
- Francis, Gill, Susan Hunston & Elizabeth Manning. 1996. *Collins COBUILD Grammar Patterns 1: Verbs*. London: HarperCollins.
- Francis, Gill, Susan Hunston & Elizabeth Manning. 1998. *Collins COBUILD Grammar Patterns 2: Nouns and Adjectives*. London: HarperCollins.
- Fraser, Bruce. 1990. Perspectives on politeness. *Journal of Pragmatics* 14(2), 219–236.
- Fraser, Bruce & William Nolen. 1981. The association of deference with linguistic form. *International Journal of the Sociology of Language* 27, 93–109.
- Gillings, Mathew & Gerlinde Mautner. 2024. Concordancing for CADS: Practical challenges and theoretical implications. *International Journal of Corpus Linguistics* 29(1), 34–58.
- Grice, Herbert P. 1975. Logic and conversation. In Peter Cole & Jerry L. Morgan (eds.), *Syntax and semantics*, 41–58. New York: Academic Press.
- Hardie, Andrew. 2012. CQPweb – combining power, flexibility and usability in a corpus analysis tool. *International Journal of Corpus Linguistics* 17(3), 380–409.
- Hartling, Lisa, Michele Hamm, Andrea Milne, Ben Vandermeer, P. Lina Santaguida, Mohammed Ansari, Alexander Tsertsvadze, Susanne Hempel, Paul Shekelle & Donna M. Dryden. 2012. Validity and Inter-rater Reliability Testing of Quality Assessment Instruments. *Agency for Healthcare Research and Quality*, Publication No. 12-EHC039-EF, Rockville MD.
- Haugh, Michael & Derek Bousfield. 2012. Mock impoliteness in interactions amongst Australian and British speakers of English. *Journal of Pragmatics* 44(9), 1099–1114.
- Hughes, Geoffrey. 2006. *An Encyclopedia of Swearing: The Social History of Oaths, Profanity, Foul Language, and Ethnic Slurs in the English-speaking World*. London: Routledge.
- Jain, Kate H. 2022. You Hoboken! Semantics of an expressive label maker. *Linguistics and Philosophy* 45(2), 365–391.
- Jay, Timothy B. 2000. *Why We Curse*. Philadelphia: John Benjamins.
- Jucker, Andreas & Daniela Landert. 2023. The diachrony of im/politeness in American and British movies (1930–2019). *Journal of Pragmatics* 209, 123–141.
- Julien, Marit. 2016. Possessive predicational vocatives in Scandinavian. *Journal of Comparative German Linguistics* 19, 75–108.
- Kleinke, Sonja & Birte Bös. 2015. Intergroup rudeness and the metapragmatics of its negotiation in online discussion fora. *Pragmatics* 25(1), 47–71.
- Lai, Xiaoyu. 2019. Impoliteness in English and Chinese online diners' reviews. *Journal of Politeness Research* 15(2), 293–322.
- Leech, Geoffrey. 1983. *Principles of Pragmatics*. London: Routledge.
- Locher, Miriam & Richard J. Watts. 2008. Relational work and impoliteness: negotiating norms of linguistic behaviour. In Derek Bousfield & Miriam Locher (eds.), *Impoliteness in Language: Studies on its interplay with power in theory and practice*, 77–99. Berlin: Mouton de Gruyter.
- Love, Robbie, Claire Dembry, Andrew Hardie, Vaclav Brezina & Tony McEnery. 2017. The spoken BNC2014: Designing and building a spoken corpus of everyday conversations. *International Journal of Corpus Linguistics* 22(3), 319–344.

- Love, Robbie & Anna-Brita Stenstrom. 2023. Corpus-pragmatic perspectives on the contemporary weakening of fuck: The case of teenage British English conversation. *Journal of Pragmatics* 216. 167–181.
- Mackey, Alison & Susan M. Gass. 2021. *Second Language Research: Methodology and Design*. 3rd edn. London: Routledge.
- McEnery, Tony. 2006. *Swearing in English: Bad Language, Purity and Power from 1586 to the Present*. London: Routledge.
- McHugh, Mary L. 2012. Interrater reliability: the kappa statistic. *Biochem Med (Zagreb)* 22(3). 276–282.
- Millwood-Hargrave, Andrea. 2000. Delete expletives? *Report for the Advertising Standards Authority, British Broadcasting Corporation, Broadcasting Standards Commission and the Independent Television Commission*. https://www.biblit.it/wp-content/uploads/2014/08/ASA_Delete_Expletives_Dec_2000-2.pdf
- Montagu, Ashley. 2001 [1967]. *The Anatomy of Swearing*. Pennsylvania: University of Pennsylvania Press.
- Rabab'ah, Ghaleb & Nusiebah Alali. 2020. Impoliteness in reader comments on the Al-Jazeera Channel news website. *Journal of Politeness Research* 16(1). 1–43.
- Ruhi, Şükriye & Yeşim Aksan (eds.). 2015. *Exploring (im)politeness in specialized and general corpora: Converging methodologies and analytic procedures*. Cambridge: Cambridge Scholars Publishing.
- Ruzaitė, J. 2023. Impoliteness categories in hateful online comments targeting migrants in Lithuania. In Victoria Guillén-Nieto, Antonio Doval Pais & Dieter Stein (eds.), *From Fear to Hate: Legal-Linguistic Perspectives on Migration*, 117–148. Berlin: De Gruyter.
- Šarić, Lilijana. 2022. Impoliteness strategies in Croatian and Serbian user comments on online news articles: a study based on readers' perceptions. *Jezikoslovlje* 23(1). 1–34.
- Sinkeviciute, Valeria. 2014. “When a joke's a joke and when it's too much”: *Mateship* as a key to interpreting jocular FTAs in Australian English. *Journal of Pragmatics* 60. 121–139.
- Teneketzi, Korallia. 2022. Impoliteness across social media platforms: A comparative study of conflict on YouTube and Reddit. *Journal of Language Aggression and Conflict* 10(1). 38–63.
- Tsoumou, Jean M. 2022. (Im)politeness on Facebook during the Covid-19 pandemic. *Journal of Politeness Research* 19(1). 249–284.
- Tsoumou, Jean M. 2023. Impoliteness among multilingual Facebook users in Congo Brazzaville. *Journal of Politeness Research* 19(2). 521–555.
- Terkourafi, Marina. 2001. *Politeness in Cypriot Greek: A Frame-Based Approach*. Cambridge, UK: University of Cambridge dissertation.
- Terkourafi, Marina. 2002. Politeness and formulaicity: evidence from Cypriot Greek. *Journal of Greek Linguistics* 3(1). 179–201.
- Terkourafi, Marina. 2005a. Pragmatic correlates of frequency of use: The case for a notion of “minimal context”. In Sophia Marmaridou, Kiki Nikiforidou & Eleni Antonopoulou (eds.), *Reviewing Linguistic Thought: Converging Trends for the 21st Century*, 209–233. Berlin: Mouton de Gruyter.
- Terkourafi, Marina. 2005b. Beyond the micro-level in politeness research. *Journal of Politeness Research* 1(2). 237–262.
- Van der Bom, Isabelle, & Sara Mills. 2015. A discursive approach to the analysis of politeness data. *Journal of Politeness Research* 11(2). 179–206.
- van Dorst, Isolde, Mathew Gillings & Jonathan Culpeper. 2024. Pragmatic variation in Britain: a corpus-based study of politeness and social variation. *Journal of Pragmatics* 227. 37–56.

- Van Olmen, Daniel, Marta Andersson & Jonathan Culpeper. 2023. Inherent linguistic impoliteness: The case of insulting YOU+NP in Dutch, English and Polish. *Journal of Pragmatics* 215. 22–40.
- Vergis, Nikos. 2017. The interaction of the Maxim of Quality and face concerns: An experimental approach using the vignette technique. *Journal of Pragmatics* 118. 38–50.
- Vergis, Nikos & Marina Terkourafi. 2015. The role of the speaker's emotional state in im/politeness assessments. *Journal of Language and Social Psychology* 34(3). 316–342.
- Watts, Richard J. 2003. *Politeness*. Cambridge: Cambridge University Press.